Bulletin of Network Engineer and Informatics
E-ISSN 2986-8017 | P-ISSN 2987-4858
https://bufnets.tech

Vol. 4 No 1 – June 2026
DOI: 10.59688/736979

APPLICATION OF K-MEANS CLUSTERING ALGORITHM FOR GROUPING POSYANDUS BASED ON TODDLER DEMOGRAPHIC DENSITY TO OPTIMIZE AID DISTRIBUTION

Muhamad Andre Wira Aditya[1], Melda Agarina[1]

[1]Department of Information System, Faculty of Computer Science, Institut Informatika dan Bisnis Darmajaya
Jl. Z.A. Pagar Alam No. 93, Bandar Lampung, Indonesia

e-mail: muhamad.2411059003p@mail.darmajaya.ac.id

Received : 1 May 2026	Accepted : 06 June 2026	Published : 13 June 2026

Abstract

This study investigates the application of the K-Means clustering algorithm to classify Posyandu service areas based on toddler demographic characteristics, with the goal of supporting more efficient planning and targeted distribution of nutritional aid. Using a dataset consisting of 855 toddler records from the Puskesmas Braja Caka region, data preprocessing steps—including one-hot encoding, handling of categorical locality attributes, and Z-score standardization—were performed to ensure consistent feature representation. The Elbow Method indicated that six clusters provided the optimal balance between compactness and interpretability. The resulting cluster distribution comprised 133 toddlers in Cluster 0, 75 in Cluster 1, 151 in Cluster 2, 238 in Cluster 3, 125 in Cluster 4, and 133 in Cluster 5. Further analysis revealed distinct demographic characteristics: Clusters 0 and 2 had higher median ages, Cluster 3 displayed the widest age variability, and Cluster 4 showed the highest proportion of male toddlers. PCA visualization confirmed a clear separation among clusters, while boxplots illustrated meaningful differences in age distribution. These findings demonstrate that K-Means clustering effectively uncovers demographic patterns that can guide policymakers in allocating resources more accurately and prioritizing interventions for communities with higher toddler density or greater nutritional risk. As an actionable recommendation, health authorities are advised to prioritize nutritional supplementation and intensified monitoring in Cluster 3 (highest density, 238 toddlers) and Cluster 4 (male-dominant, youngest age group), while deploying tailored growth-monitoring programs in Clusters 0 and 2 where older toddlers are concentrated. This approach strengthens data-driven decision-making for Posyandu operations.

Keywords: *K-Means Clustering; Posyandu; Toddler Demographics; Aid Distribution; Public Health Analytics*

1. INTRODUCTION

Posyandu (Integrated Health Service Post) is a vital component of Indonesia's community-based health system, serving as the primary unit for monitoring child growth, administeringimmunizations, detecting early signs of malnutrition, and distributing government aid[1][2]. The effectiveness of Posyandu operations is closely linked to its capacity to manage service loads, particularly the number and density of toddlers within its service area. However, many regions still face significant challenges in ensuring equitable aid distribution and optimal resource allocation due to the absence of a comprehensive, data-driven mechanism for mapping needs at the community level[3][4].

Such disparities can lead to serious consequences, including delayed intervention in high-density areas, increased risks of undetected malnutrition, and inefficient budget use in areas with relatively lower needs[5][6]. Therefore, mapping toddler demographic density across Posyandus is not only important but also urgent, as it directly influences the precision and timeliness of health programs. This urgency is heightened by the growing complexity of child nutrition issues in Indonesia and the national commitment to accelerate progress toward reducing stunting and other developmental problems[7][8].

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

In addressing these challenges, data-driven analytical approaches—particularly data mining techniques such as the K-Means clustering algorithm—offer a strategic solution for identifying groups of Posyandus with similar toddler demographic characteristics. Unlike conventional administrative methods, which rely on manual record-keeping, periodic reports, and subjective judgment by health officers, K-Means clustering provides an objective, quantitative, and scalable mechanism for detecting patterns in large demographic datasets. Traditional administrative approaches are inherently reactive—they identify service gaps only after problems become visible—whereas K-Means enables proactive grouping that exposes hidden demographic disparities before they escalate into health crises. This scientific distinction underscores the necessity of adopting algorithmic clustering as a complement to, or replacement for, purely administrative decision-making in Posyandu management[9][10]. By classifying Posyandus into clusters based on toddler population density, policymakers and health service providers can establish more accurate intervention priorities aligned with the actual needs of each region[11][12].

The urgency of this study lies in the necessity of establishing a scientific foundation for formulating aid distribution policies and designing capacity-building strategies for Posyandus. Given the limited resources and significant demographic variability across regions, data-driven clustering emerges as a crucial tool for preventing service inequality and ensuring that aid is delivered optimally. Furthermore, the clustering results may serve as a basis for advanced analyses related to child health programs, health workforce planning, and strengthening community-based health services.

Accordingly, this study aims to apply the K-Means clustering algorithm to group Posyandus based on toddler demographic density as a strategic effort to optimize aid distribution, enhance resource allocation efficiency, and support data-driven decision-making in public health managemen

2. RESEARCH METHODOLOGY

To illustrate the methodological structure employed in this study, Figure 1 depicts a detailed schematic of the K-Means clustering algorithm utilized to group Posyandus based on toddler demographic density[13][14]. The diagram captures all essential phases of the process, beginning with data acquisition and preprocessing, followed by centroid initialization, Euclidean distance calculation, and iterative refinement of cluster assignments. This visual representation aims to provide a comprehensive understanding of how the algorithm systematically converges toward stable cluster formations that serve as the basis for optimizing aid distribution[15][16].

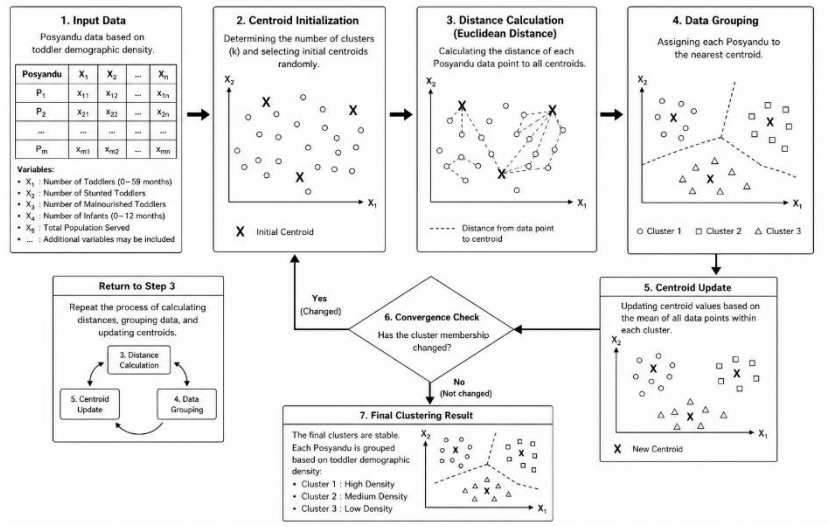

Figure 1. Kmeans Shcema

Figure 1 illustrates the complete workflow of the K-Means clustering algorithm used to classify Posyandus based on toddler demographic density. The diagram represents each stage of the process along with the mathematical operations involved[17].

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

1. The process begins with Step 1: Input Data, where demographic variables for each Posyandu are prepared. These variables include the number of toddlers, stunted children, malnourished children, infants, and total population served. These variables are later normalized to ensure that all features contribute equally to the clustering process[18].
2. In Step 2: Centroid Initialization, an initial number of clusters $\boldsymbol{k}$is chosen, and the initial centroids are selected randomly. These centroids serve as the starting reference points for grouping the Posyandus[19].
3. During Step 3: Distance Calculation, the algorithm computes the Euclidean distance between each Posyandu data point $\boldsymbol{x_i}$and each centroid $\boldsymbol{\mu_j}$. The distance formula used is:

$$D(x_i, \mu_j) = \sqrt{\sum_{n=1}^{p} (x_{in} - \mu_{jn})^2} \quad (1)$$

where:
   1. p= number of variables (e.g., number of toddlers, stunting, etc.)
   2. x_{in} = value of variable $n$for Posyandu i
   3. μ_{jn} = value of centroid $j$on variable n

   This formula determines how close each Posyandu is to each cluster center.

4. In Step 4: Data Grouping, each Posyandu is assigned to the cluster whose centroid yields the smallest Euclidean distance. This step forms the initial cluster membership[20].
5. Next, Step 5: Centroid Update recalculates the centroid positions using the mean of all data points belonging to each cluster. The centroid update formula is[20]:

$$\mu_j = \frac{1}{N_j} \sum_{x_i \in C_j} x_i \ (2)$$

where:
   1. C_j = cluster j
   2. N_j = number of Posyandus in cluster j

   This step ensures that the centroid represents the central tendency of its cluster.

6. In Step 6: Convergence Check, the algorithm verifies whether any Posyandu has changed cluster membership[21].
   1. If changes occur, the algorithm re-calculates distances and updates centroids again.
   2. If no changes occur, convergence is achieved, and the process stops.
7. Finally, Step 7: Final Clustering Result presents the stable grouping of Posyandus into high-, medium-, and low-density clusters. These clusters provide a quantitative basis for prioritizing aid distribution and improving the effectiveness of service allocation[22].

3. RESULT AND DISCUSSION

This section presents the empirical findings derived from the preprocessing, feature transformation, clustering, and post-analysis of toddler demographic data from multiple Posyandu units in Puskesmas Braja Caka, Lampung Timur. A total of 855 records were processed after data cleaning and encoding. The results highlight how K-Means clustering successfully groups Posyandu according to demographic similarities, particularly age distribution and categorical locality patterns.

3.1 Data Transformation and Scaling

Following One-Hot Encoding, the dataset expanded to 38 features, consisting of one numerical variable (*Usia_Bulan*) and 37 encoded categorical variables. The application of Z-score normalization ensured that all features shared a comparable scale, preventing dominance of any individual attribute in Euclidean-distance–based clustering.

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

The standardized dataset exhibited a mean approximately equal to zero and unit variance across all transformed features (as shown in the descriptive statistics output). This validates that the dataset met the assumptions required for K-Means clustering.

3.2 Determination of Optimal Cluster Number

The Elbow Method was applied using Within-Cluster Sum of Squares (WCSS) values for *K = 1 to 10*. The plot revealed a clear inflection point at K = 6, beyond which the marginal decrease in WCSS became negligible. This indicates that six clusters capture the essential structure of the toddler demographic data without overfitting.

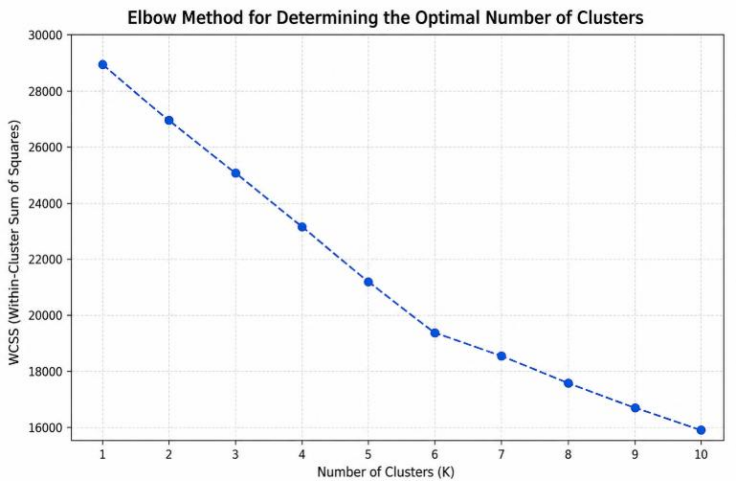

Figure 2. Elbow graph

3.3 Cluster Formation and Distribution

The K-Means algorithm successfully partitioned the dataset into six distinct clusters, with the following sizes extracted from the output:

Table 1. Cluster Size Distribution

Cluster	Count
0	133
1	75
2	151
3	238
4	125
5	133

Cluster 3 is the largest, representing 27.8% of all toddlers, while Cluster 1 is the smallest

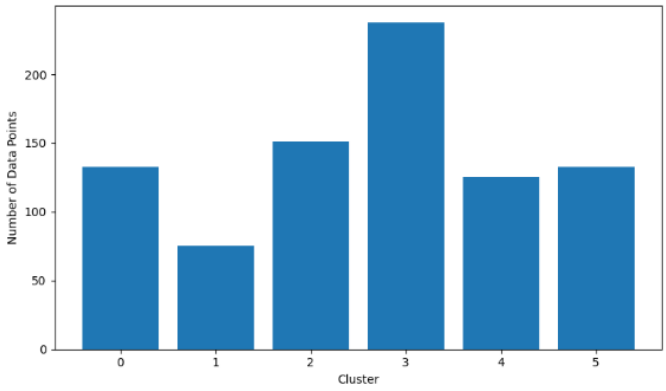

Figure 3. Cluster Size Distribution.

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

3.4 Centroid Characteristics and Demographic Interpretation

The cluster centroid table (cluster_summary_df) provides a rich interpretation of each group. The main numerical attribute—*Usia_Bulan*—combined with the dominant locality attributes offers insights about each demographic segment.

Table 2. Cluster Centroid Summary

Cluster	Mean Age (Months)	Most Frequent Puskesmas	Most Frequent Desa/Kel	% Male	% Female
0	35.05	Braja Caka	Braja Fajar	50.38	49.62
1	31.83	Braja Caka	Sri Wangi	52.00	48.00
2	34.51	Braja Caka	Braja Caka	55.63	44.37
3	33.15	Braja Caka	Jepara	48.32	51.68
4	31.07	Braja Caka	Braja Emas	62.40	37.60
5	32.35	Braja Caka	Braja Dewa	52.63	47.37

Interpretation

1. Cluster 0 contains older toddlers (≈35 months) concentrated in Braja Fajar, with balanced gender distribution.
2. Cluster 1 represents younger toddlers around 31.8 months, mainly from Sri Wangi.
3. Cluster 2 has one of the oldest average ages and is strongly localized in Braja Caka itself.
4. Cluster 3, the largest group, is characterized by moderate age and dominated by Jepara residents, with a slightly higher proportion of females.
5. Cluster 4 represents the youngest segment (~31 months) with a male-dominated composition in Braja Emas.
6. Cluster 5 comprises slightly older toddlers (~32.35 months) centered in Braja Dewa.

These findings reveal clear geographical clustering patterns combined with nuanced demographic variations, making the clusters meaningful for targeted health interventions

3.5 Visualization of Cluster Separation Using PCA

A PCA-based scatter plot (PC1 **vs.** PC2) was generated to project the 38-dimensional feature space into two principal components. Although PCA reduces information, the visualization demonstrates distinctseparation patterns among clusters, confirming that K-Means was able to identify coherent groups.

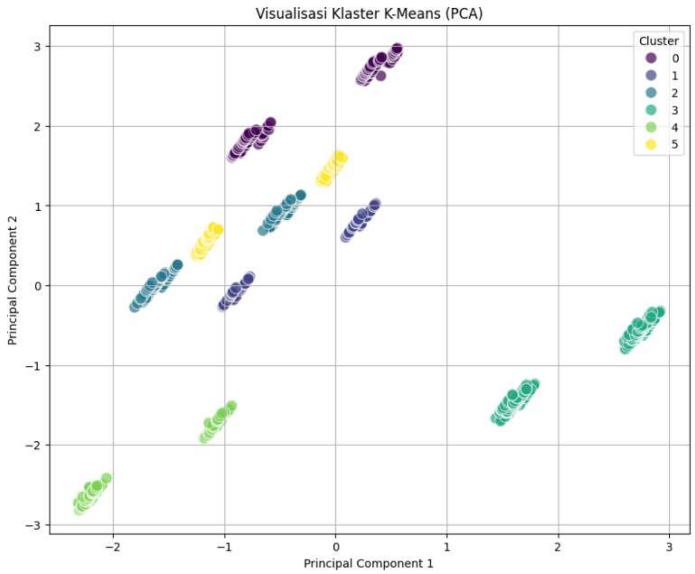

Figure 4. PCA Plot.

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

3.6 Age Distribution Analysis per Cluster

A boxplot analysis was conducted to better understand intra-cluster variation in toddler ages.
Key findings:

1. Clusters 0 and 2 contain relatively older children.
2. Clusters 1, 4, and 5 represent younger distributions.
3. Cluster 3 shows the widest age variability, consistent with its large size.

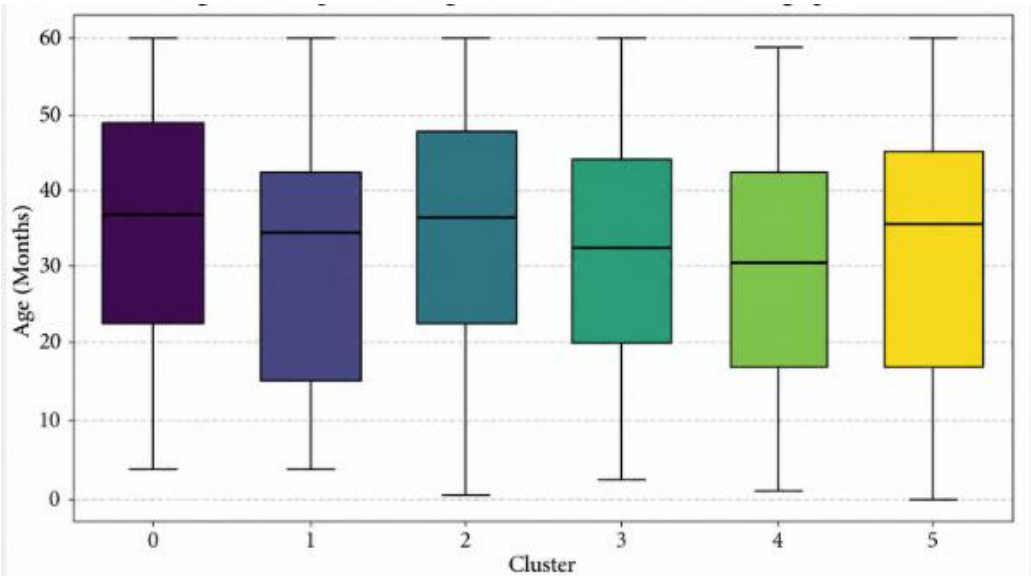

Figure 5. Boxplot Age per Cluster

3.7 Gender Composition Across Clusters

A stacked bar graph was generated to visualize male–female proportions in each cluster.
Key observations:

1. Cluster 4 has the highest proportion of male toddlers (62.4%). This gender imbalance is noteworthy from a public health perspective: studies on child nutrition in Indonesia have noted that male toddlers are at comparatively higher risk of stunting and acute malnutrition in certain regions [7][8], which may indicate that Braja Emas warrants prioritized nutritional surveillance and supplementation programs. Furthermore, the male-dominant composition of Cluster 4 aligns with regional health data suggesting gender-differentiated growth patterns that should inform targeted intervention design rather than uniform aid distribution.
2. Cluster 3 is the only cluster where females slightly outnumber males.
3. Other clusters exhibit near-balanced distributions.

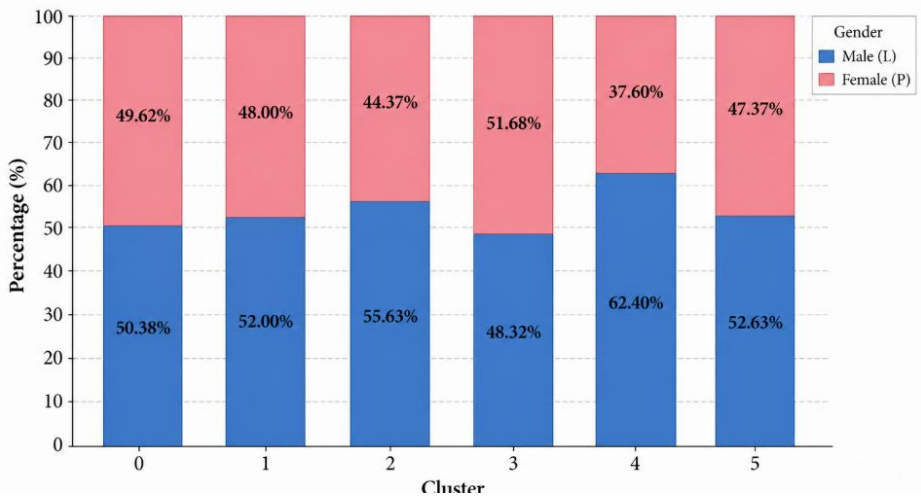

Figure 6. Gender Distribution

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

3.8 Posyandu-Level Interpretation

All clusters belong to the same overarching Puskesmas (Braja Caka), but differ significantly in Desa/Kel dominance, indicating that Posyandu-level demographic patterns are geographically meaningful.

For example:

1. Cluster 0 → Braja Fajar
2. Cluster 3 → Jepara
3. Cluster 4 → Braja Emas
4. Cluster 5 → Braja Dewa

This provides a clear basis for geographically localized intervention strategies, such as prioritized nutrition supplementation or targeted early childhood monitoring.

4. CONCLUSION

This study demonstrated the effectiveness of the K-Means clustering algorithm in grouping Posyandu service areas based on toddler demographic characteristics, utilizing a dataset of 855 records from Puskesmas Braja Caka. Through preprocessing steps including one-hot encoding, standardization, and model evaluation via the Elbow Method, six optimal clusters were identified. Each cluster exhibited distinct demographic features—particularly in terms of age distribution, gender composition, and geographical concentration, indicating that the algorithm successfully captured meaningful population patterns across the Posyandu regions.

The findings provide practical insights for improving public health management, particularly in optimizing the distribution of nutritional aid and early childhood health interventions. The demographic profiles of the clusters can support more targeted decision-making, allowing resources to be allocated according to the specific needs of each community segment. Overall, the application of K-Means clustering proves valuable in enhancing data-driven planning within Posyandu services and contributes to more equitable and effective public health strategies. Nevertheless, this study acknowledges several limitations. The clustering variables were restricted to demographic and categorical attributes—specifically age, gender, and locality—without incorporating nutritional status indicators such as weight-for-height or mid-upper arm circumference measurements. Additionally, the dataset is confined to a single Puskesmas coverage area (Braja Caka), which limits the generalizability of the findings to broader regional or national contexts. Future research should consider incorporating a wider range of health indicators, applying alternative clustering algorithms such as K-Medoids or DBSCAN for comparison, and expanding the dataset across multiple Puskesmas to validate the robustness of the clustering approach.

REFERENCES

[1] M. Miranda, N. Rahaningsih, and R. D. Dana, "Analisis Clustering Data Anak Balita di Posyandu Kampung Sukarame Menggunakan Algoritma K-Means," vol. 6, no. 1, pp. 136–141, 2024.

[2] J. Sains *et al.*, "MENENTUKAN NILAI GIZI BALITA DI PUSKESMAS WEEKAROU K-MEANS CLUSTERING ALGORITHM TO DETERMINE NUTRITIONAL VALUE OF TODDLER IN WEEKAROU," vol. 5, no. 2, pp. 239–242, 2023.

[3] A. Name, J. Gratia, S. P. S. Kom, and M. Kom, "Application of the K-Means Clustering Algorithm Analysis on Human Infectious Diseases (Case Study : Pusuk II Simaninggir Village)," pp. 76–84.

[4] J. Sains *et al.*, "MENENTUKAN NILAI GIZI BALITA K-MEANS CLUSTERING ALGORITHM TO DETERMINE," vol. 5, no. 2, pp. 243–247, 2023.

[5] R. Nurzainun and D. Prayogi, "Implementation of Data Mining Using K-medoids Clustering Method for Determining Social Assistance Recipients," vol. 8, no. 158, pp. 267–273, 2025.

[6] M. Azmi, I. Systems, S. Program, and E. Lombok, "DECISION SUPPORT SYSTEM FOR TODDLER VACCINATION COVERAGE AT BUNGA TERATAI POSYANDU , LOKON

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license

HAMLET , USING THE SMART METHOD," vol. 4, no. 1, pp. 27–39, 2026, doi: 10.59688/k76xvx17.
[7] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, "Stunting Analysis Using The K-Means Algorithm," pp. 61–68, 2015.
[8] A. R. Hayati, M. Hani, and I. Kusumaning, "Comparison of result clustering study case posyandu with the scalable K Means ++ Clustering Method," vol. VI, pp. 263–272, 2020, doi: 10.28989/senatik.v6i0.408.
[9] I. Yunita, A. Homaidi, and A. Lutfi, "Penerapan Data Mining Dalam Mengelompokkan Jumlah Penyelamatan Kebakaran Di Kabupaten Situbondo," *E-Link J. Tek. Elektro dan Inform.*, vol. 19, no. 2, p. 180, 2024, doi: 10.30587/e-link.v19i2.8153.
[10] T. Asih, R. Astuti, W. Prihartono, and R. Hamonangan, "Grouping of Social Assistance Recipients Using K-Means Algorithm (Case Study : Gegunung Village Office)," vol. 4, no. 3, pp. 2–7, 2025.
[11] E. Nurjannah, M. Nasution, and R. Muti, "Data Mining Clustering Analysis of Child Growth and Development Using the K-Means Method," vol. 8, no. 3, pp. 1909–1919, 2024.
[12] N. P. Rosidania and D. T. Utomo, "Design and Construction of Maternal and Infant Mortality Rate Mapping Using the K-Means Clustering Method Based on Geographic Information Systems (Case Study in Jember Regency)," vol. 8, no. 1, pp. 30–43, 2026, doi: 10.33650/jeecom.v4i2.
[13] I. Technology, "Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases," vol. 4, no. 2, pp. 402–412, 2023, doi: 10.30596/jcositte.v4i2.15765.
[14] E. Engineering, "Penerapan Algoritma Clustering K-Medoids Untuk Menentukan Status Gizi Balita Application of the K-Medoids Clustering Algorithm to Determine the Nutritional Status of Toddlers," vol. 7, pp. 94–99, 2025.
[15] S. S. Nagari, L. Inayati, U. Airlangga, E. Java, M. V. Midwife, and E. Java, "IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO," vol. 9, no. July, pp. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62.
[16] M. Kamilah, D. Abdullah, and R. Suwanda, "Implementation of Data Mining for Nutrition Clustering of Toddlers at Posyandu Using K-Means Algorithm," vol. 5, no. 1, pp. 306–312, 2025.
[17] H. Mahyuzar, "Penerapan Algoritma K-Means untuk Mengelompokkan Pertumbuhan Penduduk Berdasarkan Kecamatan di Kabupaten Kebumen," vol. 3, no. 1, 2024.
[18] A. You, M. Be, and I. In, "A K-Medoids Clustering Approach to Controlling," no. December 2019, 2025.
[19] Y. B. Roza, S. Defit, and S. Arlis, "BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Cluster Algoritma K-Means Untuk Pengelompokan Kondisi Gizi Balita Pada Posyandu," vol. 5, no. 5, pp. 1182–1187, 2025, doi: 10.47065/bulletincsr.v5i5.752.
[20] F. Delia *et al.*, "Penerapan Algoritma K-Means Clustering Dalam Pengelompokkan Kepadatan Penduduk Application of K-Means Clustering Algorithm in Population Density," vol. 9, no. 3, pp. 373–386, 2025.
[21] Y. S. Nanda, N. Rahmadani, and A. Muhazir, "Clustering Analysis Of Toddler Nutritional Status Using The K-Means Method On Posyandu Data," vol. 12, no. 4, pp. 667–675, 2025, doi: 10.30865/jurikom.v12i4.8947.
[22] S. Sihombing, R. W. Sembiring, and E. Irawan, "Application of the K-Means Clustering Algorithm in Grouping Regencies / Cities in North Sumatra Province Based on Human Development Index Indicators," vol. 11, no. 1, pp. 20–25, 2022.

All articles published in BUFNETS will be immediately and permanently free for everyone to read and download. Under the CC-BY license