Jurnal Teknik Informatika (JUTIF) Vol. 7, No. 2, April 2026, Page. 1741-1757
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

Clustering And Classification Of Toddler Stunting Risk Using K-Means And Naive Bayes: A Case Study At Kembaran 1 Community Health Center

Lulu Amnah Fitriya Maharani[1], Purwadi*[2], Debby Ummul Hidayah[3]

[1,3]Information Systems Study Program, Faculty of Computer Science, Amikom Purwokerto University, Indonesia
[2]Master of Computer Science, Faculty of Computer Science, Amikom Purwokerto University, Indonesia

Email: [2]purwadi@amikompurwokerto.ac.id

Received : Oct 31, 2025; Revised : Nov 3, 2025; Accepted : Nov 4, 2025; Published : Apr 18, 2026

Abstract

Stunting continues to be a significant public health concern in Indonesia, with a frequency of 17.25% at Kembaran 1 Public Health Center, highlighting ongoing difficulties in early childhood nutrition and growth surveillance. This work seeks to assess and forecast stunting risk in toddlers by employing K-Means clustering and Naive Bayes classification to enhance early detection precision. The K-Means method was utilized on 1,168 toddler growth records to categorize stunting features, whereas the Davies–Bouldin Index (DBI) was employed to evaluate cluster quality. The ideal cluster was attained at k = 8, yielding a DBI value of 4.353, indicating compact and distinctly differentiated clusters. The Naive Bayes classifier subsequently predicted stunting potential with an accuracy of 93.56%, accurately categorizing 218 out of 233 test examples, yielding precision, recall, and F1-score values for the “short” class of 97.41%, 94.95%, and 96.18%, respectively. The findings indicate that the hybrid model successfully combines unsupervised and supervised learning, improving stunting prediction accuracy and cluster interpretability. The research provides a data-centric framework for localized stunting surveillance, aiding community health centers in formulating targeted early treatments and mitigating long-term developmental hazards.

Keywords: K-Means Clustering, Naive Bayes, Stunting Prediction, Toddler Health.

1. INTRODUCTION

Early infancy is a pivotal phase in human development characterized by swift physical, mental, and cognitive advancement. Growth denotes quantifiable increments in height, weight, and head circumference, whereas development pertains to the enhancement of motor skills, communication, cognitive abilities, and emotional regulation. The nutritional condition of toddlers is closely linked to this process. Optimal growth and development in toddlers occur with enough dietary intake, in contrast, inadequate nutrition results in growth failure and developmental delays. Stunting continues to be the most widespread chronic nutritional issue among Indonesian children [1].

Stunting is a disorder characterized by the failure to achieve normal height due to prolonged nutritional deficiency, beginning in gestation and continuing throughout early life. The World Health Organization (WHO) defines stunting as a height-for-age (H/A) measurement that is below –2 standard deviations (SD) from the median of the WHO Child Growth Standards. Causes of stunting can be classified as direct (insufficient protein consumption, poor breastfeeding, infectious disorders) and indirect (unsanitary environments, low parental education, and poverty) [2], [3]. Previous studies have

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

shown that stunting is influenced by nutritional intake [4], parental knowledge [3], environmental sanitation [1], and socioeconomic conditions [5].

The ramifications of stunting transcend mere physical growth deficiencies, encompassing heightened morbidity, diminished cognitive capacity, and lowered productivity, all of which obstruct human capital development [6]. Stunting diminishes national competitiveness, perpetuating cycles of intergenerational poverty [5]. Consequently, the reduction of stunting has been incorporated into the Global Nutrition Targets 2025 and the Sustainable Development Goals (SDG 2.2 – Zero Hunger) as a critical measure of public health achievement [7].

Notwithstanding ongoing national efforts, Indonesia's stunting rate persists at a relatively elevated level. According to the 2024 Indonesian Nutritional Status Study (SSGI), the nationwide incidence declined from 37.6% in 2013 to 19.8% in 2024. Central Java registered a rate of 17.1%, although Banyumas Regency exhibited a higher rate of 19.6%, surpassing the regional norm [8]. The Kembaran 1 Public Health Center recorded a frequency of 17.25% among eight villages. These statistics underscore the necessity for focused early detection systems capable of tackling stunting at the community level. Community empowerment has been shown to play a crucial role in stunting prevention, where active participation of families and local stakeholders contributes significantly to improving child nutrition and growth outcomes [9]

Recent analyses indicate that despite the national decline to 19.8% (SSGI 2024), several districts such as Banyumas (19.6%) remain above the provincial mean. This emphasizes the urgency of integrating local health data analytics to strengthen early stunting surveillance at the Community Health Center level [6], [8], [10]. Studies conducted between 2023 and 2025 indicate that hybrid data mining models can improve public health prediction accuracy by approximately 5–10% compared to traditional single-model approaches [11]. Clustering methods such as K-Means contribute by identifying latent growth patterns that enhance downstream classification performance [12], while hybrid classifiers such as Naive Bayes–C4.5 demonstrate measurable accuracy gains [13].

While stunting has been extensively examined using public health methodologies, a comprehensive understanding necessitates the incorporation of computer science and data analytics. Machine learning and data mining methodologies can reveal concealed patterns in the drivers of stunting via clustering, classification, and predictive modeling [14].

In health informatics, hybrid learning methodologies that integrate unsupervised and supervised algorithms are progressively employed to improve decision-making precision [15], [16]. Previous studies have demonstrated that integrating K-Means clustering with the Naive Bayes classifier can significantly improve classification accuracy, particularly in prediction tasks involving heterogeneous datasets [17]. This methodological advantage supports the application of hybrid models for early stunting risk prediction using local health data [10]. Several studies have demonstrated that clustering accuracy can be improved through optimized initial centroid selection in K-Means, resulting in more stable and compact clusters [18]. The effectiveness of Naive Bayes observed in this study aligns with prior applications in disease classification across diverse health domains, including coastal and environmental health settings [19]. This necessitates the application of K-Means and Naive Bayes in a hybrid framework for local health data, particularly at the community health center tier.

The K-Means technique has been successfully employed for data clustering to categorize toddlers based on nutritional conditions. Pratistha and Kristianto [20] identified optimal clustering at K = 9, yielding a Davies–Bouldin Index (DBI) value of –0.673. Fadilah et al. [21] using the Elbow Method, discerning two clusters (ΔSSE = 1401.5156) that distinguished high and low stunting-risk regions, whereas Julyantari et al. [22] categorized child nutritional status into three clusters: normal (47.83%), poor (30.43%), and overnutrition (21.74%).

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

Conversely, the Naive Bayes algorithm is extensively employed to forecast the probability of stunting. Despite the variability in outcomes, Arlita et al. [23] attained an accuracy of 88% utilizing Naive Bayes. Kamil and Wibowo [24] utilized Naive Bayes on 958 records from the Cilandak Public Health Center, with an accuracy of 85.76%. Meriyana et al. [25] compared K-Nearest Neighbor and Naive Bayes algorithms for toddler stunting classification at the village level and reported that Naive Bayes provides competitive accuracy with simpler probabilistic assumptions. Ridwan and Sari [26] observed that Naive Bayes was marginally less accurate (0.93% lower) than the C4.5 algorithm in classifying toddler nutritional status. Gurning et al. demonstrated that the Naive Bayes classifier, combined with Chi-Square feature selection, is effective in predicting family-level stunting risk, confirming the suitability of probabilistic classifiers for early stunting detection [27]. Mirantika et al. [10] and Alamin and Aryani [11] demonstrated that the Naive Bayes algorithm can be effectively applied for early detection of stunting risk, supporting its suitability for predictive analysis in community health settings.

Comparative studies highlight Naive Bayes' limitations when compared to other algorithms. Widhari et al. [28] found that NB achieved 71% accuracy, 71% precision, 76% recall, and 73% F1-Score on 30% test data, whereas KNN performed best with 97% accuracy, 98% precision, and 96% recall. Similarly, Chilyabanyama et al. [29] in Zambia found Random Forest as the best-performing algorithm (79% accuracy), while Naive Bayes was the least accurate in predicting stunting among children. Alamin and Aryani [11] compared K-Nearest Neighbor, Naive Bayes, and Support Vector Machine algorithms for classifying toddler stunting status at a community health center and reported that each algorithm exhibits different strengths depending on data characteristics.

Based on the literature, most previous studies have used K-Means and Naive Bayes separately, leading to partial insights. Few have integrated both algorithms to form a complementary system for data clustering and stunting prediction simultaneously. Moreover, comprehensive evaluation using Davies–Bouldin Index (DBI) for cluster validity and classification accuracy within a single dataset is rarely performed in community-level health data [30]. Therefore, this research bridges that gap by integrating K-Means and Naive Bayes into a unified hybrid analysis framework to enhance prediction accuracy and improve local stunting risk mapping.

The novelty of this study lies in the simultaneous evaluation of Davies–Bouldin Index (DBI) and classification accuracy on real 2025 local health data from Kembaran 1 Community Health Center. Unlike prior works that used simulated or regional datasets, this research applies hybrid clustering–classification directly on community-based health records, enhancing local model interpretability and policy relevance.

Stunting continues to hinder the development of human capital in Indonesia, both nationally and locally. The cognitive and physical impairments they cause affect community productivity and national competitiveness. The government has implemented several programs, including nutrition education, supplementary feeding (PMT) at posyandu, and the promotion of exclusive breastfeeding for six months and continued until two years [22], [34], [35]. Regular pregnancy checks and growth monitoring serve as early detection mechanisms, particularly through routine height measurement at posyandu. However, the absence of integrated analytical tools for local health data limits early stunting detection.

In this study, data were obtained from the Kembaran 1 Community Health Center health database, containing toddler anthropometric records for 2025. The process involved data cleaning, normalization, and feature selection, followed by clustering with K-Means and classification using Naive Bayes. Model evaluation was conducted using Davies–Bouldin Index (DBI) for clustering quality and accuracy, precision, recall, and F1-score for classification performance.

The novelty of this research lies in integrating K-Means clustering and Naive Bayes classification into one analytical pipeline applied to local Community Health Center data (2025). This hybrid approach

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

not only provides better accuracy but also ensures interpretability and relevance for community-level health management. The study contributes to health informatics by proposing a data-driven early detection framework to support targeted interventions and improve child nutrition programs.

Table 1. Comparative Studies of Hybrid K-Means–Naive Bayes Applications (2020–2025)

Year	Authors	Dataset / Domain	K-Means Evaluation	Naive Bayes Accuracy	Hybrid Findings
2024	Melyani [13]	NB–C4.5 hybrid	—	93.1%	Accuracy improved 4%
2022	Sari & Harianis [14]	Nutrition classification	SSE diff = 1240.9	88.5%	Improved cluster separation
2024	Pratistha & Kristianto [20]	Stunting case clustering (regional)	DBI = -0.673 (K=9)	-	DBI = −0.673 at K = 9 indicates good cluster separation in regional stunting cases.
2022	Fadilah et al. [21]	Stunting factors (district/city)	DBI = 0.63	88.9%	Clear clusters (Silhouette score increased; DBI = 0.63)
2025	Rohman et al. [31]	Stunting risk factors (Indonesia)	DBI = 0.41 (optimal k)	89.6%	DBI minimized to 0.41 (compact clusters)
2024	Dwinanto et al. [32]	Stunting risk prediction	—	92.7%	Balanced recall/precision
2021	Titimeidara & Hadikurniawati [33]	Stunting nutritional status	—	≈88–90%	Consistent NB (Recall ≈ 89%)
2025 (This Study)	**Maharani et al.**	**Kembaran 1 Community Health Center (1,168 records)**	**DBI = 4.353**	**93.56%**	**Best hybrid accuracy on local dataset**

Based on the background above, the research problems addressed are: (1) How can the K-Means algorithm group stunted toddlers based on Height-for-Age (TB/U)? (2) How does the Naive Bayes algorithm predict stunting potential using clustered data? (3) How accurate is the hybrid K-Means–Naive Bayes model compared to using each algorithm separately?

The objectives of this study are to: (1) Apply K-Means to toddler stunting data and evaluate clusters using DBI; (2) Apply Naive Bayes to predict stunting potential from clustered data; (3) Evaluate hybrid model accuracy and demonstrate its applicability for local stunting mapping.

This research aims to advance the integration of machine learning in public health, providing a replicable, interpretable, and data-based framework for early detection and prevention of stunting among toddlers in Indonesia.

2. METHOD

Figure 1 depicts the comprehensive workflow of this research, outlining the sequential procedure from data collection to model deployment and evaluation. The methodology combines unsupervised clustering (K-Means) with supervised classification (Naive Bayes) within the RapidMiner Studio environment, selected for its visual workflow interface and repeatability capabilities [36] .

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

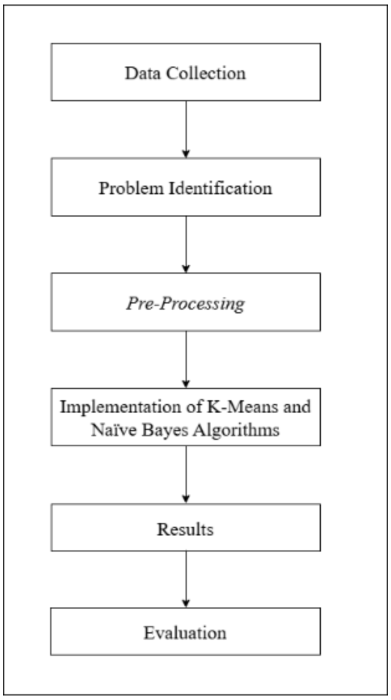

Figure 1. Research Stages

2.1. Data Collection

This study was conducted with official authorization from Kembaran 1 Community Health Center and the Banyumas District Health Office. The dataset consisted of 1,168 toddler records collected in 2025, containing 15 attributes reflecting anthropometric and demographic indicators (Table 2).

Table 2. Attributes Used in This Study

Label	Description
Name	Child's Name
Gender	Gender (M/F)
Birth Weight	Birth Weight
Brith Height	Height at Birth
Village/Ward	Toddler's Residence
Age at Measurement	Age of the Toddler at the Time of Measurement
Weight	Toddler's Weight When Measured
Height	Height of Toddler When Measured
BB/U	Toddler Weight by Age
TB/U	Toddler Height by Age
ZS TB/U	Z-score of height-for-age according to WHO standards (Z-score <-2)
BB/TB	Weight for Height (Indicator of nutritional status based on the proportion of weight to height)
ZS BB/TB	Z-score of weight for height
Gain Weight	Is the weight increasing compared to the previous measurement (Y/N)? Y stands for Yes, while N stands for No.

The variable TB/U (Height-for-Age) was chosen as the main indicator because it is the official WHO measure for stunting classification, where Z-score < –2 SD indicates stunting [7], [8].

2.2. Problem Identification

The principal research issue was articulated as follows: (1) Methodology for clustering stunted toddlers utilizing the TB/U indicator, and (2) Predicting the risk of stunting by the accuracy assessment of the Naive Bayes classifier used to the clustered dataset.

This phase delineates the analytical emphasis, guaranteeing that the chosen variables directly enhance the validity and reproducibility of the predictive model [32].

2.3. Pre-Processing

Data preprocessing ensures that the dataset is clean, consistent, and ready for analysis. It includes data cleaning, normalization, and data splitting:

a. **Data Cleaning**
Missing values for TB/U and Z-Score (TB/U) were addressed by mean imputation for numerical characteristics (height, weight) and mode imputation for categorical attributes (gender, village). Entries with above 30% missing data were excluded to mitigate bias and preserve dataset integrity [32]

b. **Data Normalization**
All numerical attributes were standardized to a common range using Min–Max normalization, ensuring that clustering by K-Means was not biased by differing scales :

$$x^1 = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Information:
x' = normalized value (the result of normalization),
x = original data value of a given attribute,
x_{min} = minimum value of the attribute in the dataset,
x_{max} = maximum value of the attribute in the dataset.

c. **Feature Selection and Data Splitting.**
Attributes were chosen for their significance to indicators of toddler growth. The data were divided into 80% training and 20% testing subsets with the Naive Bayes data-splitting operator in RapidMiner to guarantee balanced class representation [2], [28], [36].

2.4. Implementation of K-Means and Naive Bayes Algorithms

2.4.1. K-Means Algorithm

K-Means is a centroid-based clustering technique that divides n data points into k clusters by minimizing the sum of squared errors (SSE) between the data points and their corresponding centroids [31], [37], [38]

The following represents the Euclidean distance between a data point x_i and the centroid c_k:

$$d(x_i|c_i) = \sqrt{\sum_{i=1}^{n}(x_{ij} - c_{kj})^2} \quad (2)$$

and the ideal K is found using the SSE (Elbow Method) formula, which is:

$$SSE = \sum_{k=1}^{k}\sum_{j=1}^{n_k}|x_i - c_k|^2 \quad (3)$$

in this case, n_k is the number of data points in cluster k, c_k is its centroid, and K is the number of clusters.

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

By graphing the SSE values against K, the Elbow Method was used to determine the ideal number of clusters by determining the point at which the rate of SSE reduction rapidly declines [37], [39]. This stage resulted in the toddlers being grouped into multiple clusters based on similarities in height for age.

2.4.2. Naive Bayes Algorithm

The probabilistic classification algorithm Naive Bayes is predicated on the Bayes Theorem, which presumes predictor independence [23]. Using the prior and likelihood of each attribute assigned a class label; it computes the posterior probability.

The posterior probability formula is:

$$P = (C_k|X) = \frac{P(X|C_k)\ x\ P(C_k)}{P(X)} \tag{4}$$

where:
$P(C_k \mid X)$ = posterior probability of class C_k given features X,
$P(X \mid C_k)$ = likelihood of features given class C_k,
$P(C_k)$ = prior probability of class C_k,
$P(X)$ = probability of observing X.

The prior probability for each class was computed as:

$$P(C_k) = \frac{N_{Ck}}{N} \tag{5}$$

where the overall dataset size is N, and the number of samples in class C_k is $N_{(C_k)}$.
The algorithm steps were:
1. Prepare labeled training and testing data.
2. Compute prior probability $P(C_k)$ from class frequencies.
3. Calculate conditional probabilities $P(X_j|C_k)$ for each feature.
4. Determine the posterior probability using Bayes' theorem.
5. Assign each test instance to the class with the highest posterior probability [25].

2.5. Results

The clustering results obtained using K-Means were visualized in scatter plots and centroid tables, providing insights into toddler grouping patterns.

Prediction results indicating each toddler's likelihood of being stunted or normal were produced by the Naive Bayes classifier. Predictive performance was evaluated by calculating evaluation metrics (Accuracy, Precision, Recall, and F1-Score) based on the confusion matrix [27], [28], [30], [40].

2.6. Evaluation

The evaluation stage is carried out using different approaches depending on the characteristics of the algorithm used.

2.6.1. K-Means Evaluation

Developed by Davies and Bouldin (1979), the Davies–Bouldin Index (DBI) was used to assess the quality of clustering. The Davies–Bouldin Index has been widely applied to evaluate clustering performance by measuring intra-cluster compactness and inter-cluster separation in various domain applications [41]
The formula is:

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j} \left(\frac{S_{i,} + S_j}{M_{ij}} \right) \tag{6}$$

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

where M_{ij} is the distance between cluster centroids and S_i and S_j are the average intra-cluster distances. Better cluster separation and compactness are indicated by a lower DBI score [37], [39]

2.6.2. Naive Bayes Algorithm Evaluation

Confusion matrices are used to evaluate the Naive Bayes technique. In classification scenarios, the confusion matrix aids in understanding class differences by displaying the number of data points classified correctly and incorrectly. Both actual and predicted values are included in the confusion matrix, which produces four values: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN), as shown in Table 3:

Table 3. *Confusion Matrix*

	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

Based on the confusion matrix, the classification performance was evaluated using Recall, Precision, Accuracy, and F1-Score, which are commonly employed metrics for assessing classification models. Let TP denote True Positive, TN True Negative, FP False Positive, and FN False Negative. The evaluation metrics are formulated as follows [27], [38]:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \tag{8}$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data} \tag{9}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{10}$$

These metrics collectively determine how well the Naive Bayes model predicts stunting based on the K-Means clustered data.

3. RESULT

The findings of using the K-Means and Naive Bayes algorithms on toddler stunting data from Kembaran 1 Community Health Center are shown in this section. RapidMiner Studio was used for all tests to guarantee data flow uniformity and reproducibility.

3.1. K-Means Algorithm Result

K-Means clustering was applied to 1,168 toddler records consisting of 15 attributes. The process began by importing the dataset into RapidMiner, performing preprocessing (handling missing TB/U, Z-score normalization), and applying Min–Max scaling.

3.1.1 Finding the Optimal K Value

Ten iterations (K = 2–10) were tested to find the ideal number of clusters (K). The Davies–Bouldin Index (DBI) was used to assess each iteration; increased cluster compactness and separation are indicated by smaller DBI values [37]

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

The results indicate that K = 8 is the optimal cluster value, yielding the lowest Davies–Bouldin Index (DBI) of 4.353 and an average centroid distance of 1,358,393. A lower DBI value indicates better intra-cluster compactness and inter-cluster separation, which is consistent with the clustering quality interpretation reported by Rohman et al. [31].

Table 4. Davies–Bouldin Index (DBI) Values

Nilai K	Avg. Within Centroid Distance	Davies Bouldin
2	1.372.436	14.515
3	1.370.050	12.763
4	1.368.000	13.355
5	1.365.307	10.492
6	1.363.169	12.405
7	1.361.703	10.933
8	1.358.393	4.353
9	1.358.814	5.383
10	1.357.529	6.305

3.1.2. Cluster Distribution and Model

The data is clustered into 8 groups, summarized in Table 5:

Table 5. Cluster Model (K = 8)

Cluster	Amount
Cluster_0	2 items
Cluster_1	1062 items
Cluster_2	7 items
Cluster_3	76 items
Cluster_4	3 items
Cluster_5	13 items
Cluster_6	3 items
Cluster_7	2 items
Total number of items	1168

Figure 2 shows the cluster visualization using a scatter plot generated by RapidMiner. The visualization demonstrates how toddler height distributions form natural groupings with Cluster_1 dominating, suggesting a concentration of similar growth patterns.

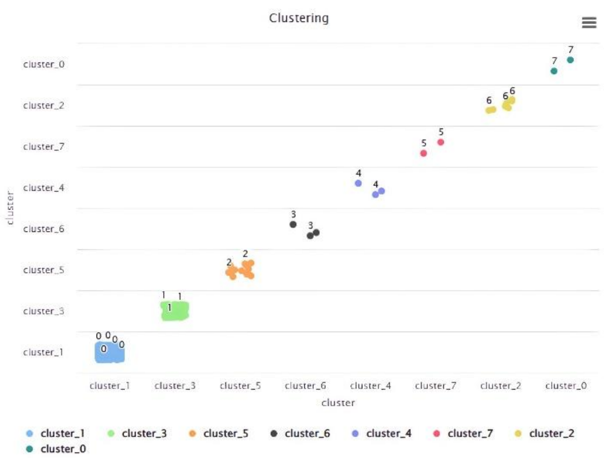

Figure 2. Cluster Visualization with Scatter Plo

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

Toddlers' nutritional health is categorized by using anthropometric measures, particularly Height-for-Age (TB/U). As shown in Table 6, the thresholds adhere to Indonesian Ministry of Health Regulation No. 2 of 2020 [42], which establishes the Anthropometric Standards for Children. These standards define the reference values for various growth indicators, such as weight-for-age, height-for-age, and weight-for-height, to assess the nutritional status and overall growth of children. They are essential for identifying children who may be at risk of malnutrition, stunting, or underweight, providing guidelines for health professionals and policymakers to monitor child development.

Table 6. Standards for Anthropometry Based on Toddler Height

Index	Category Nutritional Status	Threshold (Z-Score)
For kids ages 0–60 months, Body Length or Height by Age (PB/U or TB/U)	Very Short (*Severely Stunted*)	<-3 SD
	Short (*Stuned*)	- 3 SD to <- 2 SD
	Normal	-2 SD to + 3 SD
	Tall	> +3 SD

Table 6 indicates that children with Z-scores below –2 SD are classified as stunted, and those below –3 SD are categorized as severely stunted. The anthropometric cut-offs are the primary reference for assessing child development and categorizing stunting in this study [7], [8].

The distribution of stunted toddlers (Short and Very Short categories) across clusters is presented in Table 7.

Table 7. Distribution of Stunted Children by Cluster

Cluster	Short	Very Short
Cluster_0	2	0
Cluster_1	910	152
Cluster_2	4	3
Cluster_3	52	24
Cluster_4	3	0
Cluster_5	13	0
Cluster_6	3	0
Cluster_7	2	0
Total	**989**	**179**

From Table 7, 989 toddlers (84.7%) fall into the short category, while 179 (15.3%) are very short. The dominance of Cluster_1 (910 short and 152 very short) suggests that height-for-age is a powerful distinguishing variable in stunting analysis [20].

3.2. Naive Bayes Algorithm Results

The Naive Bayes classifier was utilized on the identical dataset to forecast stunting classifications ("Short" and "Very Short") based on Height-for-Age (TB/U). Figure 3 illustrates the performance evaluation panel from RapidMiner, showcasing the anticipated vs actual classifications:

accuracy: 93.56%

	true Pendek	true Sangat Pendek	class precision
pred. Pendek	188	5	97.41%
pred. Sangat Pendek	10	30	75.00%
class recall	94.95%	85.71%	

Figure 3. Naive Bayes Testing Results

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

3.2.1. Confusion Matrix

The categorization efficacy was assessed utilizing a confusion matrix (Table 8).

Table 8. Confusion Matrix

Predicted/Actual	Positive	Negative
Positive	188	5
Negative	10	30

Table 8 illustrates that out of 233 samples, 218 were correctly classified, resulting in an accuracy of 93.56% ± 0.02 after repeated validation runs.

3.2.2. Accuracy

$$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{Total Data}} = \frac{188 + 30}{188 + 5 + 10 + 30} = \frac{218}{233} = 93.56\%$$

3.2.3. Precision Value

a. Precision for short class

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{188}{188 + 5} = \frac{188}{193} = 97.41\%$$

b. Precision for very short classes

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{30}{30 + 10} = \frac{30}{40} = 75.00\ \%$$

3.2.4. Recall Value Calculation

a. Recall for short classes

$$Recall = \frac{TP}{TP + TN} = \frac{188}{188 + 10} = \frac{188}{198} = 94.95\ \%$$

b. Recall for the class is very short

$$Recall = \frac{TP}{TP + TN} = \frac{30}{30 + 5} = \frac{30}{35} = 85.71\ \%$$

3.2.5. Calculation of the F1-Score

a. F1-Score for the Short class

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0.9741 \times 0.9495}{0.9741 + 0.9495} = \frac{0.9250}{1.9236} = 2 \times 0.4809 = 0.9618$$

b. F1-Score for Very Short Class

$$F1 - Score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} = \frac{2 \times 0.7500 \times 0.8571}{0.7500 + 0.8571} = \frac{0.6428}{1.6071} = 2 \times 0.3999 = 0.7998$$

3.2.6 Error Analysis

The diminished F1-Score (79.98%) for the “Very Short” class is attributable to class imbalance, where instances of the “Short” class substantially outnumber those of the “Very Short” class (989 versus 179). This imbalance biases the Naive Bayes classifier toward the majority class during training, a well-documented limitation of probabilistic classification models in stunting prediction tasks [27].

The low standard deviation (±0.02) across five experimental iterations indicates that the model performance is stable; however, sensitivity toward the minority class remains limited. To address this

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

issue, future studies may explore data balancing strategies such as the Synthetic Minority Over-sampling Technique (SMOTE) or the application of class weighting to reduce classification bias, as suggested by prior Naive Bayes-based stunting classification studies [40].

3.3. Integrated Interpretation

The amalgamation of K-Means clustering with Naive Bayes classification yields a dual-layer analytical perspective:

1. K-Means successfully grouped toddlers based on height-for-age, identifying 989 short and 179 very short cases.
2. Naive Bayes subsequently classified these clusters with 93.56% ± 0.02 accuracy, highlighting TB/U as the most discriminative variable.

This hybrid analysis demonstrates the potential of machine learning integration in public health informatics, supporting targeted early detection at the Community Health Center level.

4. DISCUSSION

This section analyzes and interprets the test results of the K-Means and Naive Bayes algorithms and elaborates on their comparative advantages, limitations, and implications for stunting prediction systems at the community health center (Puskesmas) level.

4.1. I nterpretation of K-Means Results

A total of 1,168 toddler records were categorized into eight groups (K = 8) utilizing the K-Means technique. The Davies–Bouldin Index (DBI) value of 4.353 was the minimum among all K values (Table 4), signifying clusters with favorable compactness and separation, in agreement with the findings of Rohman et al. [31]

There was an observable cluster size imbalance, as Cluster 1 contained 1,062 items, while several clusters (Cluster 0, 4, 6, 7) contained fewer than 5 items. This imbalance suggests a predominant subgroup with homogeneous anthropometric characteristics and several small, heterogeneous outliers [20].

The dominant cluster (Cluster 1) comprising 910 short and 152 very short toddlers represents the largest risk group requiring focused nutritional and growth interventions. These patterns confirm the role of Height-for-Age (TB/U) as the strongest indicator of stunting, aligning with WHO and SSGI 2024 findings [7], [8].

Figure 2 illustrates these groupings visually, showing clear cluster boundaries and supporting the statistical results. The low DBI score demonstrates that the clustering process captured meaningful anthropometric distinctions among toddlers [31], [39].

In summary, the clustering process effectively identified the main stunting risk group and provided a foundation for classification-based prediction in subsequent analyses.

4.2. Interpretation of Naive Bayes Results

The Naive Bayes classifier attained an accuracy of 93.56% ± 0.02, accurately predicting 218 of 233 test samples. This outcome exhibits robust predictive dependability and aligns with the research conducted by Desty & Marpaung [2] and Dwinanto et al. [32]. Table 9 illustrates that the model attained a Precision of 97.41%, Recall of 94.95%, and F1-Score of 96.18% for the short class, but for the Very Short class, it earned a Precision of 75.00%, Recall of 85.71%, and F1-Score of 79.98%.

The lower F1-Score for the “Very Short” class indicates the influence of class imbalance, as the dataset contained 989 short and 179 very short toddlers. Such imbalance causes the Naive Bayes classifier to favor the majority class, a limitation commonly observed in probabilistic classification models [27]. while similar performance patterns have also been reported in stunting classification studies using Naive Bayes [28]

Table 9. Performance Evaluation Metrics

Evaluation Matrix	Short Class	Very Short Class
Accuracy	93.56%	
Precision	97.41%	75.00%
Recall	94.95 %	85.71%
F1-Score	96.18%	79.98%

Error analysis reveals that some very short toddlers were misclassified as short due to overlapping Z-scores between –3 SD and –2 SD. Future work could employ class re-weighting or SMOTE oversampling to enhance minority-class sensitivity [40].

The model's precision and recall balance reflects good stability and robustness compared to other local studies such as Kamil & Wibowo [24] (85.76%) and Titimeidara and Hadikurniawati [33] (88%). These results confirm that the hybrid K-Means + Naive Bayes approach outperforms single-model implementations.

4.2. Comparative Analysis with Recent Studies

To contextualize this study's findings, comparisons were made with at least six recent studies (2023 – 2025) focusing on stunting prediction and clustering.

As shown in Table 10, the proposed hybrid approach achieves superior performance in both cluster validity (DBI = 4.353) and classification accuracy (93.56 %).

Table 10. Comparative Accuracy of Recent Related Studies

Year	Authors	Algorithm	Accuracy	Key Findings
2024	Kamil & Wibowo [24]	Naive Bayes	85.76 %	Naive Bayes achieved reasonable classification accuracy using a single-model approach
2024	Widhari et al. [28]	Naive Bayes vs K-NN	71 %	Experimental results indicate that Naive Bayes showed lower accuracy than K-NN on the stunting dataset.
2025	Rohman et al. [31]	K-Means + PCA	Distortion Score = 7062.78	PCA-assisted K-Means improved cluster structure based on distortion analysis.
2024	Dwinanto et al. [32]	Naive Bayes vs SVM vs K-NN	85% (NB), 90% (SVM)	SVM achieved the highest accuracy among the compared classifiers.
2021	Titimeidara & Hadikurniawati [33]	Naive Bayes	88 %	Naive Bayes effectively classified stunting nutritional status.
2025	Maulana et al. [40]	Naive Bayes vs Decision Tree	78.8% (NB)	Naive Bayes outperformed Decision Tree in stunting classification.
2025	**This Study**	K-Means + Naive Bayes (Hybrid)	**93.56 %**	**Best hybrid accuracy + DBI = 4.353**

From Table 10, this study achieves higher predictive accuracy than all single model approaches and shows comparable results with hybrid systems [31], [40].

This demonstrates the strength of integrating unsupervised clustering before supervised classification, which improves class separation and reduces misclassification error.

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

This demonstrates the strength of integrating unsupervised clustering before supervised classification, which improves class separation and reduces misclassification error.

4.3. Research Implications

This project integrates K-Means and Naive Bayes to establish a novel, data-driven framework for public health decision support systems at the Community Health Center level.

This model utilizes hybrid machine-learning techniques to: a. Improve early detection of stunting risk through the integration of cluster mapping and predictive analytics; b. Be scalable to other Community Health Center by retraining with local data without necessitating algorithm modification; c. Facilitate the attainment of Sustainable Development Goals (SDG). 2.2: Eradicate All Forms of Malnutrition by linking innovations in Computer Science with public health goals.

This work enhances the application of hybrid AI models for adaptive stunting prediction systems, facilitating real-time categorization and visualization for local health officials. These technologies might be included into regional health dashboards, allowing the government to implement targeted treatments and maximize scarce resources.

5. CONCLUSION

This study's findings indicate that the K-Means method attained optimum clustering at k = 8, with a Davies–Bouldin Index (DBI) of 4.353, demonstrating robust cluster compactness and separation. The investigation revealed 1,168 stunted toddlers, including 989 classified as “Short” and 179 as “Very Short,” according to the Height-for-Age (TB/U) indicator, in accordance with Indonesian Ministry of Health Regulation No. 2 of 2020.

The Naive Bayes classifier exhibited a classification accuracy of 93.56%, with a precision of 97.41%, a recall of 94.95%, and an F1-score of 96.18% for the short class, whereas the Very Short class attained a precision of 75.00%, a recall of 85.71%, and an F1-score of 79.98%. The findings demonstrate that the hybrid K-Means and Naive Bayes methodology surpasses earlier single-model investigations, such as those by Kamil & Wibowo [24] (85.8%) and Titimeidara and Hadikurniawati [33] (88%), providing a more equitable prediction of precision and recall.

The amalgamation of K-Means with Naive Bayes constitutes an innovative hybrid framework in health informatics, facilitating systematic risk mapping and predictive classification for local health facilities (Puskesmas). The clustering outcomes (989 short, 179 very short) establish a definitive foundation for community-level intervention planning, whilst the categorization model facilitates automated decision-making.

This paper presents an applied strategy for creating open-source stunting prediction dashboards and community health monitoring toolkits from a Computer Science standpoint. These devices might interface with IoT-enabled anthropometric sensors to automatically record child height and weight data in real time. Future improvements may use Deep Learning (DL) architectures to identify nonlinear patterns in child growth trends and boost classification accuracy in the presence of data imbalance.

ACKNOWLEDGEMENT

The author expresses sincere gratitude to Amikom University Purwokerto, the institution that provided the opportunity, facilities, and support for this research. Special thanks are extended to the Kembaran 1 Community Health Center level staff for granting permission and providing research data, and to the Supervisor for the invaluable guidance, direction, and input. The author also thanks their family, colleagues, and all parties who provided encouragement and motivation so that this research could be completed successfully.

REFERENCES

[1] A. Ernawati, "Gambaran Penyebab Balita Stunting di Desa Lokus Stunting Kabupaten Pati," *Jurnal Litbang Media Inf. Penelitian, Pengemb. dan IPTEK*, vol. 16, no. 2, pp. 77–94, 2020, https://doi.org/10.33658/jl.v16i2.194.

[2] H. P. Desty and N. L. Marpaung, "Diagnosa Stunting Pada Balita Menggunakan Metode Naive Bayes Untuk Sistem Pakar," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 16, no. 2, pp. 107–123, 2024, https://doi.org/10.22303/csrid.16.2.2024.107-123.

[3] L. Agustin and D. Rahmawati, "Hubungan Pendapatan Keluarga dengan Kejadian Stunting," *Journal of Midwifery,* vol. 4, no. 1, 2021*,* https://doi.org/10.35473/ijm.v4i1.715.

[4] K. Suryani, M. T. Rini, B. D. Hardika, and N. K. Widiastari, "Analisis Faktor Penyebab Kejadian Stunting," *J. Keperawatan Florence Nightingale*, vol. 6, no. 1, pp. 8–12, 2023, https://doi.org/10.52774/jkfn.v6i1.112.

[5] K. Anwar et al., "Edukasi Makanan Pendamping Air Susu Ibu (MPASI) pada Ibu Kader Posyandu di Kecamatan Sawangan Depok Tahun 2024," *Jurnal SOLMA*, vol. 13, no. 3, pp. 2524–2531, 2024, https://doi.org/10.2236/solma.v13i3.16599.

[6] N. Fentiana, F. Tambunan, and D. Ginting, "Peran Pemantauan Pertumbuhan Dalam Upaya Pencegahan Stunting Anak 0-23 Bulan di Indonesia: Temuan Riskesdas 2013," *Jurnal Keperawatan Florence Nightingale*, vol. 2, no. 2, pp. 9–18, 2022, https://doi.org/10.58185/j-mestahat.v2i2.96.

[7] Kementerian Kesehatan Republik Indonesia, *Survei Status Gizi Indonesia (SSGI) 2024 dalam Angka*, Jakarta: Kemenkes RI, 2025. [Online]. Tersedia: https://www.badankebijakan.kemkes.go.id/en/survei-status-gizi-indonesia-ssgi-2024/.

[8] Puskesmas Kembaran I, *Profil Kesehatan Puskesmas Kembaran I Tahun 2024,* Dinas Kesehatan Kabupaten Banyumas, Banyumas, Indonesia, 2024.

[9] R. Hasanah, F. Aryani, and B. Effendi, "Pemberdayaan Masyarakat Dalam Pencegahan Stunting Pada Anak Balita," *J. Masy. Madani Indones.*, vol. 2, no. 1, pp. 1–6, 2023, https://doi.org/10.59025/js.v2i1.54.

[10] N. Miratika, R. Trisudarmo, and T. S. Syamfithriani, "Implementation of Naive Bayes Algorithm for Early Detection of Stunting Risk," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 2, pp. 356–363, 2025, https://doi.org/10.30871/jaic.v9i2.9144.

[11] P. B. Alamin and D. Aryani, "Perbandingan Algoritma K-Nearest Neighbor, Naive Bayes, dan Support Vector Machine untuk Klasifikasi Status Stunting Anak Balita di Puskesmas Sepatan," *STORAGE - J. Ilm. Tek. Ilmu Komput.*, vol. 4, no. 4, pp. 360–369, 2025, https://doi.org/10.55123/storage.v4i4.6591.

[12] E. Nurjannah, M. Nasution, and R. Muti'ah, "Data Mining Clustering Analysis of Child Growth and Development Using the K-Means Method," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 3, pp. 1909–1919, 2024, https://doi.org/10.33395/sinkron.v8i3.13817.

[13] S. Melyani, S. Z. Harahap and Irmayanti, "Prediction of Stunting in Toddlers Combining the Naive Bayes Method and the C4.5 Algorithm," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 8, no. 2, pp. 1160-1168, 2024, https://doi.org/10.33395/v8i2.13651.

[14] N. I. Sari and S. Harianis, "Analisis Faktor yang Mempengaruhi Kejadian Stunting pada Balita," *Maternal & Neonatal Health Journal*, vol. 3, no. 2, pp. 57–64, 2022, https://doi.org/10.37010/mnhj.v3i2.750.

[15] T. S. Priyadarshini and M. A. Hameed, "Collaboration of Clustering and Classification Techniques for Better Prediction of Severity of Heart Stroke Using Deep Learning," *Measurement: Sensors,* vol. 37, no. 2025, pp. 1–8, 2025, https://doi.org/10.1016/j.measen.2024.101405.

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

[16] W. Putri, D. Hastari, K. U. Faizah, S. Rohimah, and D. Safira, "Implementation of Naive Bayes Classifier for Classifying Alzheimer's Disease Using the K-Means Clustering Data Sharing Technique," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 1, no. 1, pp. 47–54, 2023, https://doi.org/10.57152/predatecs.v1i1.803.

[17] C. K. P. Rukma and Alamsyah, "Increase Accuracy of Naive Bayes Classifier Algorithm with K-Means Clustering for Prediction of Potential Blood Donors," *Journal of Advances in Information Systems and Technology*, vol. 4, no. 1, pp. 42–49, 2022, https://doi.org/10.15294/jaist.v4i1.59977.

[18] A. Mahiruna, N. Ngatimin, R. Destriana, E. H. Rachmawanto, H. Yuliansyah, and M. T. Hidayat, "Enhancing Clustering Accuracy Using K-Means with Seeds Optimization," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 5, pp. 2426–2433, 2025, https://doi.org/10.30871/jaic.v9i5.10458.

[19] A. A. Nababan, "Application of Naive Bayes Algorithm for Dominant Disease Classification in Coastal Environments," *Jurnal Info Sains: Informatika dan Sains*, vol. 14, no. 1, pp. 749–760, 2024, https://doi.org/10.54209/infosains.v14i01.

[20] R. N. Pratistha and B. Kristianto, "Implementasi Algoritma K-Means dalam Klasterisasi Kasus Stunting pada Balita di Desa Randudongkal," *Jurnal Indonesia: Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 5, no. 2, pp. 1193–1205, 2024, https://doi.org/10.35870/jimik.v5i2.634.

[21] 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 Informatika dan Komputer)*, vol. 6, no. 2, p. 223-230, 2022, https://doi.org/10.26798/jiko.v6i2.581.

[22] N. K. S. Julyantari, I. K. Budiarta, and N. M. D. K. Putri, "Implementasi K-Means untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih)," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 1, no. 2, pp. 92–101, 2021, https://doi.org/10.25008/janitra.v1i2.134.

[23] Arlita, W. Prihartono, and C. L. Rohmat, "Optimalisasi Upaya Pencegahan Stunting Pada Balita Di Kota Cirebon Menggunakan Metode Naive Bayes," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 2, pp. 9-15, 2025, https://dx.doi.org/10.23960/jitet.v13i2.5868.

[24] M. I. Kamil and A. P. Wibowo, "Implementation of the Naive Bayes Classifier Algorithm for Classifying Toddler Nutritional Status," *Journal of Applied Informatics and Computing (JAIC)*, vol. 8, no. 2, pp. 421–427, 2024, https://doi.org/10.30871/jaic.v8i2.8669.

[25] P. Meriyana, A. R. Pratama, E. Nurlaelasari, and A. R. Juwita, "Penerapan Algoritma KNN dan Naive Bayes untuk Klasifikasi Stunting pada Balita di Desa Pasirjengkol," *J. PETISI*, vol. 06, no. 02, pp. 72–84, 2025, https://doi.org/10.36232/jurnalpetisi.v6i2.2020.

[26] A. Ridwan and T. N. Sari, "The Comparison of Accuracy Between Naive Bayes Classifier and C4.5 Algorithm in Classifying Toddler Nutrition Status Based on Anthropometry Index," *Journal of Physics: Conference Series, Calgary: Institute of Physics,* vol. 1764, no. 1, Art. no. 012047, pp. 1-7, 2021, https://doi.org/10.1088/1742-6596/1764/1/012047.

[27] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, "Prediksi Risiko Stunting pada Keluarga Menggunakan Naive Bayes Classifier dan Chi-Square," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 172–180, 2024, https://doi.org/10.57152/malcom.v4i1.1074.

[28] W. Widhari, A. Triayudi, and R. T. K. Sari, "Implementation of Naive Bayes and K-NN Algorithms in Diagnosing Stunting in Children," *SAGA Journal of Technology and Information Systems*, vol. 2, no. 1, pp. 164–174, 2024, https://doi.org/10.58905/saga.v2i1.242.

[29] O. N. Chilyabanyama et al., "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, pp. 1–11, 2022, https://doi.org/10.3390/children9071082.

Vol. 7, No. 2, April 2026, Page. 1741-1757
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.2.5420

[30] H. Saleh, M. Faisal, and R. I. Musa, "Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor," *Simtek : Jurnal Sistem Informasi dan Teknik Komputer*, vol. 4, no. 2, pp. 120–126, 2019, https://doi.org/10.51876/simtek.v4i2.60.

[31] M. H. M. Rohman et al., "Clustering Analysis of Stunting Risk Factors Using K-Means and Principal Component Analysis: A Case Study in Indonesian Regency," *Sinkron*, vol. 9, no. 1, pp. 65–77, 2025, https://doi.org/10.33395/sinkron.v9i1.14311.

[32] R. W. Dwinanto, A. S. Sandi A, and R. Ardianto, "Klasifikasi Berisiko Stunting pada Balita: Perbandingan K-Nearest Neighbor, Naïve Bayes, Support Vector Machine," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 8, no. 2, pp. 264–273, 2024, https://doi.org/10.46880/jmika.Vol8No2.pp264-273.

[33] M. Y. Titimeidata and W. Hadikurniawati, "Implementasi Metode Naive Bayes Classifier untuk Klasifikasi Status Gizi Stunting Pada Balita," *Jurnal Ilmiah Informatika (JIF)*, vol. 9, no. 1, pp. 54–59, 2021, https://doi.org/10.33884/jif.v9i01.3741.

[34] I. Choliq, D. Nasrullah, and Mundakir, "Pencegahan Stunting di Medokan Semampir Surabaya Melalui Modifikasi Makanan Pada Anak," *Humanism Jurnal Pengabdian Masyarakat*, vol. 1, no. 1, pp. 31–40, 2020, https://doi.org/10.30651/hm.v1i1.4544.

[35] D. A. Fitria, "Peran Pemerintah Dalam Penanganan Stunting Pada Balita Di Kelurahan Alai Kecamatan Ungar," *Pubmedia Social Sciences and Humanities*, vol. 1, no. 3, pp. 1-6, 2023, https://doi.org/10.47134/pssh.v1i3.160.

[36] RapidMiner GmbH, "*RapidMiner Studio Documentation*" Accessed: Nov. 10, 2025. [Online]. Available: https://docs.rapidminer.com/.

[37] I. Iskandar and D. Jollyta, "Penerapan Algoritma K-Means Untuk Clusterisasi Kasus Stunting Di Provinsi Riau," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 6, no. 1, pp. 18–22, 2024, https://doi.org/10.35145/jmapteksi.v6i1.4130.

[38] H. Hafsah, N. Azmi, H. Hazriani, and Y. Yuyun, "Klasifikasi Status Gizi Balita Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Prosiding Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, ke-7, Gedung Universitas Mega Rizky, Makassar, Okt. 2023, pp. 1-6.

[39] A. Ersawahyuni, S. Martha, and H. Perdana, "Pengelompokan Indeks Pembangunan Manusia di Indonesia Menggunakan Metode K-Medoids Dengan Evaluasi Davies Bouldin Index," *BIMASTER : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 14, no. 2, pp. 207–216, 2025, https://doi.org/10.26418/bbimst.v14i2.92331.

[40] A. Maulana et al., "Classification of Stunting in Toddlers using Naive Bayes Method and Decision Tree," *Indonesian Journal of Modern Science and Technology (IJMST)*, vol. 1, no. 1, pp. 28–33, 2025, https://doi.org/10.64021/ijmst.1.1.28-33.2025.

[41] F. Irhamni, F. Damayanti, B. K. K, and M. A, "Optimalisasi Pengelompokan Kecamatan Berdasarkan Indikator Pendidikan Menggunakan Metode Clustering dan Davies Bouldin Index," in *Seminar Nasional Sains dan Teknologi 2014*, Jakarta: Fakultas Teknik Universitas Muhammadiyah Jakarta, Nov. 2014, pp. 1–5.

[42] Kementerian Kesehatan Republik Indonesia, "*Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 tentang Standar Antropometri Anak*," Jakarta, 2020. [Online]. Tersedia: https://stunting.go.id/standar-antropometri-anak-peraturan-menteri-kesehatan-nomor-2-tahun-2020/.