SAINSTECH
INNOVATION JOURNAL

SIJ, Volume 9, Nomor 1, May 2026, Hal. 643-648

Classification of Child Stunting Status Using the K-Nearest Neighbor (KNN) Algorithm Based on Toddler Growth Data in East Lombok Regency, West Nusa Tenggara

Asno Azzawagama Firdaus[1)*], Arif Himawan [2)], Junaedi [3)] , Anggun Sindiana [4)], Istianah [5)], Baiq Selviana Pertiwi [6)] , Baiq Wangi Narsih [7)], Kartika Yundia [8)] , Ahmad Azhari [9)]

[1), 4), 5), 6), 7), 8)] Department of Computer Science, Universitas Qamarul Huda Badaruddin Bagu, Indonesia
[2)] Pusat Riset Masyarakat dan Budaya, Badan Riset dan Inovasi Nasional, Indonesia
[3)] Department of Digital Business, Akademi Bisnis Lombok, Indonesia
[9)] Department of Informatics, Universitas Ahmad Dahlan, Indonesia

**Corresponding Author: asnofirdaus@gmail.com, Tel: +6285338237826, Fax: -*

Received ***01 June 2026*****, Revise** ***15 July 2026*****, Accepted** ***30 July 2026*****, published online on** ***31 July 2026***

Abstract: *Stunting is a major nutritional issue in Indonesia that significantly impacts children's physical growth, cognitive development, and future quality of life. Childhood stunting can be identified using nutritional status indicators—such as weight-for-age (W/A), height-for-age (H/A), and weight-for-height (W/H)—specifically by observing a Z-score below -2 for the H/A indicator. The growing volume of anthropometric data on children under five necessitates a rapid and objective method for identifying stunting status. This study aims to classify children's stunting status using the K-Nearest Neighbor (KNN) algorithm based on growth data from children under five in East Lombok Regency, West Nusa Tenggara. The dataset comprises records for 3,416 children, including information on gender, age, weight-for-age (W/A), weight-for-height (W/H), and height-for-age (H/A). Data preprocessing involved removing duplicates, handling missing values, transforming categorical data, and applying Min-Max normalization. The data was split into 70% training data and 30% testing data, with the KNN algorithm applied using k = 5. Model evaluation was conducted using a confusion matrix. The results demonstrate that the KNN model achieved an accuracy rate of 83.15%, indicating a strong capability to classify stunting status based on the children's anthropometric characteristics. These findings confirm that the KNN algorithm can serve as a tool for healthcare professionals to identify stunting more rapidly, objectively, and efficiently, thereby supporting early prevention and management efforts.*

Keywords: *Stunting, K-Nearest Neighbor, Classification, Anthropometry, Under-fives*

1. INTRODUCTION

This study on stunting in East Lombok Regency is particularly relevant given the importance of integrating local wisdom and regional management to support the well-being of communities in West Nusa Tenggara, as exemplified by the traditional values of the awig-awig, which remain deeply rooted in North Lombok [1]. Stunting is a chronic nutritional issue that remains a concern in Indonesia due to its impact on physical growth, cognitive development, and the future quality of human resources. This condition is characterized by a child's height falling below the standard for their age because of prolonged malnutrition. The effects of stunting are not limited to childhood; they can also influence productivity and learning ability, as well as increase the risk of various diseases in adulthood. Consequently, efforts to prevent and address stunting have become a priority in Indonesia's health development.

Indonesia can leverage its demographic bonus to cultivate high-quality human resources. However, the high prevalence of stunting remains a challenge, driven by various factors such as socioeconomic conditions, parental education levels, parenting practices, nutritional intake, environmental sanitation, and access to healthcare services. These factors are interconnected; therefore, addressing stunting requires an integrated approach involving multiple sectors [2].

Monitoring the growth of children under five is a crucial step in the early detection of growth disorders. Integrated Health Posts (Posyandu) and Community Health Centers (Puskesmas) serve as healthcare facilities that record anthropometric data—such as age, weight, and height—for these children. However, the process of recording and managing this data at some Posyandu locations is still performed manually; this is time-consuming, prone to recording errors, and complicates the task of compiling child development reports for the staff.

Advancements in information technology have driven the use of data mining and machine learning in the healthcare sector, including in the classification of nutritional status and stunting among children under five. Through computerized data processing, the identification of children at risk of stunting can be performed more rapidly, objectively, and accurately, thereby supporting decision-making regarding early interventions [3].

The use of data mining techniques such as K-Nearest Neighbor, Random Forest, Feature Selection, Decision Tree, Multiple Linear Regression [4][5][6][7][8] for classifying nutritional status is in line with advances in computational methods in Indonesia, including the development of the Weighted Collective Influence Graph (WCIG) method, which is used to identify influential actors in social media data [9]. One classification algorithm widely applied in health research is K-Nearest Neighbors (KNN). This algorithm operates by determining the class of a data point based on its proximity to a set of training data using Euclidean distance calculations. KNN involves a simple process, is easy to implement, and can deliver good classification performance in various studies concerning the nutritional status and stunting of children under five [10].

The study titled "Classification of Stunting Status in Toddlers in Slangit Village Using the K-Nearest Neighbor Method" demonstrates that the KNN algorithm can classify stunting status based on toddler anthropometric data with an accuracy rate of 98.89%. These results indicate that KNN can be utilized as an effective method to assist in the process of identifying stunting status in toddlers [11].

Furthermore, the study "Classification of Stunting Status in Toddlers Using K-Nearest Neighbor with Backward Elimination Feature Selection" demonstrates that the application of feature selection improves the performance of the KNN algorithm, achieving an accuracy rate of 92.20%. These results indicate that the selection of appropriate attributes can enhance the quality of classification outcomes [12].

Based on various previous studies, the K-Nearest Neighbor algorithm demonstrates a strong capability for classifying the nutritional status and stunting of toddlers. However, each study yields a different level of accuracy, as the results are influenced by the data volume, dataset characteristics, attributes employed, and the applied K-parameter value [13].

Based on the foregoing, this study aims to apply the K-Nearest Neighbor (KNN) algorithm to classify the stunting status of children under five using anthropometric data, thereby assisting healthcare workers in conducting early identification more quickly, objectively, and accurately.

2. METHODOLOGY

This study employs a data mining approach using the K-Nearest Neighbor (KNN) algorithm to classify stunting status in toddlers. The KNN algorithm was selected due to its ability to classify new data based on proximity to training data—specifically, the nearest neighbors with known class labels. The classification process utilizes anthropometric data from toddlers, including the Height-for-Age (HAZ) indicator, which serves as the basis for determining stunting status [14].

The data used is secondary data consisting of anthropometric measurements of children under five, obtained from a dataset on nutritional status and stunting. The dataset comprises records for 3,416 children, containing information on gender, date of birth, date of measurement, weight-for-age (W/A) status, weight-for-height (W/H) status, and height-for-age (H/A) status [9].

The variables used in this study consist of input variables and output variables.

The input variables include:

- Age (Toddler's age in days)
- BB/U (Weight for Age)
- BB/TB (Weight according to Height)

Output Variable:

Height-for-Age (HAZ) is used as the target class to determine the stunting status of children under five.

Research Stages:

The research was conducted in several stages, as follows:

2.1. Data Collection

Data on toddlers was collected from an existing dataset and used as the source of data for the research (doi: 10.17632/z4mrz37dm3.2) [15].

2.2. Data Preprocessing

The preprocessing stage is carried out to improve data quality prior to the classification process. The steps involved include:

a. Removing duplicate data.
b. Removing data with missing values.
c. Calculating the age of toddlers based on the difference between the measurement date and the date of birth.
d. Removing data with invalid ages.
e. Converting categorical data into numerical data using encoding

techniques.

f. Performing data normalization using the Min-Max Normalization.

2.3. Data Partitioning

The processed data is then divided into two parts, namely:

a. 70% training data

b. 30% testing data

Data partitioning was performed randomly using the random sampling method.

2.4. Classification Using K-Nearest Neighbor (KNN)

The classification process is performed using the K-Nearest Neighbor algorithm with a parameter value of k = 5. KNN operates by identifying the five nearest neighbors of the test data based on their distance to the training data, and subsequently determining the class based on the majority class among those nearest neighbors [16].

2.5. Model Evaluation

The classification results were evaluated using a confusion matrix and accuracy to determine the model's performance level.

3. RESULTS AND DISCUSSION

In addition to parenting practices, access to sanitation and adequate nutrition—which are priorities in the prevention of stunting—are also influenced by environmental stability. Climate change and ecosystem degradation are known to threaten the availability of clean water and food production, which indirectly impact children's health [17]. The study utilizes four input variables—Gender, Age, Weight-for-Age (BB/U), and Weight-for-Height (BB/TB)—while the output variable is Height-for-Age (TB/U), serving as an indicator of stunting status in toddlers. Training data is used to build the classification model, whereas testing data is used to evaluate the model's ability to classify new data. Preprocessing of 3,416 raw data points yielded 3,404 valid records ready for classification using the KNN algorithm (k=5). Model testing was conducted by splitting the dataset into 70% for training and 30% for testing. The test results are presented in the Confusion Matrix shown in the table below:

Tabel 1. Confusion Matrix Hasil Klasifikasi KNN

Prediction/ Actual	1	2	3	4	5	6
1	3	0	0	0	0	0
2	0	823	0	101	1	23
3	0	0	0	0	0	0
4	0	26	0	13	0	6
5	0	0	0	0	0	0
6	0	7	0	8	0	10

Based on the table above, most data points have been classified correctly. This indicates that the KNN model is capable of effectively recognizing data patterns based on the similarity of characteristics among data points. The model was evaluated using an accuracy score derived from comparing the predicted results with the actual values in the test dataset. The test results show that the K-Nearest Neighbor algorithm achieved an accuracy rate of 83.15%. This figure means that approximately 83 out of every 100 records regarding toddlers were correctly classified by the model. This accuracy rate indicates that the KNN algorithm performs well in classifying stunting status based on the variables of Gender, Age, Weight-for-Age (WFA), and Weight-for-Height (WFH).

To provide a more comprehensive overview of the data underlying this classification, the distribution of nutritional status among children under five used in this study is presented below:

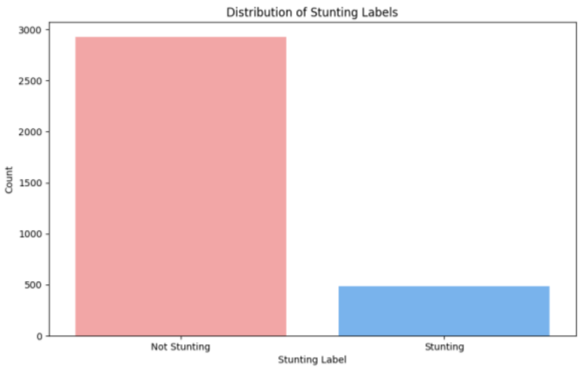

Figure 1. Number of Toddlers with Stunting Based on Height-for-Age (HAZ) Category [15]

Analysis of the pre-processed dataset reveals the distribution of nutritional status among children under five years of age, as shown in Figure 1. Of the 3,404 valid data points, the majority of children fell into the normal nutritional status category, while 482 children were identified as stunted (short and very short). This visualization provides an objective basis for the KNN model to classify stunting status, with the balanced data distribution contributing to the model's accuracy rate of 83.15%.

4. CONCLUSION

The KNN algorithm is effective for classifying stunting status in toddlers using the variables of gender, age, weight-for-age (WFA), and weight-for-height (WFH). With proper preprocessing and a parameter of k=5, the model achieved an accuracy of 83.15%. This confirms that the KNN method can serve as an objective tool for healthcare workers to facilitate more efficient early detection and management of stunting. The results of this classification are expected to serve as a tool for local governments in formulating data-driven health policies. This approach is like recommendations for natural resource management agencies to integrate satellite oceanographic data into conservation planning and community economic development [18].

REFERENCES

[1] A. Hilmawan, Marsoedi, E. Susilo, and Rudianto, "LOCAL CUSTOM VALUES AS NORTH LOMBOK COASTAL AREA MANAGEMENT," *Russ. J. Agric. Socioecon. Sci.*, vol. 72, no. 12, pp. 307–311, Dec. 2017, doi: 10.18551/rjoas.2017-12.43.

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

[3] V. Yoga Pudya Ardhana, A. Fatoni Dwi Putra, D. Teko Kumoro, M. Dermawan Mulyodiputro, and U. Qamarul Huda Badaruddin Bagu, "Capacity Building of Posyandu Cadres through Training on an Early Stunting Detection Application in Lombok."

[4] S. Lonang, A. Fatoni, D. Putra, F. Syuhada, A. A. Firdaus, and A. Masitha, "Hybrid Feature Selection for Effective Heart Disease Detection: A Multi-Algorithm Machine Learning Approach," *Scientific Journal of Informatics*, vol. 13, no. 1, 2026, doi: 10.15294/sji.v13i1.38815.

[5] Mindi Richia Putri, Asmaul Husna RS, Ahmad Fatoni Dwi Putra, Muhammad Ari Rifqi, and Syaidatussalihah, "Comparison of Multiple Linear Regression, Random Forest, and Decision Tree Methods for House Price Prediction in West Nusa Tenggara," *Journal of Innovation Information Technology and Application (JINITA)*, vol. 8, no. 1, pp. 122–130, Jun. 2026, doi: 10.35970/jinita.v8i1.3010.

[6] A. F. D. Putra, M. N. Azmi, H. Wijayanto, S. Utama, and I. G. P. W. Wedashwara Wirawan, "Optimizing Rain Prediction Model Using

Random Forest and Grid Search Cross-Validation for Agriculture Sector," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 3, pp. 519–530, Jul. 2024, doi: 10.30812/matrik.v23i3.3891.

[7] A. A. Firdaus, R. A. Faresta, and M. Yunus, "The Role of Sentiment Analysis in Election Predictions Compared to Electability Surveys," *Indonesian Journal of Modern Science and Technology (IJMST)*, vol. 1, no. 1, pp. 1–8, Jan. 2025, doi: https://doi.org/10.64021/ijmst.1.1.1-8.2025.

[8] R. Hasanah, S. Sani SR, M. Munzir, A. A. Firdaus, C. Sulton, and M. Yunus, "Sentiment Analysis of User Reviews of TikTok App on Google Play Store Using Naïve Bayes Algorithm," *Indonesian Journal of Modern Science and Technology (IJMST)*, vol. 1, no. 2, pp. 58–64, May 2025, doi: 10.64021/ijmst.1.2.58-64.2025.

[9] S. E. Youwikijaya, R. A. Kinseng, T. Sumarti, A. Hilmawan, and M. Jamilah, "Livelihood vulnerability of small fishers' households on small islands: A case in Pari Island, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1618, no. 1, p. 012003, May 2026, doi: 10.1088/1755-1315/1618/1/012003.

[10] Q. Xu, "Application of an Intelligent English Text Classification Model with Improved KNN Algorithm in the Context of Big Data in Libraries," *Systems and Soft Computing*, p. 200186, Jan. 2025, doi: 10.1016/j.sasc.2025.200186.

[11] T. Prasetiya, I. Ali, C. L. Rohmat, and O. Nurdiawan, "Klasifikasi Status Stunting Balita Di Desa Slangit Menggunakan Metode K-Nearest Neighbor," *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, vol. 5, no. 1, p. 93, Dec. 2020, doi: 10.51211/itbi.v5i1.1431.

[12] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 49, Jan. 2022, doi: 10.30865/mib.v6i1.3312.

[13] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 49, Jan. 2022, doi: 10.30865/mib.v6i1.3312.

[14] 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, Jan. 2025, doi: https://doi.org/10.64021/ijmst.1.1.28-33.2025.

[15] S. E. Youwikijaya, R. A. Kinseng, T. Sumarti, and A. Hilmawan, "Identifikasi Penguasaan dan Akses Modal Rumah Tangga Nelayan Kecil di Pulau-Pulau Kecil (Kasus Pulau Pari)," *Buletin Ilmiah Marina Sosial Ekonomi Kelautan dan Perikanan*, vol. 9, no. 2, p. 87, Dec. 2023, doi: 10.15578/marina.v9i2.12003.

[16] T. P. HERDIANTI *et al.*, "HABITAT ASSESSMENT OF WHALE SHARKS (RHINCODON TYPUS) IN SALEH BAY, INDONESIA: LINKING CHLOROPHYLL-A AND SEA SURFACE TEMPERATURE USING AQUA MODIS DATA," *Geographia Technica*, vol. 20, no. 2/2025, pp. 52–64, Apr. 2025, doi: 10.21163/GT_2025.202.04.