Stunting Analysis of Toddlers in Kota Baru, West Bekasi Using K-Nearest Neighbor and Naive Bayes

1st Latifa Firdausi
School of Computing
Telkom University
Bandung, Indonesia
latifafirdausi@student.telkomuniversity.ac.id

2nd Putu Harry Gunawan
School of Computing
Telkom University
Bandung, Indonesia
phgunawan@telkomuniversity.ac.id

***Abstract*—Stunting is a condition caused by long-term malnutrition in children, leading to impaired cognitive abilities, speech difficulties, and learning challenges. In Indonesia, stunting is a significant concern, with the country ranking 27th out of 154 in prevalence according to UNICEF and WHO data. Early detection is vital for timely intervention. This study aims to compare the performance of the K-Nearest Neighbors (KNN) and Naive Bayes models in classifying stunting status in children. The analysis was conducted on a dataset collected from the Kota Baru Health Center in Bekasi Regency, which includes various features such as age, weight, height, gender, birth weight, and height-for-age ratio. Exploratory data analysis revealed a significant prevalence of stunting among children in the dataset, highlighting the urgent need for effective classification models. The analysis found KNN to outperform Naive Bayes across evaluation metrics, particularly on a balanced dataset with a 60:40 split, where KNN achieved an accuracy of 83.33% and an F1 score of 83.37%, compared to Naive Bayes' accuracy of 73.08% and an F1 score of 70.98%. The results highlight KNN's superior ability to handle class imbalance and its consistent performance across various dataset splits. Based on these findings, KNN is recommended for stunting classification, particularly in situations involving class imbalance. Future research should explore optimization techniques such as feature selection and data augmentation to further improve model performance.**

***Index Terms*—Stunting, Machine Learning, K-Nearest Neighbor, Naive Bayes**

I. INTRODUCTION

Stunting is a condition that marks the failure of a child's growth due to long-term malnutrition, leading to a decrease in intelligence levels, speech disorders, and difficulty in grasping lessons in the usual way. This has become a significant issue in Indonesia. According to data collected by UNICEF and WHO, Indonesia ranks 27th out of 154 countries, and in terms of stunting prevalence in Asia, Indonesia ranks 5th. The factors causing stunting hinder the growth and development of children. During the COVID-19 pandemic, there was an economic downturn and an increase in stunting cases. [1].

The Ministry of Health announced a decrease in stunting rates in Indonesia from 24.4% in 2021 to 21.6% in 2022, according to sehatnegeriku.kemkes.go.id. President Joko Widodo stated that stunting is not only an issue of height. This problem is also linked to mental retardation, children's learning abilities, and long-term health risks. Although the rates are expected to decrease significantly, stunting remains a serious issue in Indonesia because the prevalence rate is always above 20%, while the government's target is 14%. Additionally, according to data from the official stunting.bekasikota.go.id website, there has been a significant decline in the incidence of stunting in Bekasi City from 2019 to 2022. In 2022, the number of children recorded with stunting reached 4,575, or about 3.4%, a drastic decrease compared to the previous year's figure of 7.9%. This number is also far below the national target of 18.4%. The reduction in stunting rates over the past year reached 5.5%, showing effective efforts in addressing stunting issues in Bekasi City.

The analysis and system developed in this research can help address the problem of stunting by utilizing Machine Learning (ML) methods, including K-Nearest Neighbors and Naive Bayes, to improve prediction accuracy. Machine Learning algorithms are particularly well-suited for stunting prediction as they can efficiently handle large datasets and identify complex patterns and risk factors that may not be apparent through traditional methods. The implementation of ML has proven to significantly enhance prediction precision, contributing positively to the prevention and management of health issues like stunting. Previous studies have shown the effectiveness of ML in healthcare fields, supporting the application of these methods to predict stunting based on various risk factors [1]. This study aims to evaluate and apply different ML methods to determine the most optimal model for predicting stunting in children.

A collaborative approach involving parents, families, healthcare workers, and public health centers is essential to prevent stunting and ensure proper child growth [2]. To address stunting, the use of prediction and classification methods such as Naive Bayes and K-Nearest Neighbor has become a topic in medical research. In this study, classification methods are used in data processing, and this method is commonly employed in data mining to categorize new data records into one of several predefined classes, with the Naive Bayes classification algorithm being one of them [3]. In addition to Naive Bayes, this research also uses prediction through the K-Nearest Neighbor method, which involves variables such as height and weight, to evaluate stunting status. The formulation of the Euclidean distance calculation shows favorable results [4].

In a previous study conducted by Setiawan and Triayudi in 2022 on the Classification of Nutritional Status of Toddlers Using Naïve Bayes and K-Nearest Neighbor based on the web, the comparison results of accuracy, recall, and precision for the Naïve Bayes Classifier were 80.60%, 80.60%, and 79.66%, respectively, while for K-Nearest Neighbor, the accuracy, recall, and precision were 91.79%, 91.79%, and 91.17%, respectively [4]. Then, in the study conducted by Prasetiya et al. in 2020 to classify the stunting status of toddlers in Suragit Village using the K-Nearest Neighbor method, an accuracy of 98.89% was achieved when the testing data consisted of 300 toddlers [4]. Meanwhile, the previous study by Nugraha et al. in their journal using the Naïve Bayes Classifier algorithm resulted in a classification accuracy of 80% [5]. This study aims to determine the best classification model by comparing the performance of two machine learning algorithms, Naïve Bayes and K-Nearest Neighbor, to predict cases of child stunting. The results of this study are expected to provide a better understanding of how classification algorithms work and enable the development of a better stunting diagnosis model [6].

This research uses a smaller dataset and addresses class imbalance with SMOTE, normalization, and hyperparameter tuning. In contrast, Widhari et al. (2023) used a larger dataset but overlooked class imbalance. This study improves upon that by focusing on the quality of imbalanced data, offering more relevant results for the Bekasi child population.

In addition, this study also aims to identify the prevalence of stunting in the Kota Baru Health Center (*Pusat Kesehatan, Puskesmas*) area, West Bekasi District, and highlight the important role of the Health Center in the efforts to prevent and address stunting in children under five years of age in the region. By utilizing medical record documentation, the Kota Baru Health Center is expected to gain access to valuable information regarding the health status of toddlers, including data on stunting incidents.

II. METHODOLOGY

A. Methodology Flow

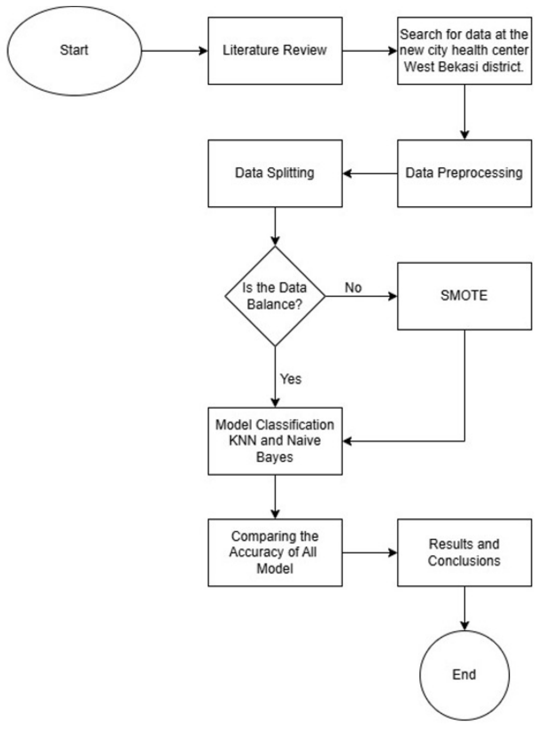

Fig. 1. Research Flowchart.

This flowchart, as seen in Fig. 1, The analysis of stunting data in children at Kota Baru Health Center follows a systematic process. First, a literature review is conducted to understand relevant theories. Next, data is collected on stunting and preprocessed, including feature selection (e.g., age, weight) and encoding categorical data. The data is then split into training and test sets, with a balance check. If imbalanced, the SMOTE technique is applied. The KNN and Naïve Bayes algorithms are used for classification, and their performance is compared through model accuracy evaluation. Finally, the results and conclusions are compiled, providing insights into algorithm accuracy and stunting conditions.

B. Dataset

This study utilizes data on child stunting collected directly through primary research at the Kota Baru Health Center, Bekasi Barat District, consisting of 193 data entries. Each entry contains information regarding the child's age, weight, height, TB/U, gender, birth weight, and birth height. The TB/U is classified into three categories based on the child's height-for-age: Short, Very Short, or Normal. These classifications reflect the extent of stunting based on the child's height relative to age standards. It is important to note that the dataset used in this study is not open source, as it was collected directly from field research and remains confidential. Table I presents examples of the stunting data used in this study.

C. Stunting

Stunting in children is a serious condition with significant health, economic, and social consequences [7]. In the future, children who experience stunting will struggle to achieve optimal physical and cognitive development for their age [8]. Determining the stunting status in children is based on the height-for-age index. The Z-score compares a child's measurement to a reference standard, indicating their nutritional status in terms of standard deviations [8]. The Z score formula is as follows [9]:

$$Z = \frac{H_{\text{children}} - \mu}{\sigma}$$

Explanation:

- Z = *z-score* for stunting.
- H_{children} = the measured height of the child.
- μ = the average height for the child's age and gender.
- σ = the standard deviation of height for the child's age and gender.

Stunting occurs when an individual's height falls below 2 standard deviations (Z-score $< -2\,\mathrm{SD}$) from the median height of a globally recognized population standard, according to the World Health Organization (WHO) (2006) [10]. Measuring the height of toddlers can be used to determine if a child is experiencing stunting. If measurements indicate signs of growth failure, immediate preventive actions can be taken [11].

D. K-Nearest Neighbor

In K-Nearest Neighbor, the value of k determines the number of neighbors for classification. An odd number is typically chosen, and the value of k increases with the training data size to minimize errors. [12]. The formula used in the implementation of the K-Nearest Neighbor algorithm is as follows [4]:

$$d_i = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Explanation:

- d_i: Euclidean distance between the stunting value of the individual i and the reference stunting value.
- x_i: Stunting value for individual i.
- y_i: Reference stunting value.

TABLE I
CHILD'S GROWTH TABLE

No.	Age	Weight	Height	TB/U	Gender	Birth Weight	Birth Height
1	33	11	84	Short	M	3	48
2	37	10	87	Short	M	3.1	50
3	10	6.4	67	Short	F	3	48
4	1	3.3	49	Very Short	F	2.95	47
5	50	13.7	89	Very Short	F	3	49
...	...	...	...	...	...	...	...
193	57	14	99	Normal	M	3.2	50

- $\sum$: The summation symbol from $i = 1$ to n, where n is the number of data points.

Below are the steps involved in implementing the K-Nearest Neighbors (KNN) algorithm [13]:

1) Determine the value of K, the number of nearest neighbors to consider.
2) For each test data:
   a) Calculate the distance between the test data and the training data.
   b) Sort the training data based on the closest distance to the test data.
   c) Take the K nearest neighbors from the training data.
   d) Determine the majority class among the K nearest neighbors.
3) Select the majority class as the prediction for the test data.

E. Naive Bayes

The Naive Bayes algorithm is a method that uses probability and statistics to estimate the likelihood of future events based on past experiences [14]. The main feature of the Naive Bayes Classifier is its strong (naive) assumption of independence between each condition or event [15]. The Naive Bayes Classifier assumes feature independence and uses Bayes' theorem to calculate class probabilities [14]:

$$P(A|B) = \frac{P(A)P(B)}{P(B|A)} \tag{2}$$

Explanation:

- $P(A|B)$: Probability that sample A occurs given that class B is known.
- $P(B|A)$: Probability that class B occurs given that sample A is known.
- $P(A)$: Prior probability of sample A.
- $P(B)$: Prior probability of class B, acting as a normalizing factor.

Below are the steps involved in implementing the Naive Bayes algorithm for classification tasks [16]:

1) For each class B:
   a) Calculate the prior probability $P(B)$ for each class based on the training data.
   b) For each feature A in the training data, calculate the likelihood $P(A|B)$ for each class B.
2) For each test data $A_1, A_2, \ldots, A_n$:
   a) Calculate the posterior probability for each class B using Bayes' Theorem:
3) Normalize the posterior probabilities if needed.
4) Select the class B with the highest posterior probability as the prediction for the test data.

F. Evaluation Model Matrix

Evaluation Matrix measures performance through accuracy, precision, recall, and f1-score. Accuracy calculates the correctness of a classification model, as shown in equation 3 [17].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Recall is used to calculate the model's performance in identifying relevant information. The equation for calculating recall is shown in equation 4 [17].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

Precision is used to calculate the ratio of relevant documents to the total number of documents. The equation for calculating precision is shown in equation 5 [17].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

F1-Score is the harmonic mean of Precision and Recall. The F1-score formula can be expressed using equation 6 [17].

$$F_1\text{-Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

G. Synthetic Minority Oversampling Technique (SMOTE)

SMOTE is a technique used to address class imbalances by generating synthetic samples for the minority class. It is applied to the training data to balance the class distribution, improving the model's ability to predict accurately, especially for the minority class, and enhancing overall performance. [18].

H. Tuning Hyperparameter

Hyperparameter tuning is performed to find the best settings for each method. In this study, two techniques are used: grid search, which exhaustively checks all possible combinations of predefined hyperparameters. This approach often provides near-optimal results, but with less time and resource consumption compared to grid search. The main goal of this tuning process is to improve the accuracy, precision, recall, and F1-score of the model, so that the resulting model can be more effective in detecting children at risk of stunting. [19].

III. RESULTS AND DISCUSSION

A. Exploratory Data Analysis (EDA)

Exploratory data plays a crucial role in uncovering patterns, trends, and relationships within the dataset. In this exploration, several features were selected, such as age, weight, height, height-for-age (TB/U), birth weight and birth height.

Table II and Fig 2 present the descriptive analysis of child growth. The average age is 32.98 years with a standard

TABLE II
EXPLORATORY DATA ANALYSIS

Variable	Age	Weight	Height	Birth Weight	Birth Height
Mean	32.98	12.63	84.98	2.97	48.16
Range	57	15	57	1.3	23
Stdev	13.43	1.96	10.26	2.76	0.29

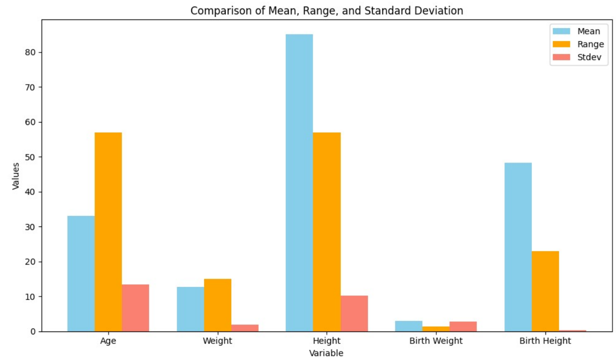

Fig. 2. Comparison Exploratory Data Analysis

deviation of 13.43 years, showing significant variation. The mean weight is 12.63 kg, and the mean height is 84.98 cm, both with moderate variability. Birth weight averages 2.97 kg, and birth height is 48.16 cm, with less variation. Skewness analysis shows age is slightly negatively skewed, weight is mildly positively skewed, and height is highly negatively skewed. Overall, the children's height for age (TB/U) status is normal, indicating appropriate growth.

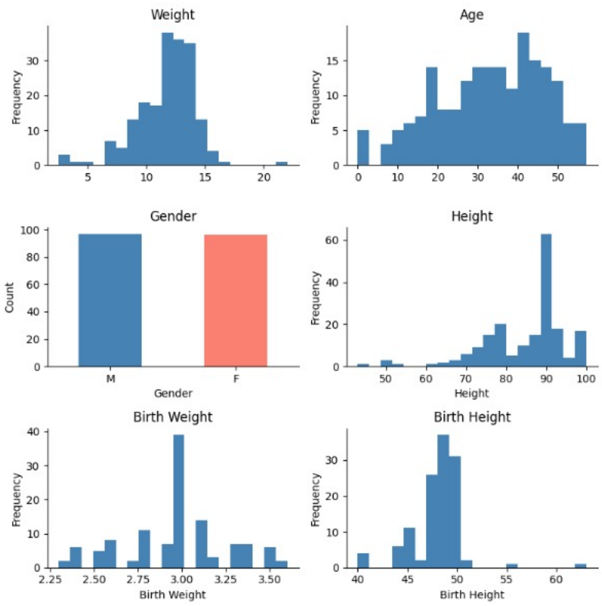

Fig. 3. Frequency Histogram on Each Selected Variable

Exploratory Data Analysis (EDA) in healthcare reveals key demographics and disease prevalence. In this study, EDA highlighted the significant prevalence of stunting, emphasizing the need for effective classification models for early detection [20]. From the exploratory data chart above (Fig. 3), this visualization illustrates the distribution of several biological variables in the stunting dataset. Body weight and birth weight show peaks around 10-15 kg and 3.0 kg, respectively. Most ages are between 30-50 years, while height and birth height are most commonly observed around 90 cm and 50-55 cm. The gender distribution is balanced, with slightly more females than males. Overall, the data reflects diverse distribution patterns, with some variables, such as height, approaching a normal distribution, which is important for understanding stunting prevalence among children in the dataset.

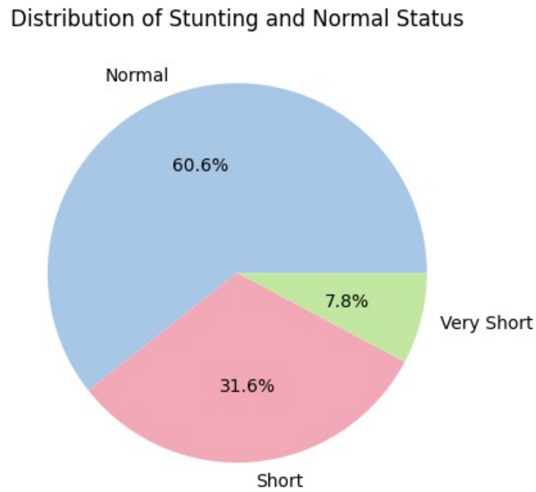

Fig. 4. Pie Chart Comparison of Stunting and Normal Status

Fig. 4 provides a snapshot of the distribution of toddlers growth status, highlighting the prevalence of both normal and stunted growth. Further analysis and consideration of relevant factors are necessary to understand the underlying causes and develop effective interventions to address the issue of stunting in toddlers.

B. Data Preprocessing

In the data processing stage, missing values were filled using manual stunting calculations to ensure completeness and accuracy. Relevant features such as Age, Weight, Height, TB/U, Gender, Birth Weight, and Birth Height were selected for analysis. TB/U was included to assess stunting status by comparing height to age. The Gender column was encoded using LabelEncoder, transforming "M" (male) to 0 and "F" (female) to 1. Additionally, feature normalization was performed using StandardScaler to standardize the features, ensuring balanced input for machine learning algorithms, which improves model accuracy and performance.

C. K-Nearest Neighbors (KNN) Model Results

With normalization, SMOTE, and tuning hyperparameter the KNN model significantly improved compared to the model without these techniques. In the 60:40 split, with k = 1, the model achieved 83.33% accuracy, 83.44% precision, 83.33% recall, and 83.37% F1 score, showing optimal performance. In the 70:30 split, with k = 6, accuracy slightly decreased to 77.59%, but the model remained strong. In the 80:20 split, with k = 1, the model achieved 76.92% accuracy. In the 90:10 split, with k = 2, accuracy dropped to 70.00%, but the model still showed solid performance.

D. Naive Bayes Model Results

With the use of SMOTE and normalization, the Naive Bayes model shows improved performance compared to the model without these techniques, although its results are slightly lower than those of KNN. In the 60:40 split, Naive

Bayes delivers solid performance with 73.08% accuracy. In the 80:20 split, accuracy increases to 76.92%, indicating that these techniques help enhance the model's performance with a smaller data proportion. In the 90:10 split, accuracy remains at 70.00%, with the continued use of SMOTE and normalization supporting the model to maintain consistent results despite changes in data distribution.

E. Comparative Analysis of Models

The approaches used to classify stunting datasets vary in their methodologies. In addition to employing different techniques, the classification process also differs when applied to imbalanced versus balanced datasets. Table III presents the results of classifying imbalanced data after model evaluation. Although the results appear favorable for the imbalanced dataset, this may actually indicate overfitting. The model may focus excessively on the majority class (class 0), neglecting the minority class. As a result, despite high accuracy, the performance for the minority class may be disproportionately low or even ignored.

TABLE III
COMPARISON SCENARIO IMBALANCED OF KNN AND NAIVE BAYES

Data Split	K-Nearest Neighbors		Naïve Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score
60:40	61.54%	55.14%	67.95%	63.85%
70:30	68.97%	63.81%	74.14%	71.78%
80:20	64.10%	59.13%	76.92%	75.34%
90:10	55.00%	50.37%	80.00%	80.06%

TABLE IV
COMPARISON SCENARIO BALANCED OF KNN AND NAIVE BAYES

Data Split	K-Nearest Neighbors		Naïve Bayes	
	Accuracy	F1 Score	Accuracy	F1 Score
60:40	83.33%	83.37%	73.08%	70.98%
70:30	77.59%	77.88%	70.69%	67.96%
80:20	76.92%	76.71%	76.92%	72.93%
90:10	70.00%	70.42%	70.00%	62.99%

Based on the performance comparison balanced Table IV between the K-Nearest Neighbors (KNN) and Naïve Bayes (NB) models across various data split scenarios, it can be concluded that the KNN model demonstrates better performance than the Naïve Bayes model in terms of accuracy and F1 score in most scenarios. In the 60:40 data split scenario, KNN achieved 83.33% accuracy and 83.37% F1 score, which is significantly higher than Naïve Bayes, which recorded 73.08% accuracy and 70.98% F1 score. KNN continues to outperform Naïve Bayes in the 70:30 and 80:20 scenarios, although its accuracy and F1 score slightly decrease, reaching 77.59% and 77.88% in the 70:30 scenario, and 76.92% and 76.71% in the 80:20 scenario. However, in the 90:10 scenario, KNN shows a noticeable decline, with 70.00% accuracy and 70.42% F1 score. Nevertheless, KNN still maintains better performance compared to Naïve Bayes in this scenario, where Naïve Bayes recorded 70.00% accuracy and 62.99% F1 score.

In Table V and VI, the KNN model delivers superior performance in the 60:40 split scenario, achieving higher accuracy, precision, recall, and F1 score. Specifically, KNN excels at classifying the Normal and Short classes, with a significant proportion of correct predictions in these categories. Naive Bayes, while accurately identifying the Normal class, struggles with the classification of Short and Very Short classes, resulting in a drop in overall performance. These observations indicate that the 60:40 split scenario yields the best results, with KNN outperforming Naive Bayes across various performance metrics.

TABLE V
CONFUSION MATRIX FOR SCENARIO KNN 60:40

Class Actual	Class Predicted		
	Normal	Short	Very Short
Normal	41	4	2
Short	3	21	1
Very Short	2	1	3

TABLE VI
CONFUSION MATRIX FOR SCENARIO NAIVE BAYES 60:40

Class Actual	Class Predicted		
	Normal	Short	Very Short
Normal	44	3	0
Short	11	11	3
Very Short	2	2	2

Table VII and VIII shows 70:30 data split scenario, KNN outperforms Naive Bayes in terms of accuracy and overall prediction performance. The KNN model is more effective at classifying the Normal and Short classes, with a higher number of correct predictions, while Naive Bayes struggles with classifying the Short class, making more prediction errors compared to KNN.

TABLE VII
CONFUSION MATRIX FOR SCENARIO KNN 70:30

Class Actual	Class Predicted		
	Normal	Short	Very Short
Normal	28	4	3
Short	4	14	0
Very Short	1	1	3

TABLE VIII
CONFUSION MATRIX FOR SCENARIO NAIVE BAYES 70:30

Class Actual	Class Predicted		
	Normal	Short	Very Short
Normal	33	2	0
Short	8	7	3
Very Short	2	2	1

In Table IX and X, both KNN and Naive Bayes show similar results, but Naive Bayes has slight difficulties classifying the Short and Very Short classes, despite performing well with the Normal class. KNN has a slight edge with more consistent predictions across these classes. This may be attributed to KNN's ability to better capture local patterns within the data due to its reliance on proximity-based classification. On the other hand, Naive Bayes assumes feature independence, which might limit its performance when features are correlated, as is often the case in stunting prediction datasets.

In Table XI and XII, both models experience a decline in performance, but KNN still shows slightly better results, with higher accuracy and F1-score performance compared to Naive Bayes, even though both face challenges in classifying rarer classes like Very Short.

TABLE IX
CONFUSION MATRIX FOR SCENARIO KNN 80:20

Class Actual	Class Predicted		
	Normal	**Short**	**Very Short**
Normal	19	4	1
Short	2	10	0
Very Short	1	1	1

TABLE X
CONFUSION MATRIX FOR SCENARIO NAIVE BAYES 80:20

Class Actual	Class Predicted		
	Normal	**Short**	**Very Short**
Normal	23	1	0
Short	5	7	0
Very Short	2	1	0

TABLE XI
CONFUSION MATRIX FOR SCENARIO KNN 90:10

Class Actual	Class Predicted		
	Normal	**Short**	**Very Short**
Normal	9	2	1
Short	2	4	0
Very Short	0	1	1

TABLE XII
CONFUSION MATRIX FOR SCENARIO NAIVE BAYES 90:10

Class Actual	Class Predicted		
	Normal	**Short**	**Very Short**
Normal	12	0	0
Short	4	0	0
Very Short	1	1	30

IV. CONCLUSION

This study evaluates the effectiveness of classification techniques in analyzing the stunting dataset, comparing results between balanced and imbalanced datasets. Classification methods, including KNN and Naive Bayes, were employed alongside data preprocessing steps such as missing value imputation, encoding, and normalization. The results indicate that models trained on imbalanced datasets, particularly Naive Bayes, achieve high overall accuracy but exhibit significant overfitting, neglecting the minority class (”very short”) with very low F1 Scores. Various data splitting scenarios (60:40, 70:30, 80:20, and 90:10) were tested, with the best results achieved on the balanced dataset in the 60:40 split using the KNN model, yielding an accuracy of 83.33% and an F1 Score of 83.37%. This research highlights challenges such as the limited dataset size, which can hinder model generalization and reliability. Future studies should explore solutions to address the limitations of small datasets to improve model performance. By tackling these challenges, future research can develop more reliable models for stunting prediction and intervention.

REFERENCES

[1] N. F. Sahamony and H. R. Terttiaavini, “Analysis of performance comparison of machine learning models for predicting stunting risk in children’s growth analisis perbandingan kinerja model machine learning untuk memprediksi risiko stunting pada pertumbuhan anak,” *Journal Homepage: https://journal. irpi. or. id/index. php/malcom*, vol. 4, no. 2, pp. 413–422, 2024.

[2] L. Yulaikhah, R. Kumorojati, D. Puspitasari *et al.*, “Upaya pencegahan stunting melalui deteksi dini dan edukasi orangtua dan kader posyandu di dukuh gupak warak desa sendangsari pajangan bantul yogyakarta,” *Journal of Innovation in Community Empowerment*, vol. 2, no. 2, pp. 71–78, 2020.

[3] J. Zeniarja, K. Widia, and R. R. Sani, “Penerapan algoritma naive bayes dan forward selection dalam pengklasifikasian status gizi stunting pada puskesmas pandanaran semarang,” *JOINS (Journal of Information System)*, vol. 5, no. 1, pp. 1–9, 2020.

[4] S. K. P. Loka and A. Marsal, “Perbandingan algoritma k-nearest neighbor dan naïve bayes classifier untuk klasifikasi status gizi pada balita: Comparison algorithm of k-nearest neighbor and naïve bayes classifier for classifying nutritional status in toddlers,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 8–14, 2023.

[5] W. C. Wahyudin, F. M. Hana, and A. Prihandono, “Prediksi stunting pada balita di rumah sakit kota semarang menggunakan naive bayes,” *Jurnal Ilmu Komputer Dan Matematika*, vol. 4, no. 1, pp. 32–36, 2023.

[6] W. Widhari, A. Triayudi, and R. T. K. Sari, “Implementation of naïve bayes and k-nn algorithms in diagnosing stunting in children,” *SAGA: Journal of Technology and Information System*, vol. 2, no. 1, pp. 164–174, 2024.

[7] D. K. D. Damayanti and M. Jakfar, “Klasifikasi status stunting balita menggunakan algoritma fuzzy c-means (studi kasus posyandu rw 01 kelurahan jepara surabaya),” *MATHunesa: Jurnal Ilmiah Matematika*, vol. 11, no. 03, pp. 533–542, 2023.

[8] V. Herliansyah, R. Latuconsina, and A. Dinimaharawati, “Prediksi stunting pada balita dengan menggunakan algoritma klasifikasi naïve-bayes,” *eProceedings of Engineering*, vol. 8, no. 5, 2021.

[9] U. R. Gurning, S. F. Octavia, D. R. Andriyani, N. Nurainun, and I. Permana, “Prediksi risiko stunting pada keluarga menggunakan naïve bayes classifier dan chi-square: Prediction of stunting risk in families using naïve bayes classifier and chi-square,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 172–180, 2024.

[10] K. M. Rajabi, W. Witanti, and R. Yuniarti, “Penerapan algoritma k-nearest neighbor (knn) dengan fitur relief-f dalam penentuan status stunting,” *Innovative: Journal Of Social Science Research*, vol. 3, no. 4, pp. 3555–3568, 2023.

[11] S. Lonang, A. Yudhana, and M. K. Biddinika, “Analisis komparatif kinerja algoritma machine learning untuk deteksi stunting,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, pp. 2109–2117, 2023.

[12] A. J. Susilo, D. Remawati *et al.*, “Implementasi naïve bayes dalam pemilihan jenis bahan pembuatan meja,” *Jurnal Ilmiah SINUS*, vol. 21, no. 1, pp. 39–50, 2023.

[13] F. A. Apriansyah, A. Hermawan, and D. Avianto, “Optimization of k value in knn algorithm for spam and ham classification in sms texts,” *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 4, no. 2, pp. 767–779, 2024.

[14] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, “Aplikasi naïve bayes classifier (nbc) pada klasifikasi status gizi balita stunting dengan pengujian k-fold cross validation,” *Jurnal Gaussian*, vol. 11, no. 1, pp. 130–139, 2022.

[15] A. Sanjaya and T. Wahyana, “Penerapan metode k-nearest neighbour untuk sistem prediksi kelulusan siswa mts nurul muslimin berbasis website,” *Journal Transformation of Mandalika*, vol. 3, no. 2, pp. 31–47, 2022.

[16] M. H. Rifqo and A. Wijaya, “Implementasi algoritma naive bayes dalam penentuan pemberian kredit,” *Pseudocode*, vol. 4, no. 2, pp. 120–128, 2017.

[17] N. Yusliani, A. Syahrini *et al.*, “Klasifikasi pertanyaan berbahasa indonesia menggunakan algoritma support vector machine dan seleksi fitur mutual information,” *JUPITER: Jurnal Penelitian Ilmu dan Teknologi Komputer*, vol. 14, no. 2-a, pp. 44–52, 2022.

[18] J. Pardede and D. P. Pamungkas, “The impact of balanced data techniques on classification model performance,” *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 401–412, 2024.

[19] L. Lindawati, M. Fadhli, and A. S. Wardana, “Optimasi gaussian naïve bayes dengan hyperparameter tuning dan univariate feature selection dalam prediksi cuaca,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 7, no. 2, pp. 237–246, 2023.

[20] H. W. Dhany, S. Sutarman, and F. Izhari, “Exploratory data analysis (eda) methods for healthcare classification,” *Journal of Intelligent Decision Support System (IDSS)*, vol. 6, no. 4, pp. 209–215, 2023.