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

Vol. 4 No 1 – May 2026
DOI: 10.59688/732114

PERFORMANCE ANALYSIS OF NAÏVE BAYES CLASSIFIERS BASED ON THE INFORMATION GAIN-BASED FEATURE SELECTION WITH MULTICOLLINEARITY ANALYSIS

Luh Putu Risma Noviana[1], I Gede Aris Gunadi[2], I Made Gede Sunarya[3]

[1,2,3]Department of Computer Science, Faculty of Computer Science, Universitas Pendidikan Ganesha
Jalan Udayana Nomor 11, Banjar Bali, Kec. Buleleng, Kabupaten Buleleng, Bali

e-mail: risma.noviana@undiksha.ac.id[1],igedearisgunadi@undiksha.ac.id[2], sunarya@undiksha.ac.id[3]

Received : 28 April 2026	Accepted : 26 May 2026	Published : 30 May 2026

Abstract

This study aims to analyze the performance of the Naïve Bayes Classifier algorithm by comparing several feature selection methods, namely Information Gain-based feature selection, Multicollinearity-based feature selection, and a combination of Information Gain and Multicollinearity. The dataset used in this study consists of 337 toddler stunting cases obtained from Kintamani I and VI Public Health Centers. The experiment was conducted using four testing scenarios: (1) Naïve Bayes Classifier without feature selection, (2) Naïve Bayes Classifier with Information Gain feature selection, (3) Naïve Bayes Classifier with Multicollinearity feature selection, and (4) Naïve Bayes Classifier with a combination of Information Gain and Multicollinearity feature selection. All experiments used a data split of 70% training data and 30% testing data, while model performance was evaluated using a confusion matrix.

In the Information Gain feature selection stage, several features achieved the highest gain values, namely BPJS with a gain value of 1.0, immunization with a gain value of 1.0, age with a gain value of 0.842, maternal pregnancy history with a gain value of 0.791, and smoking habits with a gain value of 0.756. These features were retained in the final combined model because they contributed the most to the stunting classification process. In addition to improving predictive performance, the combination of Information Gain and Multicollinearity was also able to reduce feature redundancy, resulting in a more stable classification model.

The results showed that the accuracy of the Naïve Bayes Classifier without feature selection was 90.10%, the Naïve Bayes Classifier with Information Gain feature selection achieved 95.05%, the Naïve Bayes Classifier with Multicollinearity feature selection achieved 93.07%, and the Naïve Bayes Classifier with a combination of Information Gain and Multicollinearity achieved the highest accuracy of 96.04%. These findings indicate that the combination of Information Gain and Multicollinearity produced the best performance among all tested methods. In addition, a coefficient of determination (R Square) test was conducted using SPSS, resulting in a value of 0.577, indicating that 57.7% of stunting classification was influenced by independent variables such as age, BPJS, immunization, smoking habits, and maternal pregnancy history, while the remaining 42.3% was influenced by other factors outside the scope of this study. The results also indicate that the Naïve Bayes algorithm combined with Information Gain feature selection and multicollinearity testing can be used as a stable and effective approach for early stunting classification to support decision-making in public health services.

Keywords: *Stunting, Naïve Bayes Classifier, Information Gain, Multicollinearity, Data Classification.*

1. INTRODUCTION

Stunting is one of the major health problems among toddlers that remains a primary concern in Indonesia because it affects physical growth, cognitive development, and the quality of future human resources. Based on data from the Basic Health Research (Riskesdas) and reports from the Ministry of Health of the Republic of

Indonesia, the prevalence of stunting in Indonesia is still relatively high, requiring accurate and measurable data-driven intervention efforts. Therefore, an information technology-based approach is needed to support the process of identifying and classifying stunting more effectively

The dataset used in this study was obtained from Kintamani I and Kintamani VI Public Health Centers, Bangli Regency, Bali. This dataset has unique characteristics because it represents rural public health data with diverse nutritional conditions, uneven attribute distribution, and the potential for multicollinearity among anthropometric variables. These conditions make the dataset more challenging compared to general datasets commonly used in stunting classification research.

Several previous studies have shown that the Naïve Bayes Classifier method has relatively good performance in classifying health-related data. The Naïve Bayes Classifier method was able to achieve an accuracy rate of 88% in the classification of toddler nutritional status [2]. The implementation of Information Gain feature selection was able to improve the accuracy of the Naïve Bayes method from 80% to 90% [9]. The combination of Naïve Bayes and Information Gain also improved classification performance with an average accuracy of 80% [14]. Gain-based feature selection can improve the efficiency and accuracy of classification models. Meanwhile, statistical approaches such as multicollinearity testing are also important to ensure that there is no high correlation among independent variables that could affect the model results. Recent studies have also demonstrated that machine learning approaches can improve the early detection and prediction of stunting prevalence through classification and optimization models [1], [10].

Information Gain and Multicollinearity are two complementary methods in the feature selection process. Information Gain functions to measure how much contribution a feature provides to the classification process, enabling the selection of the most relevant attributes with the highest predictive capability for stunting status. Thus, the use of Information Gain can improve model accuracy because only the most important features are utilized in the classification process. However, even if a feature has high predictive value, there may still be strong relationships or correlations among independent variables that can cause information redundancy. Therefore, the Multicollinearity method is used to detect and reduce excessively high relationships among features through tolerance value testing and Variance Inflation Factor (VIF) analysis. The application of Multicollinearity helps maintain the stability of the classification model by eliminating highly correlated features, making the model more optimal, consistent, and less prone to performance degradation due to data redundancy. In other words, Information Gain focuses on optimizing predictive capability, while Multicollinearity plays a role in maintaining model stability by reducing redundancy among features. The combination of these two methods is expected to produce better classification performance compared to using only a single feature selection method.

Although various studies have demonstrated performance improvements through the use of Naïve Bayes and feature selection methods, there are still limitations in examining the combination of multiple feature selection methods simultaneously, particularly between Information Gain andMulticollinearity. In addition, there have been limited studies that comprehensively evaluate the performance of these methods using several testing scenarios on real-case datasets, especially stunting data at the public health center level. Therefore, evaluating the performance of the methods is important to determine the level of accuracy, precision, and contribution of each feature in improving classification results. This evaluation is also necessary to ensure that the model used is not only accurate but also stable and reliable in decision-making.

Based on these problems, this study aims to analyze and compare the performance of the Naïve Bayes Classifier method without feature selection, with Information Gain feature selection, with Multicollinearity feature selection, and with a combination of both methods in classifying stunting status. The testing was conducted using toddler data from Kintamani I and VI Public Health Centers with a data split of 70% training data and 30% testing data. The results of this study are expected to contribute to the development of more accurate classification methods and serve as a reference for data-driven stunting case management in the health sector.

2. RESEARCH METHODOLOGY

2.1 Research Design

This study uses an experimental quantitative approach with the supervised learning method to classify stunting status in toddlers. The model used is the Naïve Bayes Classifier with several feature selection scenarios, namely no feature selection, Information Gain, Multicollinearity, and a combination of both. The research design includes the stages of data collection, preprocessing, feature selection, modeling, evaluation, and analysis of results [2], [4], [14].

2.2 Data Source and Collection Method

The data used in this study are secondary data obtained from: Kintamani I and VI Health Centers, Bangli Regency, Bali. The number of datasets used was 337 data for toddlers consisting of data on health and environmental conditions. Data collection methods include:

1. Documentation: taking data on medical records of toddlers from health centers
2. Observation: verification of data related to field conditions
3. Literature study: to support the selection of research methods and variables [1], [8]

The table shown in the article contains only a portion of the data (example 10 data) as an illustration, while all 337 data are used in the training and testing of the model. To support research transparency, the full dataset can be made available through an external repository

Table 1. Data of Toddlers of Kintamani Health Center I and VI
[Source: Author, 2022]

No.	BB	TB	Age	BPJS	Air Clean	stump and	Latrines Healthy	Immunity	Smoking	HISTORY Mother Pregnant	Classification
1	11,9	84	2,11	0	1	0	1	1	0	0	Stunting
2	11	84	2,6	1	1	0	1	1	1	0	Stunting
3	9,3	81,3	2,3	0	1	0	1	1	0	0	Stunting
4	10,8	85	4,7	0	1	0	1	1	1	0	Stunting
5	12	97	4,6	1	1	0	1	1	1	0	Stunting
6	13,5	93	3,8	1	1	0	1	1	1	0	Stunting
7	12	91	3,11	0	1	0	1	1	1	0	Stunting
8	11,8	92	3,11	0	1	0	1	1	1	0	Stunting
9	10,5	92	1,6	0	1	0	1	1	1	0	Stunting
10	9	76	0,8	0	1	0	1	1	1	0	Stunting

2.3 Research Parameters and Variables

The variables in this study consist of dependent and independent variables used in the classification process. The selection of variables refers to WHO standards for assessing the nutritional status of toddlers.

The variables in this study consisted of:

1. Dependent Variable (Target): Stunting status (Stunting = 1, Normal = 0)
2. Independent Variables (Features): Weight, Height, Age, BPJS, Clean Water, Worms, Healthy Latrines, Immunization, Smoking, History of Pregnant Women [5], [10]

Categorical data is normalized into numerical form (0 and 1), while numerical data is used directly in the classification process.

2.4 Research Tools and Instruments

The tools used in this study include:

1. Software: RapidMiner for classification modeling and Information Gain calculation processes.
2. Supporting Software: SPSS for multicollinearity testing and coefficient of determination analysis. Multicollinearity testing was conducted using SPSS software.
3. Hardware: Laptop/PC with standard specifications for data processing

The main instruments in this study are the toddler dataset and the Naïve Bayes Classifier algorithm model.

2.5 Research Stages

The research stages include data preprocessing, feature selection, modeling, and evaluation. Feature selection is performed using Information Gain and multicollinearity analysis to avoid redundancy among variables. This approach is consistent with feature selection concepts aimed at improving model performance [4], [9].

The stages of research are carried out systematically from start to finish as follows:

1. Problem Identification

   Identify stunting problems and data classification needs.

2. Data Collection
   Collected datasets from Kintamani I and VI Health Centers.
3. Preprocessing Data:
   a. Data cleaning
   b. Data transformation (normalization and encoding)
   c. Data sharing of training (70%) and testing (30%) [13]
4. Feature Selection
   a. Information Gain: selects features based on the highest gain value
   b. Multicollinearity: testing relationships between variables using Tolerance and VIF values
      Tolerance and VIF Calculation:
      Tolerance = $1 - R^2$

$$VIF = \frac{1}{\text{Tolerance}} \tag{1}$$

   Criteria:
   a. Tolerance > 0.10 → multicollinearity does not occur
   b. VIF < 10 → multicollinearity does not occur

5. Classification Modeling
   Using the Naïve Bayes Classifier algorithm with 4 scenarios:
   a. No feature selection
   b. With Information Gain
   c. With Multicollinearity
   d. Combination of Information Gain and Multicollinearity
6. Model Evaluation
   Using the Confusion Matrix to calculate:
   a. Accuracy
   b. Precision
   c. Recall
   d. F1 Score
7. Analyze the results
   Compare the performance of each method to determine the best model.
8. Conclusion
   Draw conclusions based on the results of model evaluation.

2.6 Research Flow Design (Flowchart)

The research flow follows the standard data mining process, starting from problem identification, data collection, data preprocessing, feature selection, model building, evaluation, and conclusion drawing. The classification model used in this study is the Naïve Bayes Classifier, which is recognized as an effective probabilistic classification method for handling health-related datasets [11]. The overall research process is illustrated in Figure 1.

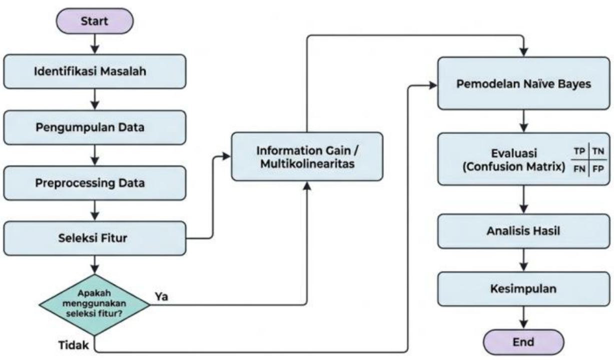

Figure 1. Naive Bayes Modeling Flowchart
[Source:Author, 2023]

During the preprocessing stage, several processes were carried out to prepare the dataset before classification. These processes included data cleaning, handling missing values, data transformation, and normalization of continuous variables. Continuous variables such as weight, height, and age were normalized to ensure compatibility with the probability distribution assumptions of the Naïve Bayes algorithm and to reduce differences in data scale among attributes. In this study, Min-Max Normalization was applied to transform the values of continuous variables into a range between 0 and 1 using the following formula:

$$\chi^{I} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Where x represents the original value $\boldsymbol{X_{min}}$, is the minimum value of the attribute, and $\boldsymbol{X_{max}}$ is the maximum value of the attribute. This normalization process helps prevent attributes with larger numeric ranges from dominating the classification process and improves the stability and performance of the Naïve Bayes model.

After preprocessing, feature selection was conducted using Information Gain and Multicollinearity methods. Information Gain was used to select the most relevant features based on their contribution to stunting classification, while Multicollinearity testing was applied to identify and reduce highly correlated independent variables using tolerance values and Variance Inflation Factor (VIF). The selected features were then used in the Naïve Bayes modeling process. Model evaluation was performed using a confusion matrix to measure classification performance in terms of accuracy, precision, true positive, true negative, false positive, and false negative values [7].

3. RESULT AND DISCUSSION

3.1 Data Description

The dataset consists of toddler health data used for stunting classification based on WHO standards. The selected variables represent both health and environmental conditions affecting toddlers. The dataset used in this study consisted of 337 data from toddlers obtained from the Kintamani I and VI Health Centers. Each data has several parameters that are used for the classification of stunting status. The explanation of the main parameters is as follows:

1. BB (Weight): measured in kilograms (kg). Example: BB = 11.9 means that the weight of a toddler is 11.9 kg.
2. TB (Height): measured in centimeters (cm). Example: TB = 84 means that the height of the toddler is 84 cm.
3. Age: expressed in decimal years. Example: Age = 2.11 means that the age of toddlers is 2.11 years (approximately 2 years and 1 month).

Other parameters such as BPJS, clean water, immunization, and history of pregnant women are categorical variables that are converted into numerical form (0 = no, 1 = yes). The basis for the use of this parameter refers to the WHO standard in assessing the nutritional status of toddlers, especially the anthropometric indicator TB/U (Height by Age) which is used to determine stunting status.

3.2 Test Methods

The testing method applies a supervised learning approach with a training and testing data split [12]. Model evaluation is conducted using a confusion matrix, which is a common evaluation method in classification tasks. The testing method in this study uses a supervised learning approach with the Naïve Bayes Classifier algorithm. The dataset is divided into:

1. 70% data training
2. 30% data testing

Model evaluation was carried out using the Confusion Matrix, which consisted of:

1. True Positive (TP): stunting data predicted to be true
2. True Negative (TN): normal data that is predicted to be true
3. False Positive (FP): normal but predicted data of stunting
4. False Negative (FN): stunting data but predicted to be normal

From the confusion matrix, the following evaluation metrics are calculated:

1. Accuracy = $\frac{TP + TN}{TP + TN + FP + FN}$
2. Precision= $\frac{TP}{TP + FP}$
3. Recall = $\frac{TP}{TP + FN}$
4. F1 Score = $2 \times \frac{Precision \times Recall}{Precision + Recall}$

This method is used because it is able to provide a comprehensive picture of the model's performance, not only in terms of accuracy but also in terms of classification errors.

3.3 Discussion

The results show that the Naïve Bayes method provides good performance in classifying stunting, consistent with previous studies. The application of Information Gain significantly improves classification accuracy.

The test was carried out using 4 scenarios, namely:

1. Experiment Using the Naïve Bayes Classifier Method

   The experiment was conducted using the Naïve Bayes Classifier method by utilizing 10 features, namely body weight, height, age, BPJS ownership, access to clean water, history of helminth infection, availability of a healthy latrine, immunization status, smoking habits in the environment, and maternal pregnancy history. The dataset used consisted of 337 toddler records. The validation process was carried out using a validation percentage technique, with 70% of the data used for training and 30% for testing. The model performance was evaluated using a Confusion Matrix, which produced classifications in the form of true stunting and true normal. Based on the test results, the model achieved an accuracy of 90.10%, a precision of 83.33%, and a recall of 100%. The recall value of 100% indicates that the model successfully identified all stunting cases, although there may still be some misclassification in normal cases, which affects the precision value [15]. The evaluation results can be seen in Figure 2.

accuracy: 90.10%

	true Stunting	true Normal	class precision
pred. Stunting	41	0	100.00%
pred. Normal	10	50	83.33%
class recall	80.39%	100.00%	

Figure 2. Accuracy Performance Matrix of the Naïve Bayes Classifier Model
[Source:Author, 2023]

2. Experiment with Information Gain Feature Selection using the Naïve Bayes Classifier Method

   This experiment was conducted by applying the Naïve Bayes Classifier method combined with the Information Gain feature selection technique. A total of 10 features were used in this study, namely body weight, height, age, BPJS ownership, access to clean water, availability of a healthy latrine, history of helminth infection, immunization status, smoking habits in the environment, and maternal pregnancy history. The criteria for N-Gain values in this study were categorized into three levels: gain > 0.7 is classified as high, 0.3 ≤ gain ≤ 0.7 as moderate, and gain < 0.3 as low. Based on the results of the Information Gain calculation, three features were identified as having a very high level of influence, namely BPJS, immunization, and smoking, each with a gain value of 1. A gain value of 1 indicates that these features have a very significant influence on the classification process and can be considered highly optimal in distinguishing between data classes. Meanwhile, the features of body weight, height, age, access to clean water, history of helminth infection, availability of a healthy latrine, and maternal pregnancy history obtained a gain value of 0. This indicates that these features do not provide a significant contribution to the classification process in the dataset used. Furthermore, the classification test was carried out using the Naïve Bayes Classifier method with a total of 337 toddler records. The validation process employed a validation percentage technique, with 70% of the data used as training data and 30% as testing data. The model performance was evaluated using a Confusion Matrix, which resulted in classifications in the form of true stunting and true normal. Based on the test results, the model using Information Gain feature selection achieved an accuracy of 95.05%, a precision of 90.91%, and a recall of 100%. These results indicate an improvement in model performance compared to the model without feature selection, particularly in terms of accuracy and precision, while recall remains optimal. The evaluation results can be seen in Figure 3.

accuracy: 95.05%

	true Stunting	true Normal	class precision
pred. Stunting	46	0	100.00%
pred. Normal	5	50	90.91%
class recall	90.20%	100.00%	

Figure 3. Accuracy Performance Matrix of the Naïve Bayes Classifier with Information Gain Feature Selection
[Source:Author, 2023]

3. Experiment with Multicollinearity Testing using the Naïve Bayes Classifier Method
   This experiment was conducted by applying the Naïve Bayes Classifier method while considering multicollinearity testing among the research variables. A total of 10 features were used in this study, namely body weight, height, age, BPJS ownership, access to clean water, availability of a healthy latrine, history of helminth infection, immunization status, smoking habits in the environment, and maternal pregnancy history. The dependent variable in this study is stunting status (stunting and normal), while the independent variables consist of all the aforementioned features. Multicollinearity testing was performed using IBM SPSS Statistics by examining the Tolerance and Variance Inflation Factor (VIF) values. Based on the results shown in Figure 4.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	0,762	0,212		3,597	0,000		
	BB	-0,008	0,017	-0,043	-0,488	0,626	0,157	6,349
	TB	-0,002	0,004	-0,044	-0,484	0,629	0,153	6,532
	Umur	0,058	0,016	0,152	3,540	0,000	0,673	1,485
	BPJS	-0,410	0,045	-0,355	-9,132	0,000	0,819	1,221
	Merokok	0,519	0,043	0,476	11,977	0,000	0,784	1,276
	RiwayatIbu Hamil	0,362	0,100	0,128	3,613	0,000	0,989	1,011

Figure 4. Multicollinearity Test Results Using IBM SPSS Statistics
[Source:Author, 2023]

Variables have Tolerance values greater than 0.10 (Tolerance > 0.10) and VIF values less than 10.0 (VIF < 10.0). This indicates that there is no evidence of multicollinearity among the independent variables in relation to the dependent variable in the model. Furthermore, a t-test was conducted to determine the effect of each independent variable on the dependent variable. The results show that, among the variables analyzed, four features have a significant effect on stunting, while two features do not show a significant effect. Subsequently, the classification process was carried out using the Naïve Bayes Classifier method with a total of 337 toddler records. The validation process employed a validation percentage technique, with 70% of the data used as training data and 30% as testing data. The model performance was evaluated using a Confusion Matrix, which produced classifications in the form of true stunting and true normal. Based on the test results, the model considering multicollinearity testing achieved an accuracy of 93.07%, a precision of 87.72%, and a recall of 100%. These results indicate that the model has an excellent ability to identify all stunting cases (optimal recall), although there are still some misclassifications in normal cases that affect the precision value. The evaluation results can be seen in Figure 5.

accuracy: 93.07%	true Stunting	true Normal	class precision
pred. Stunting	44	0	100.00%
pred. Normal	7	50	87.72%
class recall	86.27%	100.00%	

Figure 5. Accuracy Performance Matrix of the Naïve Bayes Classifier with Multicollinearity Feature Selection
[Source:Author, 2023]

4. Experiment with Information Gain Feature Selection and Multicollinearity Testing using the Naïve Bayes Classifier Method
This experiment was conducted by combining the Information Gain feature selection technique and multicollinearity testing in the implementation of the Naïve Bayes Classifier method. Based on the results of feature selection and multicollinearity testing, five main features were identified and used in the classification process, namely age, BPJS ownership, immunization status, smoking habits in the environment, and maternal pregnancy history. The dataset used in this study consisted of 337 toddler records. The validation process was carried out using a validation percentage technique, with 70% of the data used as training data and 30% as testing data. The model performance was evaluated using a Confusion Matrix, which produced classifications in the form of true stunting and true normal. Based on the test results, the model that utilized the combination of Information Gain feature selection and multicollinearity testing achieved an accuracy of 96.04%, a precision of 92.59%, and a recall of 100%. These results indicate that the combination of both feature selection approaches significantly improves the model performance, particularly in terms of accuracy and precision, while maintaining an optimal recall in detecting all stunting cases. The evaluation results can be seen in Figure 6.

accuracy: 96.04%	true Stunting	true Normal	class precision
pred. Stunting	47	0	100.00%
pred. Normal	4	50	92.59%
class recall	92.16%	100.00%	

Figure 6. Accuracy Performance Matrix of the Naïve Bayes Classifier with Information Gain and Multicollinearity Feature Selection
[Source:Author, 2023]

Table 2. Accuracy Comparison Between Current Research and Previous

Research	Method	Dataset	Accuracy
Previous Study Ref [2]	Naïve Bayes Classifier	Toddler Stunting Dataset	88%
Previous Study Ref [6]	Decision Tree Algorithm	Toddler Stunting Dataset	91%
Previous Study Ref [10]	Random Forest Algorithm	Stunting Classification Dataset	94%
This Study	Naïve Bayes + Information Gain	337 Toddler Records from Kintamani I and VI Public Health Centers	95.05%
This Study	Naïve Bayes + Multicollinearity	337 Toddler Records from Kintamani I and VI Public Health Centers	93/07
This Study	Naïve Bayes Classifier features Information Gain + Multicollinearity	337 Toddler Records from Kintamani I and VI Public Health Centers	96.04%

Based on Table 2, the proposed method achieved the highest accuracy of 96.04% compared to several previous studies. This indicates that the combination of Information Gain feature selection and multicollinearity testing was able to improve classification performance as well as model stability.

4. CONCLUSION

Based on the results of the research that has been conducted, it can be concluded that the Naïve Bayes Classifier method can be effectively used to classify stunting status in toddlers with a high level of accuracy. The implementation of feature selection has proven to have a significant effect on improving model performance, where the method without feature selection produced an accuracy of 90.10%, while the use of Information Gain increased the accuracy to 95.05%, and the use of Multicollinearity achieved an accuracy of 93.07%. The combination of these two feature selection methods produced the best performance with an accuracy of 96.04%.
In addition, the coefficient of determination (R Square) result of 0.577 indicates that 57.7% of the independent variables used in this study were able to explain the stunting classification, while the remaining percentage was influenced by other factors outside the model. This finding shows that the feature selection and feature processing stages play an important role in improving the accuracy and stability of the classification model. The results also indicate that the Naïve Bayes classification model with Information Gain feature selection and multicollinearity testing was able to produce high accuracy and stable performance in stunting classification. Multicollinearity analysis successfully reduced redundant variables, thereby improving the consistency and reliability of the model.
Therefore, the combination of Information Gain and Multicollinearity methods is recommended as an optimal approach to improve the performance of the Naïve Bayes Classifier. Future research is suggested to include additional variables related to socioeconomic factors, nutritional consumption patterns, environmental sanitation, and family health history in order to improve the predictive capability of the model. Furthermore, future studies are also recommended to compare the proposed method with other classification methods such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Neural Network to obtain a more comprehensive evaluation of classification performance.
Since stunting datasets often have imbalanced class distributions, future research is also recommended to apply imbalance handling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) [3], Adaptive Synthetic Sampling (ADASYN), or undersampling methods to improve the quality of training data and reduce classification bias toward the majority class. The implementation of these techniques is expected to improve the accuracy, precision, recall, and stability of the classification model so that prediction results become more optimal and reliable in supporting decision-making in the health sector.

REFERENCES

[1] A. Maulana, M. Ilham, and S. Lonang, "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, 2025. doi: 10.64021/ijmst.1.1.28-33.

[2] A. Rizki, D. Lestari, and P. W. Utama, "Implementation of Naïve Bayes Classifier for Stunting Classification in Toddlers," Journal of Informatics and Health Technology, vol. 5, no. 1, pp. 45–53, 2024.

[3] D. Sartika, Y. Florensia, and M. Utami, "The Effect of the SMOTE Method on the Classification of Toddler Nutritional Status Using the Naïve Bayes Method," Jurnal Sisfokom, vol. 14, no. 3, pp. 238–245, 2025. doi: 10.32736/sisfokom.v14i3.2381.

[4] H. M. S. Pasaribu, D. Abdullah, and L. R. Magdalena, "Classification of Stunting Status Using the Naive Bayes Classifier Algorithm with Backward Elimination Feature Selection," JINAV: Journal of Information and Visualization, vol. 6, no. 1, pp. 12–19, 2025.

[5] H. R. G. Turnip, M. R. N. Habibi, and A. W. Wijayadi, "Klasifikasi Penentuan Stunting Menggunakan Metode Naïve Bayes," Jurnal Sains dan Teknologi, vol. 2, no. 1, pp. 64–74, 2025.

[6] M. Arimbawa, S. Rahmawati, and A. Putra, "Implementation of Decision Tree Algorithm for Early Stunting Classification in Toddlers," International Journal of Data Science and Health Informatics, vol. 6, no. 1, pp. 21–30, 2024.

[7] M. F. Muchtar, R. Laila, D. Shinta, and H. M. Y. Pasca, "Perbandingan Algoritma Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," The Indonesian Journal of Computer Science, vol. 13, no. 4, pp. 50–59, 2024. doi: 10.33022/ijcs.v13i4.4055.

[8] M. I. Prasasti and D. Normawati, "Sistem Pakar Deteksi Dini Status Stunting Pada Balita Menggunakan Metode Naive Bayes," Jurnal Media Informatika Budidarma, vol. 7, no. 3, pp. 1276–1286, 2023. doi:

10.30865/mib.v7i3.6443.
[9] M. Yunus, M. K. Biddinika, and A. F. Rozie, "Optimasi Algoritma Naïve Bayes Menggunakan Fitur Seleksi Backward Elimination untuk Klasifikasi Prevalensi Stunting," Decode: Jurnal Pendidikan Teknologi Informasi, vol. 3, no. 2, pp. 278–285, 2023. doi: 10.51454/decode.v3i2.188.
[10] N. P. Sari, A. Fauzi, and K. M. Azhar, "Implementation of Random Forest Algorithm for Stunting Classification in Toddlers," Journal of Artificial Intelligence and Health Informatics, vol. 7, no. 1, pp. 60–72, 2024.
[11] N. W. Azani and M. Afdal, "Implementation Of Naïve Bayes Classifier And Support Vector Machine For Stunting Classification," The Indonesian Journal of Computer Science, vol. 13, no. 4, pp. 40–49, 2024. doi: 10.33022/ijcs.v13i4.4040.
[12] R. Ratnasari, A. J. Wahidin, and T. Andika, "Deteksi dini stunting pada anak berdasarkan indikator antropometri dengan menggunakan algoritma machine learning," Jurnal Algoritma, vol. 21, no. 2, pp. 378–385, 2024.
[13] S. Melyani, S. Z. Harahap, and I. Irawan, "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. doi: 10.33395/sinkron.v8i2.1355.
[14] T. Hardiani and R. N. Putri, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita," Digital Transformation Technology, vol. 4, no. 1, pp. 621–627, 2024. doi: 10.47709/digitech.v4i1.448.
[15] M. Y. Titimeidara and W. Hadikurniawati, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," Jurnal Ilmiah Informatika, vol. 9, no. 1, pp. 54–59, 2021.