Indonesian Journal of Modern Science and Technology (IJMST)
Vol. 1, No. 1, January 2025, pp. 28-33

Classification of Stunting in Toddlers using Naive Bayes Method and Decision Tree

Adrian Maulana [1,*], Muhammad Ilham [2], Syahrani Lonang [3], Nazaruddin Insyroh [4], Apolonia Diana Sherly da Costa [5, 6], Florence Jean B. Talirongan [7], Furizal [8], Asno Azzawagama Firdaus [9]

[1,2,3] Department of Computer Science, Universitas Qamarul Huda Badaruddin, Central Lombok 83562, Indonesia
[3] Department of Information System, Universitas Qamarul Huda Badaruddin, Central Lombok 83562, Indonesia
[4] Department of Law and Digital Technologies, Leiden University, Leiden, Netherland
[5] Guest Scientist of Interdisciplinary Natural Disaster Risk Management at the Institute for Landscape Ecology, University of Münster, Münster, Germany
[6] Geography Study and Spatial Planning, Department of Geography, University of Porto, Portugal
[7] College of Computer Studies, Misamis University, Ozamis City 7200, Philippines
[8] Department of Research and Development, Peneliti Teknologi Teknik Indonesia, Sleman 55281, Indonesia

ARTICLE INFO

Article history:

Received October 25, 2024
Revised November 15, 2024
Published January 31, 2025

Keywords:

decision tree; machine learning; naïve bayes; stunting;

ABSTRACT

Child stunting is a health problem that has a major impact on their physical growth and brain development. This study aims to create a model that can predict the risk of stunting using machine learning technology, in order to provide assistance quickly. Using data from 7,573 children, which included information such as age, weight, height gender and breastfeeding status, we tried two methods, Naive Bayes and Decision Tree. As a result, Naive Bayes was more accurate and the success rate reached 92%, compared to Decision tree which was only 88%. With this model, it is hoped that health workers will find it easier to find children at risk of stunting, so that preventive action can be taken earlier. This research aims to provide technology-based solutions to overcome the problem of stunting in the community.

Corresponding Author:

Adrian Maulana, Department of Computer Science, Universitas Qamarul Huda Badaruddin, Central Lombok, Indonesia
Email: adrianreey88@gmail.com

1. INTRODUCTION

Stunting is a disorder of health or growth and development of children due to malnutrition and repeated infections characterized by length or height below the standard. Where the examination of stunting status in toddlers is vulnerable to health and is done manually, so it takes a long time. It is hoped that a system that can help with the classification process so that it can detect toddlers who are stunted or not with a fast and accurate process. There are several factors that can cause stunting, such as poor infant care, the role of a mother who does not understand the nutritional needs of infants, limited health services, lack of family access to provide nutritious food, and lack of access to clean water [1].

Nutrient deficiencies in fetuses and young children can significantly affect their development, so it is important to ensure adequate nutrition to support healthy growth to avoid potential impediments to cognitive, motor, and physical development, as well as long-term impacts on an individual's health and quality of life. One of the most important factors in children's growth is adequate nutrition, especially the 4 important stunting-

preventing nutrients needed by toddlers, exclusive breastfeeding, age of complementary feeding, history of infectious diseases and genetic factors [2].

Stunting describes a chronic undernutrition status during growth and development from the beginning of life. This situation is represented by a height-for-age (TB/U) z-score of less than -2 standard deviations (SD) based on WHO growth standards (MOH, 2014). Globally, about 1 in 4 children under five are stunted. Malnutrition at an early age increases infant and child mortality rates, causes sufferers to get sick easily and have poor posture as adults. Their cognitive abilities are also reduced, resulting in long-term economic losses for Indonesia. The incidence of stunting in toddlers is more common among toddlers aged 12-59 months than among toddlers aged 0-24 months [3].

Studies show that stunting is a serious problem in Indonesia as it affects around 30.8% of children under the age of five, nearly eight million children in 2018 (Ilman & Wibisono, 2019). Many mothers and caregivers have a limited understanding of stunting, including its causes and prevention (Syam et al., 2022). Triggering factors for stunting involve poor maternal nutrition, inadequate breastfeeding, inappropriate complementary feeding, and lack of hygiene and sanitation [4].

Data Mining is a term used to describe knowledge discovery in databases. And in this research using Naive Bayes and Decision tree algorithms [5]. Naive Bayes has strong intuition between features whereas Decision Tree is easy to interpret and intuitive [6]. The Naive Bayes classifier operates on the principle of Bayes' theorem and falls under the category of supervised learning. It assumes that the presence (or absence) of one feature is independent of the presence (or absence) of other features within a class [7]. This method is very simple and fast because it assumes that document features are independent of each other.

2. METHODS

In this research there are several processes in data collection that are carried out. Here is the process of the process.

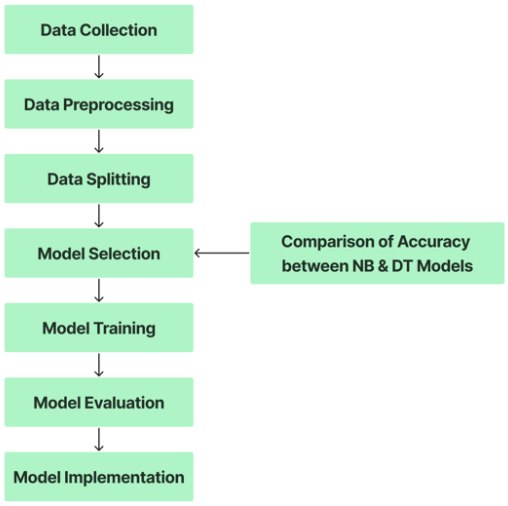

Fig. 1. Stage workflow research

Fig 1 the stages of a machine learning workflow, beginning with Data Collection, followed by Data Preprocessing to clean and prepare the data. Next, the data is split into training and testing sets in the Data Splitting stage. At the Model Selection step, a comparison of accuracy between Naïve Bayes (NB) and Decision Tree (DT) models is conducted to determine the best-performing model. The chosen model then undergoes Model Training, followed by Model Evaluation to assess its performance. Finally, the model is deployed in the Model Implementation phase for real-world use.

2.1. Data Collecting

The data collection process starts with finding a dataset via Kaggle [8] and checking the dataset description to see the amount of data and the number of samples in the dataset. My data consists of 8 columns and 10,000 rows ranging from infants to toddlers.

2.2. Cleaning

The data undergoes preprocessing steps to ensure it is cleaned and prepared [9]. Preprocessing eliminates unnecessary characters and words that are irrelevant to the document. This stage simplifies and enhances text processing [10]. Cleaning carried out in the preprocessing stage is the process of removing the same data (duplicates), checking for data inconsistencies and data that has errors will be corrected and removing data that does not have a value in one of its attributes. The following is a graph after the data cleaning process:

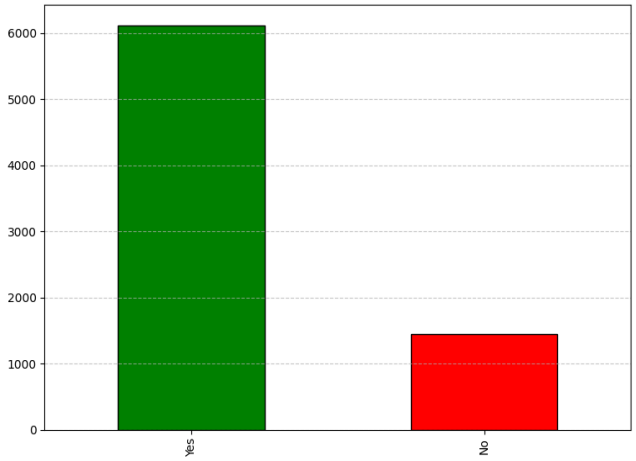

Fig. 2. Label distribution of stunting

2.3. Data Split

Data that has passed the preprocessing stage is then divided into two parts, namely model training (train) and model testing (testing). The data to be used is usually taken 80% for the model training process and the rest is made for testing data.

Table 1. Split data

Training (80%)	Testing (20%)
6.060 Data	1.513

2.4. Model Selection

At this stage, the model chosen is the Machine Learning algorithm that suits the classification problem to be solved. The models I chose were Naive Bayes, and Decision tree. And in this problem based on the data that becomes my label is the Stunting column which requires Binary Classification (yes / no).

2.5. Model Training

Model training is the process by which machine learning algorithms are trained using training set data to learn relevant patterns and build predictive models. In your dataset, features such as Age, Birth Weight, Body Weight, and Breastfeeding can be selected as relevant for the prediction of Stunting status.

Examples of some training data.

Table 2. Data training example

Age	Birth Weight	Body Weight	Breastfeeding	Stunting
17	3.0	10.0	No	No
11	2.9	2.9	No	Yes

2.6. Model Evaluation

After all the processes are completed, the data will be processed and calculated the level of precision, recall, f1 score, and support using the naive bayes and decision tree methods so that it can produce accuracy between the 2 models.

a. Naive Bayes model

Naive Bayes is a probabilistic machine learning algorithm based on Bayes' Theorem, frequently used for classification tasks. It demonstrates a lower error rate compared to many other classifier algorithms [11]. Consider the Naive Bayes formula as follows:

$$P(v \text{ I } a) = P(v)\, P(a \text{ I } v)$$

Description:
v = target label (class)
a = word with unknown class
The probability P(a I v) describes the likelihood of event v occurring if it is known that event a has occurred. P(v), called the prior, is the base probability of the occurrence of a particular target label value without considering the value of the feature (predictor variable). Meanwhile, P(a I v) known as the likelihood, is the probability of a certain feature value occurring when the target label value is known. Finally, P(a), known as evidence, is the probability of occurrence of a set of feature values (predictor variables) [12].

b. Decision Tree

Decision tree is one of the simple and interesting classification methods. This method uses a tree structure, where each node represents a decision based on an attribute, and the branches show the results of the decision. This process starts from the root of the tree and continues through the branches until it reaches the final node (leaf), which gives the final result or decision [13].

To calculate the matrix of the naive bayes and Decision Tree model, the formula as below is needed [14][15]:

$$Accuracy = True\ Positive + \left(\frac{True\ Negative}{True\ Positive}\right) False\ Positive + False\ Negative + True\ Negative$$

$$F1 - Score = 2\ x\ Precision \frac{Recall}{Precision} + Recall$$

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

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

2.7. Model Implementation

This research can be made as a mobile application or web platform that is easily used by health workers. They simply enter the child's data, and the prediction results will immediately appear. All results will be displayed in the form of a simple dashboard, making it easier for users to see data patterns and trends clearly. However, it needs a little improvement and refinement so that the application can run well

3. RESULTS AND DISCUSSION

3.1. Naïve Bayes Classification

The confusion matrix generated by the Naive Bayes model shows the classification performance on two classes, namely Class 0 and Class 1. Based on the matrix, the model successfully predicted correctly 109 data for Class 0 (True Negative) and 1085 data for Class 1 (True Positive). However, the model also made mistakes by predicting 159 data from Class 0 as Class 1 (False Positive) and 162 data from Class 1 as Class 0 (False Negative). This shows that The model is superior in classifying data in Class 1 compared to Class 0, which can be seen from the high number of True Positive compared to False Negative.

Table 3. Performance evaluation table Naïve Bayes

Stunting	Precision	Recall	F1-Score	Support
No	0.40	0.41	0.41	268
Yes	0.87	0.87	0.87	1,247

Overall, the accuracy of the model was calculated by dividing the total correct predictions (True Positive + True Negative) by all the data tested. This model has an accuracy of 78.8%. Although the accuracy is good, there is still a significant amount of error in the predictions, especially for Class 0. Therefore, additional analysis using other metrics such as precision, recall, or F1-score is needed to evaluate the model's performance more thoroughly.

3.2. Decision Tree Classification

Confusion matrix from the Decision Tree model shows the classification results for two classes, namely Class 0 and Class 1. In Class 0, the model is able to correctly predict 96 data (True Negative), while for Class 1 there are 960 data predicted correctly (True Positive). However, there are some errors, namely 172 data from Class 0 predicted as Class 1 (False Positive) and 287 data from Class 1 predicted as Class 0 (False Negative). This shows that the model is more reliable in classifying data in Class 1 than Class 0, although there is still a weakness in detecting Class 0 well.

Table 3. Performance evaluation table Decision Tree

Stunting	Precision	Recall	F1-Score	Support
No	0.26	0.36	0.30	268
Yes	0.79	0.77	0.81	1,247

The accuracy of the model is calculated by dividing the number of correct predictions (True Positive + True Negative) by the total data tested. From a total of 1515 data, the model successfully predicted 1056 data correctly, resulting in an accuracy of 69.7%

Based on the two results above, the accuracy of the two models above can be classified using the bar chart method. The bar graph is one of the visual presentations of data that can make it easier for anyone who reads, to be able to understand the meaning or results of the data presented on the graph shows an image in the form of a bar and is used to see the comparison of the data.

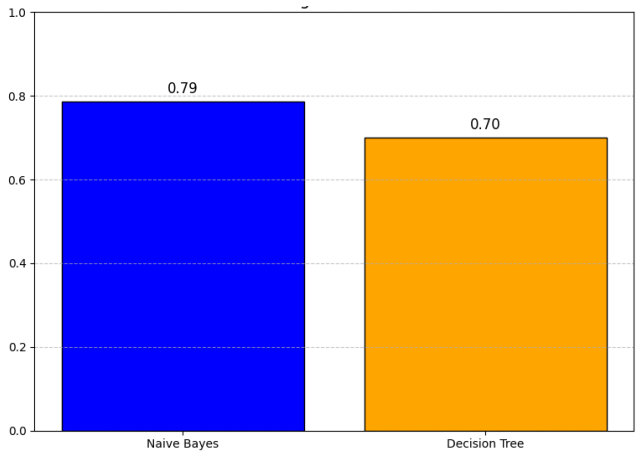

Fig. 3. Comparison model accuracy

Figure 3 shows in the performance of two classification algorithms, Naïve Bayes and Decision Tree, based on a specific evaluation metric, likely accuracy. The Naïve Bayes algorithm achieves a performance score of 0.79, represented by a blue-colored bar, while the Decision Tree algorithm scores 0.70, shown with an orange-

colored bar. This indicates that Naïve Bayes performs better than Decision Tree in this evaluation. The chart provides a clear visual comparison, highlighting the difference in effectiveness between the two models.

4. CONCLUSION

From the evaluation results, the Naive Bayes model has an accuracy rate of 78.8%, while the Decision Tree model shows an accuracy of 69.7%. This shows that Naive Bayes is superior in predicting the overall data with higher accuracy and lower number of errors than Decision Tree. However, Naive Bayes still shows weakness in detecting Class 0, as seen from the high number of False Positive (159). Meanwhile, the Decision Tree model suffers from a greater weakness, particularly in the number of False Negatives (287), which shows the model's difficulty in accurately classifying Class 1. These shortcomings caused the overall performance of Decision Tree to be less adequate than Naive Bayes. Therefore, for this data, the Naive Bayes model is recommended due to its better performance in classifying both classes.

REFERENCES

[1] M. Yunus, M. K. Biddinika, dan A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *J. Online Inform.*, vol. 8, no. 1, hal. 99–106, Jun 2023, doi: 10.15575/join.v8i1.1062.

[2] T. Hardiani and R. N. Putri, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita," Digit. Transform. Technol., vol. 4, no. 1, pp. 621–627, 2024, doi: 10.47709/digitech.v4i1.4481.

[3] W. C. Wahyudin, F. M. Hana, and A. Prihandono, "Prediksi Stunting Pada Balita Di Rumah Sakit Kota Semarang Menggunakan Naive Bayes," J. Ilmu Komput. dan Matemtika, vol. 2019, pp. 32–36, 2023.

[4] F. A. Sany, "Penerapan Sistem Pakar untuk Deteksi Stunting," J. Ilm. Ecosyst., vol. 23, no. 3, pp. 602–609, 2023, doi: 10.35965/eco.v23i3.3774.

[5] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, "Penerapan Data Mining Untuk Analisis Stunting Pada Balita," vol. 13, no. November, pp. 1717–1728, 2024.

[6] Asno Azzawagama Firdaus *et al.*, "Application of Sentiment Analysis as an Innovative Approach to Policy Making: A review," *Journal of Robotics and Control (JRC)*, vol. 5, no. No. 6, pp. 1784–1798, 2024.

[7] M. M. Dakwah, A. A. Firdaus, Furizal, and R. A. Faresta, "Sentiment Analysis on Marketplace in Indonesia using Support Vector Machine and Naïve Bayes Method," Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), vol. 10, no. 1, pp. 39–53, 2023, doi: 10.26555/jiteki.v10i1.28070.

[8] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," Infotek J. Inform. dan Teknol., vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.

[9] A. A. Firdaus, A. Yudhana, and I. Riadi, "Indonesian presidential election sentiment: Dataset of response public before 2024," *Data Brief*, vol. 52, p. 109993, 2024, doi: 10.1016/j.dib.2023.109993.

[10] A. A. Firdaus, A. Yudhana, and I. Riadi, "Public Opinion Analysis of Presidential Candidate Using Naïve Bayes Method," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, no. 2, pp. 563–570, May 2023, doi: 10.22219/kinetik.v8i2.1686.

[11] A. A. Firdaus, A. Yudhana, and I. Riadi, "Prediction of Presidential Election Results using Sentiment Analysis with Pre and Post Candidate Registration Data," 2024, doi: 10.23917/khif.v10i1.4837.

[12] F. R. Yuliardi, F. Fauzi, and T. W. Utami, "Analisis Sentimen Opini Masyarakat Terhadap Stunting Pasca Debat Cawapres Pertama 2024 Dengan Algoritma Bootstrap Aggregating Naïve Bayes," pp. 1129–1139, 2024.

[13] P. Meilina, "Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi," J. Teknol. Univ. Muhammadiyah Jakarta, vol. 7, no. 1, pp. 11–20, 2015, [Online]. Available: jurnal.ftumj.ac.id/index.php/jurtek.

[14] S. S. Berutu, H. Budiati, J. Jatmika, and F. Gulo, "Data preprocessing approach for machine learning-based sentiment classification," J. Infotel, vol. 15, no. 4, pp. 317–325, 2023, doi: 10.20895/infotel.v15i4.1030.

[15] A. Azzawagama Firdaus, A. Yudhana, and I. Riadi, "Prediction of Indonesian Presidential Election Results using Sentiment Analysis with Naïve Bayes Method," Jurnal Media Informatika Budidarma, vol. 8, no. 1, pp. 41–50, 2024, doi: 10.30865/mib.v8i1.7007.