



Nutritional Status Classification Of Stunting In Toddlers Using Naive Bayes Classifier Method

Risky Devandra Hartana

Faculty of Science & Technology, Yogyakarta University of Technology

Enny Itje Sela

Faculty of Science & Technology, Yogyakarta University of Technology

Corresponding author: Riskydevandra@gmail.com

ABSTRACT. Stunting in toddlers is one of the prevalent issues of malnutrition in Indonesia. The causes of Stunting are diverse, and one contributing factor is the insufficient nutritional intake required for toddlers. The categorization of Stunting nutritional status in toddlers is crucial to identify those experiencing Stunting, enabling appropriate interventions to prevent more serious health problems in the future. This research aims to develop a classification model for short nutritional status in toddlers using the Naive Bayes Classifier method. The data utilized in this study originate from anthropometric measurements of toddlers in the Malebo area, Kandangan, Temanggung, Central Java. The anthropometric data include weight, height, and age of the toddlers. This data is then processed using the Naive Bayes Classifier method to classify the nutritional status of Stunting in toddlers. The results of this research are expected to assist in identifying toddlers experiencing Stunting, facilitating appropriate interventions to prevent more serious health issues in the future. Additionally, the Naive Bayes Classifier method employed can be applied in similar studies to enhance the quality of life, especially for children in Indonesia, particularly in the Malebo area, Kandangan, Temanggung, Central Java.

Keywords: Stunting, Nutritional Status, Classification, Intervention, Anthropometric Measurements.

INTRODUCTION

Stunting is a serious public health issue in Indonesia, particularly among children in rural and urban areas. According to data from the Badan Pusat Statistik (BPS) in 2020, the prevalence of stunting in Indonesia remains high at around 27.7% among children aged 0-59 months. The Central Java province itself has a relatively high prevalence of stunting, approximately 25.3% among children aged 0-59 months.

Malebo Village, Kandangan, Temanggung is a sub-district in the Temanggung Regency, Central Java, which still faces health issues related to stunting, as reported in the (Kebijakan, Kesehatan, dan Ri 2023). Therefore, research is needed to identify the factors influencing the nutritional status of stunting in toddlers in that area.

Based on previous studies, several research projects have explored contributing factors to the nutritional status of stunting in toddlers in various regions of Indonesia. A study by Rahmadhita found that factors such as low maternal education, low socio-economic status, inadequate feeding patterns, and poor sanitation are associated with an increased risk of stunting in children. (Rahmadhita 2020).

Furthermore, research conducted in rural areas of Central Java indicates that factors such as maternal nutritional status, complementary feeding practices, and household sanitation

Received: February 21, 2024; Accepted: March 19, 2024; Published: April 30, 2024

* Risky Devandra Hartana: Riskydevandra@gmail.com

influence the nutritional status of stunting =in toddlers (Ridho Nugroho, Nur Sasongko, dan Kristiawan 2021). Similar findings have been reported by other studies in urban areas, emphasizing the importance of these factors in addressing the issue of stunting (Putri dkk. t.t.).

Several studies related to the classification of stunting nutritional status include research conducted by Syahrani Lonang and Dwi Normawati titled “Klasifikasi Status *Stunting* Pada Balita Menggunakan *K-Nearest Neighbor* Dengan *Feature Selection Backward Elimination*”. This study, based on 1000 data points with 9 attributes, including 243 stunted infants and 757 non-stunted toddlers, achieved an accuracy of 91.90% using the K-Nearest Neighbor method with 9 attributes. The addition of Backward Elimination increased accuracy to 92.20% with 8 attributes (Lonang dan Normawati 2022).

Another study by Muhamad Amirudin and Alz Danny Wowor titled “Analisis Perbandingan Klasifikasi Balita Beresiko *Stunting* Menggunakan Metode *Support Vector Machine* dan *Decission Tree*”. The study used 10,000 data points from kaggle.com with 6 attributes, the study compared the Support Vector Machine and Decision Tree methods, resulting in 83% accuracy for Support Vector Machine and 78% for Decision Tree. The study claimed that the Support Vector Machine method is effective in analyzing the risk of stunting in toddlers (Amirudin dan Wowor 2023).

Lastly, a study by Dyta Kresna Devi Damayanti and Muhammad Jakfar titled “Klasifikasi Status *Stunting* Balita Menggunakan Algoritma *Fuzzy C-Means* (Studi Kasus Posyandu Rw 01 Kelurahan Jepara Surabaya)”. The study used 134 data points with 3 attributes, creating 3 clusters: severe stunting, stunting, and normal. The study achieved 50% accuracy, with 6 toddlers in the severe stunting cluster, 56 in the stunting cluster, and 72 in the normal cluster (Kresna, Damayanti, dan Jakfar 2023).

Building on these previous studies, the current research aims to classify stunting nutritional status using the Naïve Bayes Classifier method. By deepening our understanding of contributing factors to stunting and combining them with the Naive Bayes classification method, this research may provide new insights into addressing stunting issues in Indonesian children. It is hoped that the results of this research can serve as a foundation for the development of more effective and specific interventions to reduce the prevalence of stunting in toddlers, particularly in the Malebo Village, Kandangan, Temanggung area, contributing to the overall improvement of the quality of life and well-being of Indonesian children.

RESEARCH METHODS

The research stages represent the process carried out throughout the research. The study is conducted in phases, starting from data collection to classification using the Naïve Bayes Classifier. The following are the stages or steps of the research.

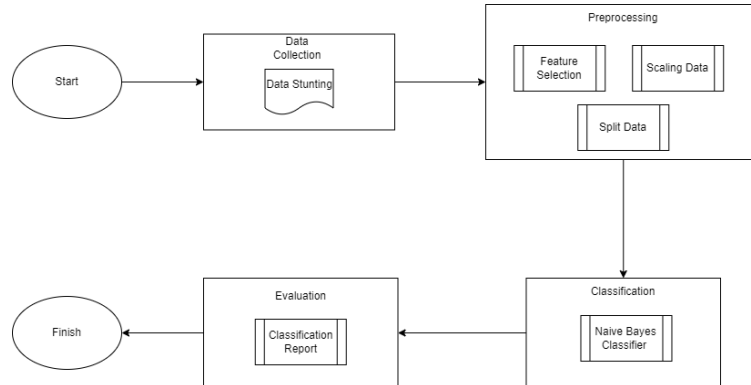


Figure 1 Research Stages

Stunting

Stunting is a condition of growth failure that occurs in toddlers as a result of chronic nutritional deficiencies and recurrent infections, particularly during the first 1000 days of life (Rozaq dan Purnomo 2022). The occurrence of stunting in both toddlers and children can be attributed to several triggering factors, such as maternal nutrition during pregnancy, infant illnesses, and inadequate nutritional intake for the infant (Yuwanti, Himawati, dan Mulya Susanti 2022). A mother plays a crucial role in determining stunting development, considering that a child's early development begins during pregnancy.

DATA COLLECTION

The data used in this research consists of nutritional information on toddlers obtained from the Malebo Community Health Center, Temanggung, Central Java. The data collection period spans from January 16, 2023, to March 16, 2023, and comprises 182 rows. The dataset features include age, Child_order, Income, Occupation, Education, gender, and Label. An example of stunting nutritional status data can be observed in **Table 1**.

Table 1 Stunting Nutrition Status Data

Age	Gender	Education	Occupation	Income	Child_order	Label
22	Perempuan	Kuliah	Karyawan	Rp. 2,000,000 - Rp. 4,999,999	3	Normal
20	Perempuan	Kuliah	Wirausaha/Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	4	Normal
18	Perempuan	SMA	Wirausaha/Wiraswasta	Rp. 1,000,000 - Rp. 1,999,999	3	Normal
17	Laki-laki	Kuliah	Lainnya	Rp. 2,000,000 - Rp. 4,999,999	2	Normal
18	Laki-laki	SD	Petani/Buruh	Rp. 2,000,000 - Rp. 4,999,999	3	Normal
16	Laki-laki	Tidak Sekolah	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	1	Normal
16	Laki-laki	Kuliah	Wirausaha/Wiraswasta	Rp. 2,000,000 - Rp. 4,999,999	2	Normal
15	Laki-laki	Tidak Sekolah	Lainnya	Rp. 1,000,000 - Rp. 1,999,999	2	Normal
15	Perempuan	Kuliah	Lainnya	Rp. 2,000,000 - Rp. 4,999,999	1	Normal
18	Perempuan	SMP	Karyawan	Rp. 1,000,000 - Rp. 1,999,999	2	Normal

Feature Selection

Feature selection is the process of choosing a subset of relevant features to be used in model construction (AlNuaimi dkk. 2022). The primary goal of feature selection is to identify and retain the most significant and informative features while removing redundant, irrelevant, or less important ones (Budiman, Sunyoto, dan Nasiri 2021). This helps enhance the efficiency and effectiveness of machine learning models by reducing dimensionality, decreasing computational complexity, and often improving the model's generalization and interpretability.

Feature Scaling

Feature scaling is a crucial step in building machine learning models. One key aspect of feature scaling involves scaling, normalization, and standardization, which entails transforming data to make it more suitable for machine learning modeling (Ozsahin dkk. 2022).

Data Splitting

Data splitting is the process of dividing a dataset into two or more parts, typically used for training and testing a model. The data is divided into a training set, used to train the model, and a test set, used to assess the model's performance (Afifah 2022).

Naïve Bayes Classifier (NBC)

The Naive Bayes Classifier is one of the Supervised Learning algorithms for classification, utilizing a probability-based approach that calculates the likelihood values of each attribute to achieve effective results in a short amount of time (Setyanto dan Sela 2023). This method assumes that all attributes in the data are conditionally independent (Ardiyanto

dan Sela 2024), although this assumption is often not met in the real world. Nevertheless, Naive Bayes Classifier remains a popular and effective method for various classification problems. Naive Bayes can be calculated using Equation (1).

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (1)$$

Description:

$P(C_k|X)$ = the probability that data with attributes X belongs to the target class Ck. In the context of this research, the target class can be "Stunting" or "Normal".

$P(X|C_k)$ = the probability that data with attributes X given the target class Ck. This assumes that each attribute is conditionally independent of the target class. In this case, attributes can include age, height, and weight.

$P(C_k)$ = the prior probability of the target class Ck. This is the probability of the occurrence of the target class Ck in the training data.

$P(X)$ = the probability of data X, i.e., the probability of the occurrence of data X in the training data.

Gaussian Naïve Bayes Model

The Gaussian Naive Bayes Model is one variation of the Naive Bayes Classifier used when the attributes in the data are continuous and assumed to follow a normal (Gaussian) distribution (Irawan Saputra dan Hakim 2022). This method is beneficial for classifying data with numerical attributes, as is the case in this research involving attributes such as age, Child_order, Income, Occupation, Education, and gender. The process can be computed using Equation (2).

$$P(X|C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(X-\mu_k)^2}{2\sigma_k^2}} \quad (2)$$

Description:

$P(X|C_k)$ = the probability of numerical attribute X given the target class Ck.

μ_k = the mean value of attribute X in the target class Ck.

σ_k = the standard deviation of attribute X in the target class Ck.

e = Euler's constant, and π is the constant pi.

Classification Report

A classification report is a document that presents the performance evaluation of a classification model based on the information found in the confusion matrix (Raihan, Allaam, dan Wibowo 2021). In this context, the classification report provides a detailed overview of

the model's ability to effectively categorize each class. Information within the classification report includes precision, recall, f1-score, and support (AMAN KHARWAL 2021).

Precision is a parameter that measures the extent to which a class is predicted as positive by the model and how many of them are truly positive (Jefriyanto dkk. 2023). The calculation of precision can be performed using Equation (3).

$$precision = \frac{TP}{(TP + FP)} \tag{3}$$

Recall is a parameter that measures the extent to which the model can find all classes that should be positive (Sheth, Tripathi, dan Sharma 2022). The calculation of recall can be performed using Equation (4).

$$recall = \frac{TP}{(TP + FN)} \tag{4}$$

F1-score is the average of precision and recall values (Chicco dan Jurman 2020). The calculation of the f1-score can be done using Equation (5).

$$f1 - score = 2 * \frac{recall * precision}{recall + precision} \tag{5}$$

RESULTS AND DISCUSSION

This testing was conducted using Gaussian Naive Bayes with a dataset size of 182, and the dataset was divided into 70% training data and 30% testing data, utilizing 7 attributes that have been transformed into numerical form. The following are the results of the Gaussian Naive Bayes classifier model.

Akurasi dan Classification Report					
Akurasi: 0.8727272727272727					
Classification Report:					
precision	recall	f1-score	support		
	0	0.95	0.88	0.91	42
	1	0.69	0.85	0.76	13
accuracy				0.87	55
macro avg		0.82	0.86	0.84	55
weighted avg		0.89	0.87	0.88	55

Figure 2 Classification Report

Based on **Figure 2**, the highest precision accuracy is found in the stunting category (1), amounting to 95%, and the lowest in the normal category (0), which is 69%. Then, for recall accuracy, the highest is in the normal category (0) at 88%, and for stunting (1), it is neutral at 85%. Furthermore, for the f1-score accuracy, the highest is in the normal category (0) at 91%,

and the lowest accuracy is in stunting (1) at 76%. The overall accuracy result for the entire model is 87%.

Implementation of the Program

a. Initial System Interface

On the system's initial page, users can input stunting data that will be classified by the system. The system will automatically perform preprocessing and classification. The data used adheres to specified attributes, including age, gender, education, occupation, income, and child_order.

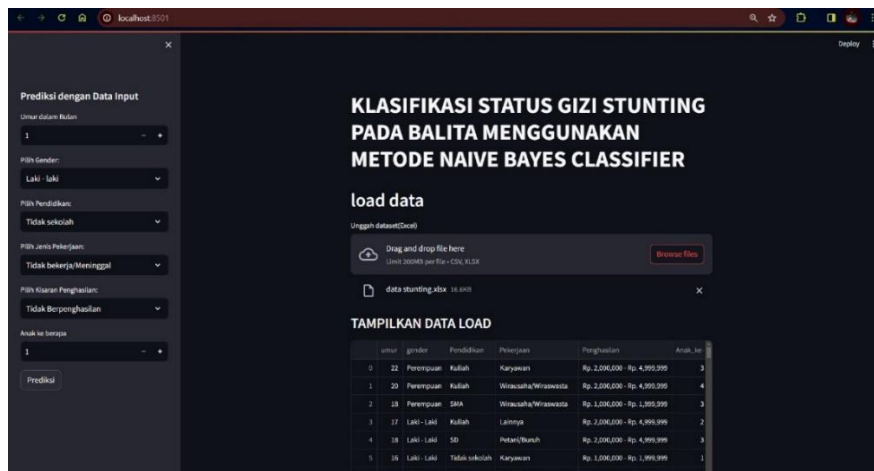


Figure 3 Initial System Interface

b. Preprocessing Display

In the preprocessing display, users can view data that has been preprocessed by the system. The preprocessing steps include label encoding, removing duplicate data, and filling missing values with the mean value.

Data Hasil Preprocessing

	umur	gender	Pendidikan	Pekerjaan	Penghasilan	Anak_ke-	Label
0	22	0	4	2	4	3	C
1	20	0	4	3	4	4	C
2	18	0	3	3	3	3	C
3	17	1	4	5	4	2	C
4	18	1	1	1	4	3	C
5	16	1	0	2	3	1	C
6	16	1	4	3	4	2	C
7	15	1	0	5	3	2	C
8	15	0	4	5	4	1	C
9	18	0	2	2	3	2	C

Figure 4 Preprocessing Display

c. Prediction Display

In the prediction display, users can input data by filling out the provided form. The user-inputted data will be classified by the Naïve Bayes model that has been created, and the result will be displayed to indicate whether the inputted data falls into the stunting or normal category.

Figure 5 Prediction Display

CONCLUSION

Based on the analysis conducted, this system can provide quite satisfactory results in the Classification of Nutritional Status of Stunting in Toddlers Using the Naïve Bayes Classifier method. The accuracy level of the testing data obtained from the best model, with a

data split ratio of 70%:30%, is 87.27% in classification. This accuracy demonstrates the model's ability to provide relatively good predictions.

To enhance the performance of the Naïve Bayes Classifier in predicting nutritional status in toddlers, the following recommendations can be considered. Firstly, the use of a larger dataset can significantly improve the accuracy of classifying the nutritional status of stunting in toddlers. Secondly, adjusting parameters in the Naïve Bayes model can be explored to improve accuracy and achieve better classification results. It is crucial to experiment with training using various parameter settings. Finally, it is recommended to use a computer device with high GPU capacity, as this can impact the training process, especially with larger datasets.

REFERENCES

- Afifah, Lutfia. 2022. "Evaluasi Model Machine Learning: Train/Test Split." *ilmudatapy.com*.
- AlNuaimi, Noura, Mohammad Mehedy Masud, Mohamed Adel Serhani, dan Nazar Zaki. 2022. "Streaming feature selection algorithms for big data: A survey." *Applied Computing and Informatics* 18(1–2). doi: 10.1016/j.aci.2019.01.001.
- AMAN KHARWAL. 2021. "Classification Report in Machine Learning." *The Clever Programmer*.
- Ardiyanto, Agus, dan Enny Itje Sela. 2024. "Analisis Sentimen Opini Mahasiswa Terhadap Aplikasi Portal Mahasiswa UTY Menggunakan Metode Naïve Bayes Classifier." *Jurnal Indonesia : Manajemen Informatika dan Komunikasi* 5(1):525–40. doi: 10.35870/JIMIK.V5I1.508.
- Budiman, Setiawan, Andi Sunyoto, dan Asro Nasiri. 2021. "Analisa Performa Penggunaan Feature Selection untuk Mendeteksi Intrusion Detection Systems dengan Algoritma Random Forest Classifier." *SISTEMASI* 10(3):754–60.
- Chicco, Davide, dan Giuseppe Jurman. 2020. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." *BMC Genomics* 21(1). doi: 10.1186/s12864-019-6413-7.
- Irawan Saputra, Dede, dan Dadang Lukman Hakim. 2022. "Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler." *EPSILON: Journal of Electrical Engineering and Information Technology* 20(2). doi: 10.55893/epsilon.v20i2.94.
- Jefriyanto, Jefriyanto, Nur Ainun, Muchamad Arif, dan Al Ardha. 2023. "Application of Naïve Bayes Classification to Analyze Performance Using Stopwords." *Journal of Information System, Technology and Engineering* 1(2):49–53. doi: 10.61487/JISTE.V1I2.15.
- Kebijakan, Badan, Pembangunan Kesehatan, dan Kementerian Kesehatan Ri. 2023. *Status Gizi SSGI 2022*.
- Kresna, Dyta, Devi Damayanti, dan Muhammad Jakfar. 2023. "KLASIFIKASI STATUS STUNTING BALITA MENGGUNAKAN ALGORITMA FUZZY C-MEANS (STUDI KASUS POSYANDU RW 01 KELURAHAN JEPARA SURABAYA)." *MATHunesa: Jurnal Ilmiah Matematika* 11(3):533–42. doi: 10.26740/MATHUNESA.V11N03.P524-533.

- Lonang, Syahrani, dan Dwi Normawati. 2022. "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination." *JURNAL MEDIA INFORMATIKA BUDIDARMA* 6(1):49–56. doi: 10.30865/MIB.V6I1.3312.
- Ozsahin, Dilber Uzun, Mubarak Taiwo Mustapha, Auwalu Saleh Mubarak, Zubaida Said Ameen, dan Berna Uzun. 2022. "Impact of feature scaling on machine learning models for the diagnosis of diabetes." dalam *Proceedings - 2022 International Conference on Artificial Intelligence in Everything, AIE 2022*.
- Amirudin, Muhamad, dan Alz Danny Wowor. 2023. "Analisis Perbandingan Klasifikasi Balita Beresiko Stunting Menggunakan Metode Support Vector Machine dan Decision Tree." *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media* 3(1):581–91.
- Putri, Raiy, Pratama Sari, Maria Montessori Prodi, Pendidikan Pancasila, dan Dan Kewarganegaraan. t.t. "Upaya Pemerintah Dan Masyarakat Dalam Mengatasi Masalah Stunting Pada Anak Balita." *129 Journal of Civic Education* 4(2):2021.
- Rahmadhita, Kinanti. 2020. "Permasalahan Stunting dan Pencegahannya." *Jurnal Ilmiah Kesehatan Sandi Husada* 11(1). doi: 10.35816/jiskh.v11i1.253.
- Raihan, M., Rafiiful Allaam, dan Agung Toto Wibowo. 2021. "Klasifikasi Genus Tanaman Anggrek Menggunakan Metode Convolutional Neural Network (CNN)." *e-Prceeding of Engineering* 8(2).
- Ridho Nugroho, Muhammad, Rambat Nur Sasongko, dan Muhammad Kristiawan. 2021. "Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia." 5(2):2269–76. doi: 10.31004/obsesi.v5i2.1169.
- Rozaq, Abdul, dan Ari Joko Purnomo. 2022. "View of CLASSIFICATION OF STUNTING STATUS IN TODDLERS USING NAIVE BAYES METHOD IN THE CITY OF MADIUN BASED ON WEBSITE." *Journal of Computing and Information Technology* 69–76. Diambil 16 Februari 2024 (<https://ejournal.nusamandiri.ac.id/index.php/techno/article/view/3337/951>).
- Setyanto, Riky Ananda, dan Enny Itje Sela. 2023. "Klasifikasi Kematangan Buah Ciplukan dengan Metode Naive Bayes dan Ekstraksi Fitur GLCM." *Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi* 12(3).
- Sheth, Vraj, Urvashi Tripathi, dan Ankit Sharma. 2022. "A Comparative Analysis of Machine Learning Algorithms for Classification Purpose." dalam *Procedia Computer Science*. Vol. 215.
- Yuwanti, Laily Himawati, dan Meity Mulya Susanti. 2022. "View of Pencegahan Stunting pada 1000 HPK." *Jurnal ABDIMAS-HIP Pengabdian Kepada Masyarakat* 35–39. Diambil 16 Februari 2024 (<https://akbidhipekalongan.ac.id/e-journal/index.php/abdimaaship/article/view/166/166>).