Received November 6th 2025; accepted December 19th 2025. Date of publication December 31st 2025
Digital Object Identifier: https://doi.org/10/25047/jtit.v12i2.452

Child Stunting Risk Analysis through Machine Learning Models using XGBoost Algorithm

NURUL RENANINGTIAS[1], ATIK PRIHATININGRUM[2], HARDIANSYAH[3], YUDI SETIAWAN[1,] ARIE VATRESIA[5]

[1]Program Studi Sistem Informasi, Fakultas Teknik, Universitas Bengkulu, Indonesia
[2]Program Studi Arsitektur, Fakultas Teknik, Universitas Bengkulu, Indonesia
[3]Program Studi Teknik Sipil, Fakultas Teknik, Universitas Bengkulu, Indonesia
[5]Program Studi Informatika, Fakultas Teknik, Universitas Bengkulu, Indonesia

CORESPONDING AUTHOR: NURUL RENANINGTIAS (nurulrenaningtias@unib.ac.id)

ABSTRACT Stunting is a chronic nutritional disorder that significantly affects child growth, development, and the overall quality of future human resources. According to the 2024 Indonesian Nutritional Status Survey (SSGI), the prevalence of stunting remains high at 19.8%, equivalent to approximately 4.48 million children under five. Early detection of stunting risk is essential for timely and data-driven interventions. This study employed the CRISP-DM methodology, encompassing business understanding, data collection, preparation, modeling, and evaluation phases. The dataset was processed through cleaning, variable encoding, and stunting status classification based on WHO standards. An XGBoost-based predictive model was developed and evaluated using accuracy, precision, recall, and F1-score metrics. The model achieved 98% accuracy in predicting stunting risk. Feature importance analysis revealed that height is the most influential variable determining stunting risk.

KEYWORDS: stunting, machine learning, risk prediction, XGBoost

I.INTRODUCTION

Stunting is a serious challenge to public health development in Indonesia, directly impacting the long-term quality of human resources (HR). Stunting is defined as a condition of growth failure in toddlers due to chronic malnutrition that persists from pregnancy to two years of age (the First 1000 Days of Life) [1]. The impact of stunting is not limited to childhood but can persist into adulthood, affecting productivity and overall quality of life [2]. Children who experience stunting are at risk of decreased cognitive abilities, low academic achievement, and weakened immunity to disease [3]. Furthermore, stunting can hamper national economic growth by reducing future human resource productivity [4].

According to the World Health Organization (WHO), the first 1.000 days of life, starting from conception until the age of two, is a critical period for growth and development. Malnutrition during this period can lead to stunting, a condition in which a child's height is below the standard for their age [5]. Stunting not only affects physical growth but can also impact cognitive development and long-term health outcomes [3]. Research has shown that children who experience malnutrition at an early age have a higher risk of becoming stunted [6]. Therefore, it is important to understand one of the key factors contributing to stunting, namely nutritional intake.

Stunting is a chronic nutritional problem that remains a global challenge, especially in developing countries. In Indonesia, according to data from the 2024 Indonesian Nutritional Status Survey (SSGI), the prevalence of stunting remains high, reaching 19.8%, equivalent to 4.482.340 children under five, although this figure has decreased compared to previous years. In Bengkulu Province, stunting remains a serious concern, particularly in areas with limited access to health services and nutrition information [7].

Although efforts to reduce stunting have been implemented in various regions of Indonesia through various nutrition and education interventions, the main challenge remains the ability to detect stunting risk quickly, accurately, and data-drivenly. Currently, measuring children's nutritional status is generally done manually and using simple descriptive statistics, requiring manual interpretation based on WHO anthropometric standards. This process is not only spend a lot of time but also carries the risk of misinterpretation, especially in areas with limited health personnel.

With the development of information technology and artificial intelligence, particularly machine learning, there is a significant opportunity

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

to increase the effectiveness and efficiency of the process of classifying and predicting stunting risk [8]. One algorithm that has proven superior in classification tasks with simple data structures and medium volumes is Extreme Gradient Boosting (XGBoost) [9]. XGBoost is a decision tree-based algorithm capable of handling various types of data, producing accurate predictions, and identifying the variables most influential on the predicted outcome [10][11].

This research will develop a predictive model capable of classifying children into four categories: severe stunting, stunted, normal, and tall. Furthermore, the model will provide a feature importance analysis, identifying the relative contribution of each variable (age, weight, height, and gender) to the classification. This research is important and strategic because it directly contributes to the development of an information technology-based early detection system for stunting, which is not only practical but also replicable and can be further developed.

.

II.METHOD

The development method used in this research is based on CRISP-DM (Cross-Industry Standard Process for Data Mining). Several phases applied in this research are Business Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation [12]

Stunting is a serious health problem in Indonesia, characterized by growth failure in children caused by various complex factors. According to data from the Ministry of Health, the prevalence of stunting remains quite high, especially in areas with low socioeconomic levels (Ministry of Health of the Republic of Indonesia, 2024). This condition has a significant impact on the quality of human resources, considering that stunting affects not only a child's physical growth but also their cognitive development and future potential. This research focuses on developing an early prediction method for stunting risk using a machine learning approach. The primary goal is to create an accurate and reliable predictive model to detect potential stunting in children as early as possible. Using the XGBoost algorithm, this study seeks to leverage the power of complex data analysis to identify risk factors contributing to stunting.

This research aims to develop a predictive model capable of analyzing various variables influencing stunting, providing a tool for health workers and policymakers to conduct early interventions for children at risk of stunting, and providing methodological contributions to the development of machine learning models for child health issues. Specifically, the research's technical requirements include comprehensive dataset collection, accurate data preprocessing, relevant feature selection, and the development of an XGBoost model with high accuracy and sensitivity.

The data required for this research consist of child nutritional status records representing the classification target, along with data on the child's height, sex, and age in months.

Data preprocessing was performed to clean and prepare the dataset before use for modeling. In this study, the data preparation stage encompassed several approaches to improve data quality and model accuracy.

1. Handling Missing Values, the first step in data preparation is addressing missing values in dataset variables, such as age, height, gender, and nutritional status. This data handling is crucial to ensure data completeness
2. Encoding Categorical Variables, encoding variables is performed to ensure optimal use of the variables in machine learning models. Categorical variables such as gender and nutritional status must be converted into numeric format for proper interpretation within the model.
3. Detecting and Handling Outliers, data containing outliers in variables such as height or age can disrupt the prediction process and reduce model accuracy. Values that differ significantly from the dataset are called outliers. Outliers are handled using z-score or interquartile range calculations.
4. Dataset Division, this stage divides the dataset into training and test data with an 80:20 split. The training data is used to build the model, while the test data is used to evaluate the model's performance on new data.
5. Oversampling or undersampling, to ensure the data being tested has the same ratio, an oversampling or undersampling step is required. Equalizing the data ratio helps generate accurate predictions. With an equal number of data classes, the model can better learn patterns to detect appropriate nutritional status targets based on the given variables.

The modeling process was then carried out using the XGBoost algorithm. The modeling process involved three stages: preprocessing, training, and testing. After the preprocessed data was trained, the results were evaluated to measure the model's reliability [13][14]. The steps in the XGBoost algorithm are as follows:

1. The average target value is calculated for the initial prediction and the corresponding initial residual error.
2. A model is trained with the independent variables and residual errors as data to obtain predictions.
3. Additive predictions and residual errors are calculated using multiple learning rates from

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

the previous output predictions obtained from the model.

4. Steps 2 and 3 are repeated several times until the required number of models are created.
5. The final prediction from boosting is the sum of all previous predictions made by the model.

After the model was obtained, it was evaluated using testing data, which was then evaluated against the weights obtained from the model training process. Validation was performed using cross-validation. Techniques such as cross-validation are used to avoid overfitting and ensure that the model performs well on previously unseen data [15]. Evaluation was also conducted using a confusion matrix and a classification report to measure the performance of a classification model [16]. A regression model was also run to identify which variables had the most significant influence on nutritional status [17].

The evaluation was conducted to test and measure the model's reliability. The results demonstrated the model's accuracy in predicting children's nutritional status, which is then useful in analyzing the risk of stunting in children. Feature importance from XGBoost was also used to assist this analysis. Feature Importance is used to analyze the importance of each feature in predicting nutritional status. This feature in XGBoost measures the relative contribution of each predictor variable to nutritional status. Using this feature, we can see the influence of each variable on the target in explaining variations in nutritional status.

III.RESULT AND DISCUSSION

This study used a stunting dataset obtained from the Kaggle platform . This dataset is the source of information used by researchers to analyze the risk of stunting in children. The dataset used has comprehensive and representative data coverage, including nutritional status records children representing the classification target, along with height data body, type gender, and age children in units month. In data understanding phase analyzed in the dataset include:

Age a numeric variable measured in months, indicates a child's developmental stage. This dataset contains a range of values for the age variable, from zero to 60 months. The age distribution of children in the dataset is as follows .

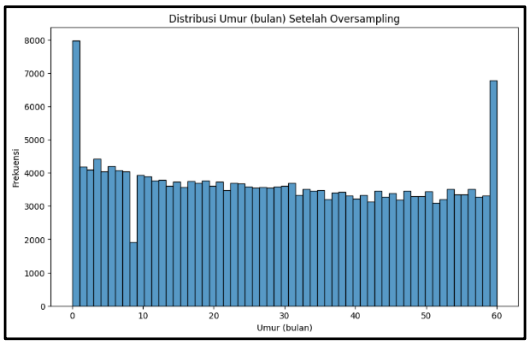

FIGURE 1. Distribution of age

Height a numeric variable that describes a child's physical size. The minimum value of the height variable is 40 cm and the maximum value is 128 cm. with an average (mean) height is 88.7cm. Height information in the dataset can be seen as follows .

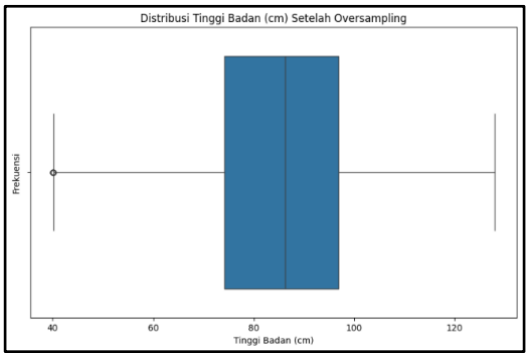

FIGURE 2. Boxplot of Height Distribution

Figure 2 shows the distribution of children's heights in the dataset. The boxplot above shows that the dominant distribution of children's heights is in the range of 75 to 97 cm.

Gender a categorical variable in the dataset, is a variable that can influence the risk of stunting. Gender distribution displayed on the following image .

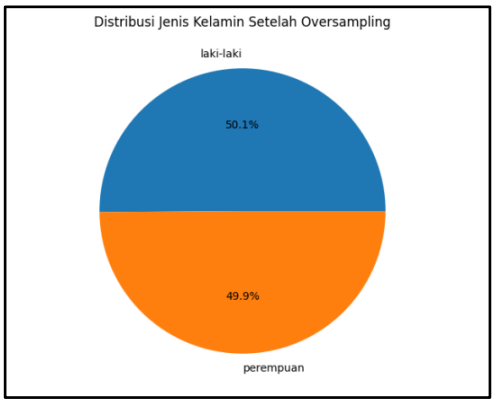

FIGURE 3. Distribution of Gender Dataset

Figure 3 shows that the gender variable is divided into two types: male and female . This gender variable contains data with a division of

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

50.1% of the variables containing male data and 49.9 % of the variables containing female data.

Nutritional Status is the target variable in this study. Nutritional status is a key variable in determining children's growth conditions. The frequency distribution for children's nutritional status in the dataset is as follows.

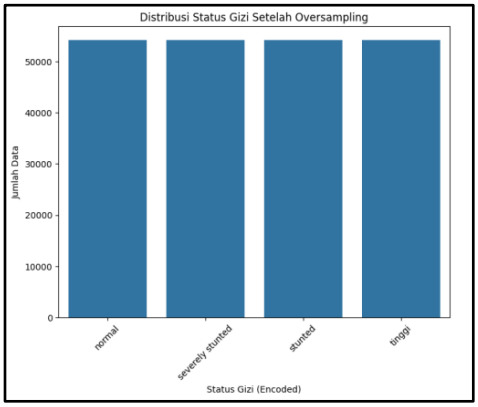

FIGURE 4. Frequency of Nutritional Status

Figure 4 shows that the dataset used predominantly contains nutritional status. normal , high nutrition, stunted nutrition and severely stunted. Futhermore the better understand the data used a comparison was made to see the relationship between each variable and the variables of height and nutritional status. The following is the relationship between these variables.

1. Relationship between Height and Age

The relationship between height and age is known to have a pattern in which the higher the age value, the higher the height value. However, some data fall outside this pattern. Correlation testing was conducted on these two variables, with a Pearson correlation coefficient of 0.8431 and a Spearman correlation coefficient of 0.8434.

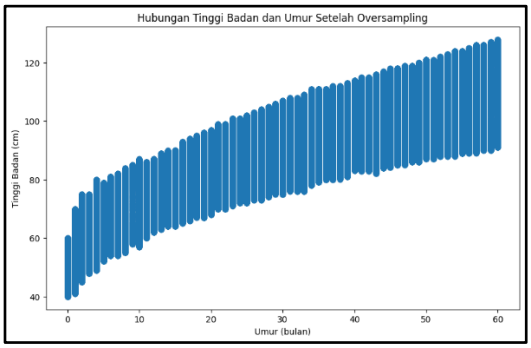

FIGURE 5. Height Relationship and Age

2. Height Distribution by Gender

The distribution of height by gender shows that the dominant height data ranges from 75 to 97 cm. Furthermore, females have a wider data distribution, reaching a top figure of 128 cm and a bottom figure closer to 40 cm compared to males.

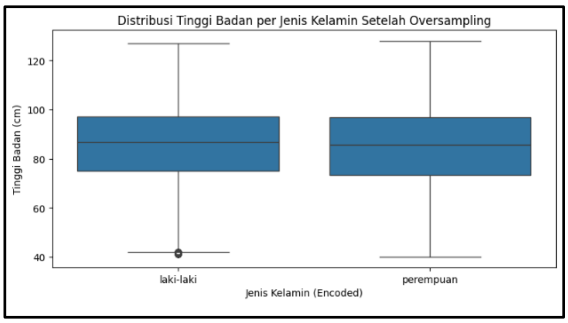

FIGURE 6. Distribution of Height per Gender

3. Distribution of Height per Nutritional Status

The height distribution graph for each nutritional status shows that each nutritional status has a distribution of data in order from lowest to highest, namely severely stunted, stunted, normal, and tall. In general, the dominant data distribution is in the height range of 60 to 110 cm.

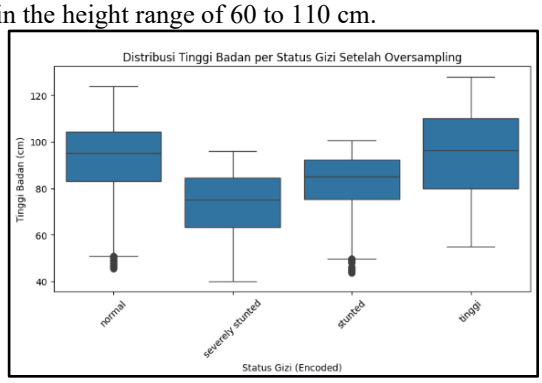

FIGURE 7. Distribution of Height per Nutritional Status

4. Age Distribution per Nutritional Status

Distribution of age per nutritional status, seen in general, the distribution of height is divided into two types. The first type, where lower height figures predominantly represent severely stunted or tall data. The second type, where higher height figures predominantly represent stunted and normal data.

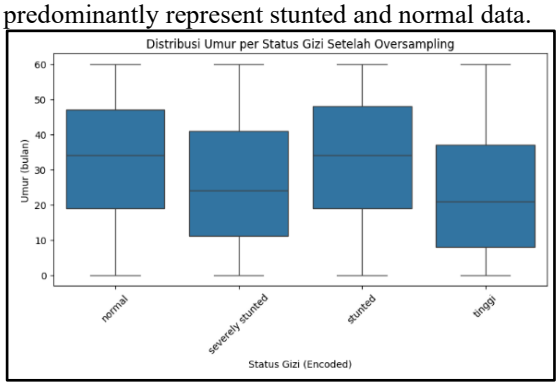

FIGURE 8. Age Distribution per Nutritional Status

5. Relationship between Gender and Nutritional Status

The relationship between gender and nutritional status and done Chi-square testing was performed on both variables.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

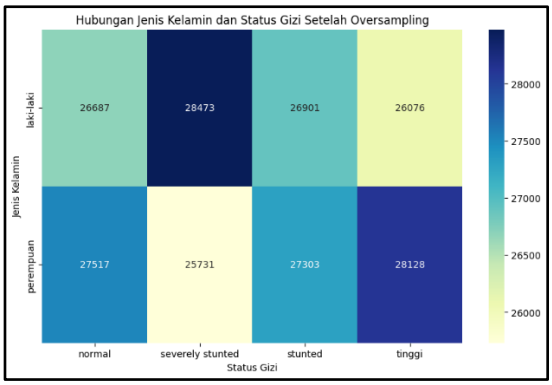

FIGURE 9. Relationship between Gender and Nutritional Status

In this study, data preparation stages were carried out for handling missing values, changes of categorical variables, detecting and handling outliers, dataset distribution, and undersampling.

For each categorical variable, coding will be carried out with the aim of ensuring that the variable can be read and then processed in a model that uses LabelEncoder. The values for the variables gender and nutritional status are converted into numbers according to Table 1 below.

TABLE 1. Changes variable value

Symbol	Before	After
Gender	Man	0
	Woman	1
Status Nutrition	Normal	0
	Severely Stunted	1
	Stunted	2
	Tall	3

Outliers are values that differ significantly from the existing variable data set. In this study, the outlier detection process was performed using the z-score and the Interquartile Range (IQR) method. These two methods were used to provide a more comprehensive and robust approach than using either method alone. This combination not only improves the accuracy of outlier detection but also provides deeper insight into the data, enabling better analysis and more informed decision-making. The results obtained from outlier detection are as follows:

TABLE 2. Outlier detection results

Methods	Variables	Value Description
Z-score	Tall body	Number of outliers 0
		Outliers percentage 0.00%
	Age	Number of outliers 0
		Outliers percentage 0.00%
IQR	Tall body	Number of outliers 38
		Outliers percentage 0.03%
	Age	Number of outliers 0
		Outliers percentage 0.00%

The dataset was divided into two subsets: training data and testing data . The split ratio was 80% for training and 20% for testing. The split was performed using the nutritional status variable as the target (y). The test variables (x) in this case, namely age, height and gender. After dividing the dataset into training and test data, it's necessary to evenly distribute the number of classes across nutritional status. This distribution ensures the model can effectively recognize each class.

In this study, modeling was carried out using the eXtreme Gradient Boosting (XGBoost) algorithm. The model was trained to identify data patterns and then make accurate predictions related with nutritional status child. The parameters used for configuration are model is learning_rate = 0.1 which controls the contribution of each tree, max_depth = 6 which limits the complexity of each tree, n_estimators = 200 which determines the number of trees to build the ensemble model, subsample = 0.8 means the model uses 80% of the data for each tree (preventing overfitting), objective = 'multi:softprob ' which produces probabilities for each class, and num_class which dynamically determines the number of classes in the dataset.

Evaluation was conducted to see the performance and accuracy of the XGBoost model. This stage is carried out evaluation use cross validation, confusion matrix, classification report and feature importance. the results of cross-validation carried out with 5 folds (k-fold), where each fold produces a different accuracy score. The average score (Mean CV Score) is 0.9899 and a very low standard deviation of 0.0005 shows excellent model performance with consistency across various data subsets and indicates model stability with almost the same accuracy score in each fold so that the conclusion is that the model is not overfitting or underfitting. The confusion matrix shows that the model has very good performance with most of the predictions being on the main diagonal, which indicates predicted and actual data match. This shows that most of the samples were classified correctly. Classification report model for child nutritional status classification demonstrated excellent and consistent performance across all categories. With an overall accuracy of 98%, it was found that The model is able to classify nutritional status with a high level of precision and recall for each class.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

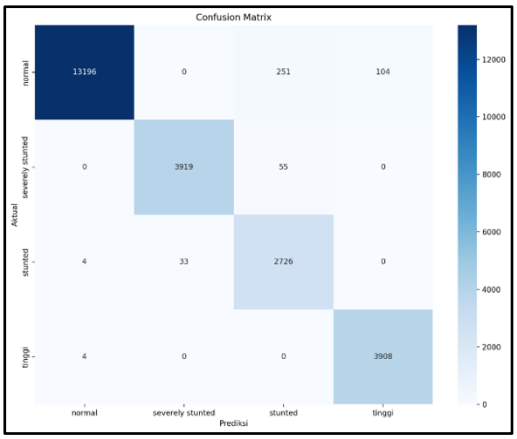

FIGURE 10. Results of confusion matrix

There are three features that are considered in predicting nutritional status, namely age, height, and gender. Through the feature importance of the model used, we can see which features or variables have the most significant influence on predicting children's nutritional status. Based on the three features, height has the highest importance score. The highest value, approaching 0.47, indicates that height is the most significant factor in predicting nutritional status. Furthermore, age has an importance score slightly above 0.41, indicating that age is also an important influence, although less significant than height. Furthermore the gender feature had the lowest importance score, below 0.09. This indicates that gender has a relatively small influence. Overall, this model shows that height and age are the most important factors in predicting nutritional status.

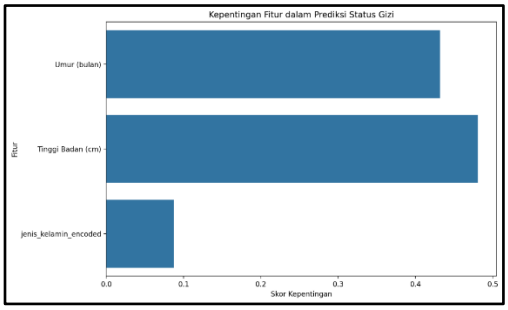

FIGURE 11. Feature Importance

After the evaluation stage, an external testing stage is carried out as a form of validation. Testing using new data outside the existing dataset. At this stage, external testing is performed as an additional validation step and to assess the model's accuracy in predicting new data.

```
Masukkan jenis kelamin anak (Laki-laki/Perempuan): laki-laki
Masukkan umur anak (bulan): 36
Masukkan tinggi badan anak (cm): 95

Hasil Prediksi untuk Data Baru:
   Umur (bulan) Jenis Kelamin  Tinggi Badan (cm) status_gizi_prediksi
0            36     laki-laki               95.0               normal

Probabilitas Prediksi:
[[9.9656087e-01 2.3349990e-04 2.7383415e-03 4.6727760e-04]]
```

```
Masukkan jenis kelamin anak (Laki-laki/Perempuan): laki-laki
Masukkan umur anak (bulan): 36
Masukkan tinggi badan anak (cm): 80

Hasil Prediksi untuk Data Baru:
   Umur (bulan) Jenis Kelamin  Tinggi Badan (cm) status_gizi_prediksi
0            36     laki-laki               80.0     severely stunted

Probabilitas Prediksi:
[[1.2099432e-05 9.9853945e-01 1.4438349e-03 4.6240775e-06]]
```

```
Masukkan jenis kelamin anak (Laki-laki/Perempuan): laki-laki
Masukkan umur anak (bulan): 36
Masukkan tinggi badan anak (cm): 85

Hasil Prediksi untuk Data Baru:
   Umur (bulan) Jenis Kelamin  Tinggi Badan (cm) status_gizi_prediksi
0            36     laki-laki               85.0              stunted

Probabilitas Prediksi:
[[5.8882260e-03 1.3537772e-01 8.5799432e-01 7.3975121e-04]]
```

```
Masukkan jenis kelamin anak (Laki-laki/Perempuan): laki-laki
Masukkan umur anak (bulan): 36
Masukkan tinggi badan anak (cm): 110

Hasil Prediksi untuk Data Baru:
   Umur (bulan) Jenis Kelamin  Tinggi Badan (cm) status_gizi_prediksi
0            36     laki-laki              110.0              tinggi

Probabilitas Prediksi:
[[9.7432584e-03 3.4720415e-05 4.2909310e-06 9.9021775e-01]]
```

FIGURE 12. Testing results

Based on the results of external testing, it is known that the model successfully determined the nutritional status for each data provided.

Previous studies related to stunting risk prediction generally used basic machine learning algorithms such as Logistic Regression, Decision Tree, Naïve Bayes, and Random Forest. Some studies used logistic regression to model stunting, but this approach is only capable of binary classification and does not provide a comprehensive analysis of the contribution of each feature to the prediction results [17]. In addition, most previous studies only focused on identifying determinants of stunting, not on developing multi-class prediction models for four categories of nutritional status [8][15]. Compared with traditional models such as logistic regression, the XGBoost algorithm has been shown to be superior for tabular health data due to its ability to handle complex feature interactions, overcome missing patterns, and produce higher prediction accuracy and stability [11].

The results of the study showed that the XGBoost model was able to achieve an accuracy of 98%, which is a significant improvement compared to previous studies which generally only ranged from 80–90% with algorithms such as Decision Tree, SVM, or Random Forest [8][15][9]. In the context of stunting prediction, applying the binary logistic regression method that only separates two categories (stunting vs. non-stunting) produces lower performance than XGBoost [17]. This study successfully predicted four categories of nutritional status with high precision and recall across all classes, as shown by the results of the classification report and confusion matrix. Model stability was also confirmed through k-fold cross-validation with a mean CV value of 0.9899 and SD 0.0005, which is much better than previous studies which generally do not apply layered validation and tend to be

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

susceptible to overfitting. External validation using new data further strengthens the consistency of model performance, an approach that is still rarely used in machine learning studies related to stunting. Furthermore, feature importance analysis revealed that height was the most dominant predictor, followed by age, while gender had a relatively small contribution. These findings align with global epidemiological studies, such as those conducted by UNICEF-WHO.

IV.CONCLUSION

This study successfully developed a predictive model for assessing the risk of stunting in children using the XGBoost algorithm. Based on the evaluation results from the confusion matrix and classification report, the XGBoost model demonstrated an effectiveness of 98% in predicting the nutritional status of children under five. The variable found to have the most significant influence on the prediction of nutritional status was height. Future studies are recommended to conduct further validation of the model using data from different regions or time periods to ensure its generalization capability under various conditions. Since height proved to be the most significant predictor, it is also suggested that future research expand the range of input variables to include environmental, socioeconomic, parenting, and health factors, so that the model can produce more comprehensive results and not rely solely on a single physical indicator.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to all parties who have contributed to the completion of this research, and especially to the faculty of engineering, University of Bengkulu, for providing financial support for this study.

REFERENCE

[1] M. Simamora, R. Sipayung, J. Sinaga, and A. A. Sutrisna, "Kejadian Stunting dengan Kemampuan Kognitif Anak Usia Sekolah," Jurnal Online Keperawatan Indonesia, vol. 6, no. 1, pp. 29–36, 2023. doi: 10.51544/keperawatan.v6i1.4304.

[2] N. Muthiah, Efektivitas Implementasi Kebijakan Penanganan Stunting di Indonesia dan Aspek Penting yang Perlu Dimasukkan dalam RUU KIA. Jakarta: The Indonesian Institute, 2022.

[3] E. Sumartini, "Dampak Stunting Terhadap Kemampuan Kognitif Anak," in Prosiding Seminar Nasional Kesehatan: Peran Tenaga Kesehatan Dalam Menurunkan Kejadian Stunting, 2020, pp. 127–134.

[4] Awaludin, "Analisis Bagaimana Mengatasi Permasalahan Stunting di Indonesia," Jurnal Kedokteran, vol. 35, no. 4, p. 60, 2019.

[5] J. Julaecha, "Edukasi Periode Emas 1000 Hari Pertama Kehidupan," Jurnal Abdimas Kesehatan (JAK), vol. 2, no. 3, p. 163, 2020. doi: 10.36565/jak.v2i3.109.

[6] H. Kurniati, R. Djuwita, and M. Istiqfani, "Literature Review: Stunting Saat Balita sebagai Salah Satu Faktor Risiko Penyakit Tidak Menular di Masa Depan," Jurnal Epidemiologi Kesehatan Indonesia, vol. 6, no. 2, 2023. doi: 10.7454/epidkes.v6i2.6349.

[7] Y. Theresiana, D. Novira, J. H. Nurdian, H. Pansori, and F. B. M. Harianja, "Analisis Kejadian Stunting, Gizi Buruk dan Gizi Kurang pada Balita di Desa Sari Mulya Cengri Kabupaten Seluma Bengkulu Tahun 2024," Student Scientific Journal, vol. 2, no. 2, pp. 109–114, 2024.

[8] A. Roihan, P. A. Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper," IJCIT (Indonesian Journal on Computer and Information Technology), vol. 5, no. 1, pp. 75–82, 2020. doi: 10.31294/ijcit.v5i1.7951.

[9] K. D. K. Wardhani and M. Akbar, "Diabetes Risk Prediction Using Extreme Gradient Boosting (XGBoost)," JOIN: Jurnal Online Informatika, vol. 7, no. 2, pp. 244–250, 2022. doi: 10.15575/join.v7i2.970.

[10] R. N. Alifah et al., "Perbandingan Metode Tree Based Classification untuk Masalah Klasifikasi Data Body Mass Index," Indones. J. Math. Nat. Sci., vol. 47, no. 1, 2024. Available: https://journal.unnes.ac.id/journals/JM/index.

[11] G. Abdurrahman, H. Oktavianto, and M. Sintawati, "Optimasi Algoritma XGBoost Classifier Menggunakan Hyperparameter GridSearch dan Random Search pada Klasifikasi Penyakit Diabetes," INFORMAL: Informatics Journal, vol. 7, no. 3, p. 193, 2022. doi: 10.19184/isj.v7i3.354.

[12] A. Rianti, N. W. A. Majid, and A. Fauzi, "CRISP-DM: Metodologi Proyek Data Science," in Prosiding Seminar Nasional Teknologi, 2023, pp. 107–114.

[13] B. Mesut, A. Başkor, and N. Buket Aksu, "Role of Artificial Intelligence in Quality Profiling and Optimization of Drug Products," in A Handbook of Artificial Intelligence in Drug Delivery, Elsevier, 2023, pp. 35–54. doi: 10.1016/B978-0-323-89925-3.00003-4.

[14] A. F. Nugraha, Y. Pristyanto, and I. Pratama, "Penanganan Missing Values untuk Meningkatkan Kinerja Model Machine Learning pada Data Telemarketing," Pseudocode, vol. 7, no. 2, pp. 165–171, 2020. doi: 10.33369/pseudocode.7.2.165-171.

[15] W. A. Firmansyach, U. Hayati, and Y. A. Wijaya, "Analisa Terjadinya Overfitting dan Underfitting pada Algoritma Naive Bayes dan Decision Tree dengan Teknik Cross Validation," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 1, pp. 262–269, 2023. doi: 10.36040/jati.v7i1.6329.

[16] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier dan Confusion Matrix pada Analisis Sentimen Berbasis Teks di Twitter," Jurnal Sains Komputer & Informatika (J-SAKTI), vol. 5, no. 2, pp. 697–711, 2021.

[17] S. Ningsih, M. R. Madonsa, S. L. Mahmud, I. Djakaria, and S. K. Nasib, "Implementasi Regresi Logistik Biner Stratifikasi pada Pemodelan Stunting untuk Anak Balita di Kabupaten Gorontalo," Jambura Journal of Probability and Statistics, vol. 5, no. 1, pp. 19–23, 2024. doi: 10.37905/jjps.v5i1.19793.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

© 2025 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/