International Journal Of Health, Engineering And Technology (IJHET) E-ISSN 2829 - 8683
Volume 5, Number 1, May 2026, *Page. 413 - 417*
Email : editorijhess@gmail.com

Classification Of Stunting In Toddlers Using The Random Forest Method

Sutarti Dwi Yanti[1)], Menur Wahyu Pangestika [2)]*, Alda Cendekia Siregar[3)]
[1,2,3)] Informatics, Faculty of Engineering and Computer Science, Muhammadiyah University of Pontianak

*Corresponding Author
Email : sutartidwiyanti@gmail.com

Abstract

Monitoring and data collection related to stunting at several community health centers (Puskesmas) in Ketapang Regency play a crucial role in assessing optimal growth and development of fetuses and newborns. One of the recurring issues in Ketapang Regency is the inaccuracy and inconsistency in monthly stunting data collection. This study aims to design a stunting classification model for children under five using the random forest method and to evaluate its classification accuracy. The research follows several stages problem identification, literature review, data collection, data processing, testing, and drawing conclusions. The performance of the random forest method is assessed to determine the impact of each stage on the model's classification ability. Evaluation metrics are derived from the confusion matrix. The confusion matrix results show a recall of 97%, precision of 96%, F1-score of 96%, and an accuracy of 91%, indicating that the random forest method performs excellently in classifying nutritional status.

Keywords: Classification, Confusion Matrix, Data Mining, Random Forest, Stunting.

INTRODUCTION

According to the World Health Organization (WHO), stunting affects more than 150 million children under the age of five worldwide, with long-term impacts on health, education, and economic potential. In developing countries, stunting represents not only a health issue but also a social and economic problem, caused by multiple factors such as maternal nutrition, access to nutritious food, sanitation, and healthcare services. This complexity often renders traditional approaches ineffective. In Indonesia, stunting remains a serious concern. The growth stages of toddlers are highly influenced by nutritional intake from birth. Children aged 12–59 months constitute a vulnerable group to health and nutritional disorders. Inadequate nutritional intake, particularly protein and energy, may lead to growth impairment. Malnutrition and stunting are closely interrelated, and if not properly addressed, they may reduce a child's capabilities and overall health quality.

Ketapang Regency is one of the regencies located in West Kalimantan Province, covering an area of approximately 31,588 km². The prevalence of stunting in this region reaches 23.6%, which is considered relatively high (classified under the "orange" category) and ranks eleventh among regions in West Kalimantan. Early detection of stunting is crucial to prevent more severe long-term consequences. This study was also conducted due to existing challenges in the collection and processing of stunting data at Kedondong Community Health Centre (Puskesmas) in Ketapang Regency, where data management is still performed manually using Microsoft Excel. This results in slow, less accurate analysis and makes early detection difficult. With advancements in large-scale data processing technologies, data mining offers a promising solution for healthcare data management. Therefore, a data mining–based classification system is required to assist healthcare professionals in monitoring nutritional status and detecting stunting risks more quickly, accurately, and efficiently.

Previous studies on stunting classification have applied various methods. One study entitled *"Application of the Naïve Bayes Classifier Algorithm for Classifying Stunting Nutritional Status in Toddlers"* used parameters such as gender, age, weight, height, mid-upper arm circumference (MUAC), height-for-age Z-score (HAZ), and height-for-age index (H/A) as labels (normal, short, and very short), achieving an accuracy of 94.65%, indicating strong classification performance [3].

Another study, *"Web-Based Stunting Detection System in Toddlers Using the Random Forest Method"*, employed parameters including gender, age, birth weight, birth length, body weight, body length, breastfeeding status, and stunting status (yes/no). This study successfully developed a web-based stunting detection model using the Random Forest method, with the best model obtained from a 90:10 data split after hyperparameter tuning, achieving an accuracy of 85%. These findings demonstrate that the Random Forest method has significant potential in addressing complex classification problems.

The first study, entitled *"Classification of Hypertension Disease Using the Random Forest Method"*, was conducted by Novianti in 2024. The focus of this study was to develop a classification model for hypertension using the Random Forest method. The results indicated that the implementation of the Random Forest method in hypertension classification achieved a high level of accuracy, reaching 98% on the training data and 95% on the testing data. Findings from the Exploratory Data Analysis (EDA) were consistent with previous literature, identifying risk factors such as gender, age, and blood pressure. Despite the presence of data imbalance, preprocessing steps, including SMOTE, were effective in addressing this issue.

The second study, entitled *"Classification of Diabetes Risk Levels Using the Random Forest Algorithm"*, was conducted by Nur Halizah Alfajr in 2024. This study examined the performance of the Random Forest method in classifying diabetes risk levels. The model achieved an accuracy of 98%, precision of 100%, recall of 96%, and an F1-score of 98%. The testing data comprised 33% of a total of 1,026 records, indicating that the Random Forest method is highly effective in performing classification tasks.

The third study, entitled *"Implementation of Random Forest in the Classification of Cardiovascular Diseases"*, was conducted by I Ketut Adian Jayaditya in 2023. Based on the findings, the Random Forest algorithm was found to be applicable for classifying cardiovascular diseases, achieving an accuracy of 69.65%, precision of 69.62%, recall of 69.46%, and an F1-score of 69.54% prior to hyperparameter tuning. After hyperparameter tuning, the performance improved, with accuracy increasing to 73.06%, precision to 75.15%, recall slightly decreasing to 68.72%, and the F1-score rising to 71.79%.

The fourth study, entitled *"Classification of Heart Disease Using the Random Forest Classifier"*, was conducted by Hidayat in 2023. This study involved twelve experimental evaluations, with the best result obtained in the sixth experiment, which used an 80:20 data split and achieved an accuracy of 94% .

Based on the aforementioned problems and previous research findings, the Random Forest method has demonstrated strong capability in performing accurate classification. Therefore, this study applies the Random Forest method to classify stunting in toddlers. The resulting model is expected to assist healthcare professionals in conducting initial screening more efficiently and accurately.

RESEARCH METHODS

The research methodology employed in this study can be seen in Figure 1.

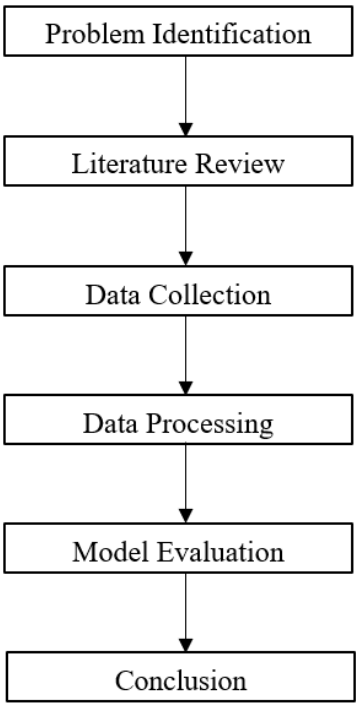

Figure 1 Research Flow Diagram

Problem Identification

The initial step undertaken is problem identification. The problem identified in this study is how to classify stunting using the Random Forest method.

Literature Review

The literature review is the process of collecting references relevant to the research topic. The sources utilised include journals, books, and other scholarly materials from previous studies to support the completion of this research.

Data Collection

Data collection is the process of obtaining the data required for conducting the research. The dataset used in this study was obtained from a community health centre (Puskesmas). The parameters in the dataset include gender, age, body weight, body height, and status. The hardware specifications used to support this research include an AMD Ryzen 3 processor, 4 GB of RAM, and a 1 TB HDD. The software utilised in this study includes Python version 3.12.9, Visual Studio Code version 1.102.0, Google Chrome, Microsoft Excel, and Jupyter Notebook.

Data Processing

Data processing is a data mining procedure aimed at exploring data and identifying useful patterns or information. In this study, data processing is conducted using the Random Forest method. The stages of data processing using the Random Forest method are illustrated in the flowchart shown in Figure 2.

Model Evaluation

The evaluation in this study is conducted using a confusion matrix by dividing the dataset into training and testing sets. Within the confusion matrix method, four performance indicators are assessed, namely accuracy, precision, sensitivity (recall), and F1-score

International Journal Of Health, Engineering And Technology (IJHET) E-ISSN 2829 - 8683
Volume 5, Number 1, May 2026, *Page. 413 - 417*
Email : editorijhess@gmail.com

Conclusion

The conclusion is the output stage that presents a summary of the model's performance.

RESULTS AND DISCUSSION

Figure 2 illustrates the distribution of nutritional status risk. The total dataset used in this study consists of 1,992 records. There are six classes in the classification of nutritional status: class 0 indicates normal nutrition, class 1 indicates a risk of overnutrition, class 2 indicates overnutrition, class 3 indicates obesity, class 4 indicates undernutrition, and class 5 indicates severe malnutrition. The format of the results of research and discussion is not separated, considering the number of pages

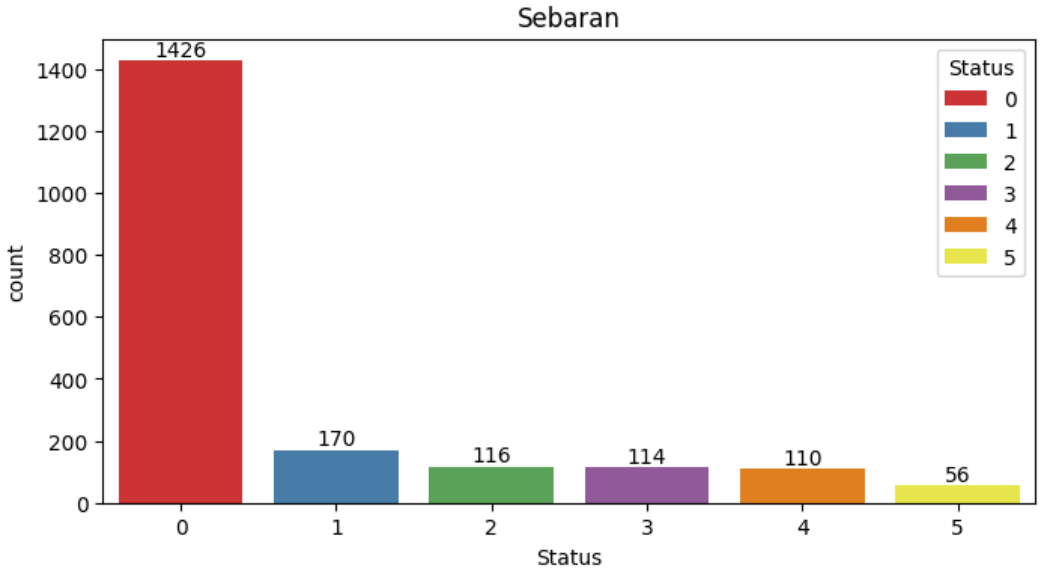

Figure 2 Bar Chart of Nutritional Status

Next, a correlation check between features is performed using a correlation matrix, as shown in Figure 3.

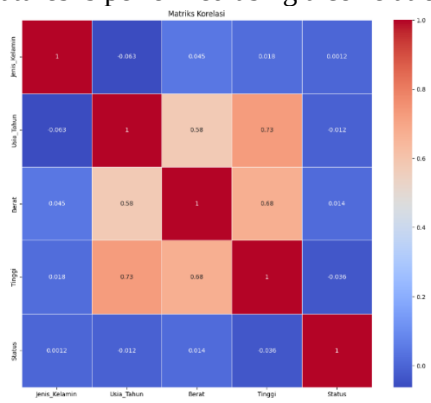

Figure 3 Correlation Matrix

Figure 3 indicates that the closer the value is to one (1), the stronger the correlation between features, whereas values approaching negative one (−1) indicate a weaker correlation. The age feature has a correlation of 0.58 with weight and 0.73 with height. The weight feature has a correlation of 0.68 with height. Next, the dataset is split into 80% for training data and 20% for testing data. This division aims to utilise the majority of the data for training the model while reserving a portion for evaluating its performance. Consequently, the model can be tested on data that were not used during training, providing a more objective estimate of its ability to perform classification on new data. The results of the evaluation can be seen in Table 1.

Table 1

No	Evaluation Matrics	Result (%)
1	*Accuracy*	93%
2	*Precision*	96%
3	*Recall*	99%
4	*F1-Score*	97%

CONCLUSION

The results of this study were obtained through several stages, starting from data input, exploratory data analysis (EDA), and preprocessing, which includes handling missing values, checking for duplicated data, and examining correlations between features. These stages are followed by separating features and target variables, splitting the data into training and testing sets, implementing the method, and evaluating the model. The results of testing the Random Forest method, along with the modelling stages, were evaluated to determine the impact of each stage on the model's ability to perform classification. The evaluation metrics were derived from the confusion matrix. The results indicate an accuracy of 93%, recall of 99%, precision of 96%, and an F1-score of 97%, demonstrating that the Random Forest method performs exceptionally well in classifying the nutritional status of toddlers.

REFERENCES

F. Maula Hidayat, Kusrini, and Ainul Yaqin, "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Status Gizi Stunting Pada Balita," *Dielektrika*, vol. 11, no. 2, pp. 107–118, 2024, doi: 10.29303/dielektrika.v11i2.384.

Hidayat, Andi Sunyoto, and Hanif Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan *Random forest* Clasifier," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 7, no. 1, pp. 31–40, 2023.

I. G. A. G. Jayaditya, I. K. A, dan Kadyana, "Implementasi *Random forest* pada kLasifikasi Penyakit Kardiovaskular dengan Hyperparameter Tuning Grid Search," *Jnatia*, vol. 2, no. 1, pp. 219–226, 2023.

M. W. Pangestika and A. C. Siregar, "Reduced Rule Base Pada Sistem Pakar Untuk Diagnosa Penyakit Balita Gizi Buruk Di Kalimantan Barat," *Cybernetics*, vol. 3, no. 01, 2019, doi: 10.29406/cbn.v3i01.1818.

N. Cholifatul Izza and A. Irma Rizmayanti, "Analisis Rekam Medis dengan Metode Data Mining untuk Memprediksi Faktor Risiko Stunting dalam Kesehatan Masyarakat," 2024.

N. H. Alfajr and S. Defiyanti, "Prediksi Penyakit Jantung Menggunakan Metode *Random forest* Dan Penerapan Principal Component Analysis (Pca)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3s1.5055.

Novianti, S. Putri Agustini Alkadri, and I. Fakhruzi, "Klasifikasi Penyakit Hipertensi Menggunakan Metode *Random forest*".

S. N. Syafa Iswahyudi and R. Eka Putra, "Sistem Deteksi Stunting pada Balita Berbasis Web Menggunakan Metode *Random forest*," *J. Informatics Comput. Sci.*, vol. 6, no. 03, pp. 755–764, 2025, doi: 10.26740/jinacs.v6n03.p755-764.