Random Forest Algorithm for Toddler Nutritional Status Classification Website

Maylia Fatmawati [1]*, Bambang Agus Herlambang [2]*, Noora Qotrun Nada [3]*
* Informatics, Universitas PGRI Semarang
liafatmawati530@gmail.com [1], bambangherlambang@upgris.ac.id [2], noora@upgris.ac.id [3]

Article Info

Article history:

Received 2024-09-09
Revised 2024-09-20
Accepted 2024-09-26

Keyword:

Random Forest,
Toddler Nutritional Interventions,
Toddler Nutritional Status,
Website.

ABSTRACT

Accurate data processing is essential for classifying toddler nutritional status on a website platform. The Random Forest algorithm is particularly effective in this context due to its ability to manage large datasets and mitigate overfitting. This study leverages Flask as the web framework to ensure responsiveness and adaptability, optimizing the data processing experience for users. Using secondary data comprising 120,999 records, the research aims to answer: "What factors affect the accuracy of the Random Forest model in classifying toddler nutritional status?" Model evaluation yielded excellent performance metrics, with accuracy, precision, recall, and F1-score values of 99.91%, 100%, 100%, and 100%, respectively. These results highlight the informative attributes in the dataset, such as age, gender, and height, that enhance classification accuracy. The Flask-based website enables users, such as healthcare professionals and policymakers, to input essential data points and receive instant classification results, thereby supporting prompt and informed responses to nutritional health issues. This study confirms that the Random Forest algorithm, combined with an intuitive web interface, effectively classifies toddler nutritional status with high accuracy.

I. INTRODUCTIONS

Nutritional status is one of the main indicators in assessing the health condition of individuals, especially toddlers who are in their growth phase. Optimal nutrition ensures that a child receives a balanced diet, which is essential for supporting physical growth, cognitive development, and immune function. Conversely, poor nutritional status can lead to various health problems, such as undernutrition, malnutrition, obesity, and stunting [1]. Indonesia continues to face a high prevalence of stunting, indicating that a significant proportion of toddlers are experiencing chronic nutritional issues. Nutritional problems among toddlers remain a major concern in Indonesia. According to data from the Ministry of Health (KEMENKES), the prevalence of stunting has shown a downward trend over the past few years, with a rate of 21.6% recorded in 2023, down from 24.4% in 2021 and 26.92% in 2020 [2]. Despite these improvements, stunting, as a form of chronic nutritional problem, still requires strategic and collaborative efforts from various stakeholders to be comprehensively addressed. The classification of toddlers' nutritional status is typically carried out through measurements of weight, height, and age, which are then compared to WHO standards. This classification process aims to detect the nutritional status of toddlers early so that preventive measures against chronic nutritional conditions can be implemented promptly. However, this manual method requires time and specialized expertise. Therefore, there is a growing need for a system that can classify toddlers' nutritional status quickly and accurately to reduce cases of malnutrition. Web-based technology, integrated with machine learning algorithms, presents a promising solution to address this challenge [3].

In recent years, numerous studies have explored the application of the Random Forest algorithm, particularly in the areas of stunting and child health. For instance, Muhammad Syauqi et al. [4] developed a predictive model for stunting prevalence among toddlers in East Java using the Random Forest algorithm, achieving a low error rate of 1.02. Similarly, Muhammad Ramadani Akbar Ariyadi et al. [5]

developed a stunting classification model for toddlers in Blitar Regency using a Random Forest classifier, yielding an accuracy of 90%, a precision of 71.4%, and a recall of 62.5%. Another study by Putri Handayani et al. [6] applied the Random Forest algorithm to classify toddlers' nutritional status with an accuracy of 88.6%, using a data split of 10% for testing and 90% for training. Additionally, the Random Forest algorithm has been applied in combination with genetic algorithms to classify toddlers' nutritional status at Puskesmas Cakru. For example, Ersya Nadia Candra et al.

[7] demonstrated that optimizing Random Forest with genetic algorithms resulted in an accuracy of 89.58%, highlighting its suitability for classifying toddlers' nutritional status.

The Random Forest algorithm was selected for this study due to its superior performance in handling large datasets and its ability to manage overfitting. Random Forest is a robust ensemble learning technique that builds multiple decision trees and combines their results to enhance classification accuracy [8]. This approach makes it particularly well-suited for handling complex datasets, such as those involving toddlers' nutritional status, where different factors interact in intricate ways. Additionally, Random Forest mitigates overfitting by averaging the results of multiple trees, thereby reducing variance and improving generalization to unseen data.

While existing studies have demonstrated the effectiveness of Random Forest in various contexts related to child health, the present study aims to fill gaps in the literature by applying this algorithm to classify toddlers' nutritional status on a web-based platform. One limitation in prior studies is the narrow scope of features used in classification models. In this study, three features—age, gender, and height—are used to classify nutritional status. These features were chosen based on their direct relevance to WHO growth standards, as they represent key indicators in assessing whether a child is stunted or malnourished. However, it is acknowledged that additional features, such as socioeconomic status, parental education, or dietary intake, could also influence nutritional status. Future research could expand the set of features to include these variables for a more comprehensive analysis. For this study, the focus remains on the three core variables due to their universal applicability and availability in most health datasets.

By applying the Random Forest algorithm and focusing on key features, this study aims to determine the factors that influence the classification of nutritional status in toddlers by implementing the website-based Random Forest algorithm. Health care providers can use this tool to proactively address potential nutritional problems and tailor interventions to meet the specific needs of each child. As a result, this classification system is expected to improve child health outcomes by enabling early detection and intervention.

II. Method

The flowchart of the research methodology is shown in the diagram below.

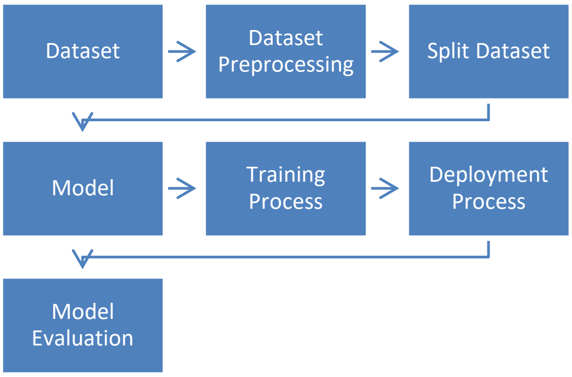


Figure 1. Research Method

A. Dataset

In this study, the researchers used a secondary dataset obtained from the Kaggle platform. The dataset comprises 120,999 records of children's nutritional status, which served as the basis for our analysis. We divided the dataset with a proportion of 80% for training data and 20% for testing data. This data splitting ratio was chosen because it is a common and optimal method to avoid overfitting. The data we used has four attributes: age (in months), gender, height (in centimeters), and nutritional status (classified as severely stunted, stunted, normal, or tall). Nutritional status serves as the target variable. The main consideration during the preprocessing phase was class imbalance, particularly due to the higher prevalence of normal nutritional status compared to stunted or severely stunted cases. The dataset consists of 67,755 children with normal nutritional status, 19,869 with severely stunted status, 19,560 with tall status, and 13,815 with stunted status. One of the main challenges with this dataset is the class imbalance, where "normal" nutritional status cases significantly outnumber the other classes, such as "severely stunted" or "stunted." To address the class imbalance, we employed the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset by generating synthetic samples from the minority class, ensuring that the model is not biased towards the majority class [9]. This step is crucial to prevent biased predictions and enhance the model's ability to detect stunting in children. To ensure the interpretability of the Random Forest model, feature importance values and SHAP (Shapley Additive Explanations) were used. According to the SHAP technique results, the age feature had a value of -0.23344736, the gender feature had a value of -0.07754895, and the height feature had a value of 0.22254059, indicating that the height feature had the most significant impact due to its substantial positive influence, increasing the likelihood of predicting towards the target class. This interpretability is essential in clinical settings, as it allows healthcare providers to understand the basis of the model's predictions and make informed decisions based on the model's output. All data records input for the classification process must be in numerical format. Accurate data entry according to the format is necessary to ensure that the input data can be processed correctly, resulting in optimal outcomes.

B. Dataset Preprocessing

Dataset preprocessing is a critical step in preparing the data for analysis. In this study, several preprocessing steps were applied to the dataset containing 120,999 records of children's nutritional status. The preprocessing steps we conducted included data cleaning, converting categorical data to numerical values, data normalization, handling class imbalance, and feature selection. Data cleaning was performed by checking the data to ensure the absence of missing values. Records with unrealistic values, such as heights or ages beyond logical limits, were either removed or corrected. It was found that there were no missing values in the dataset. Next, the gender column was converted from text form (e.g., "male" or "female") to numerical values (e.g., 0 for male and 1 for female) so that it could be used by the machine learning model. Subsequently, data normalization was carried out; features such as age and height were normalized using the min-max scaling technique to ensure that the features were within a uniform range, preventing the model from being sensitive to differences in feature scales. The next step was addressing the dataset's class imbalance by applying SMOTE (Synthetic Minority Over-sampling Technique). SMOTE was used to generate synthetic data from the minority class, so the model was not overly biased towards the majority class (children with normal nutritional status). The final step was determining the three main features used in this study, namely age, gender, and height, as these three features are relevant to the WHO standards in determining whether a child is stunted or not.

C. Split Dataset

The dataset was split into two parts to train and test the model, with 80% allocated for training data and 20% for testing data. The larger portion of the dataset (80%) was used to train the model. This data included various patterns that enabled the model to learn effectively. The remaining 20% was used to test the model's performance on new data that the model had not seen during training. This is crucial to ensure that the model can generalize well to new data and does not overfit.

D. Model

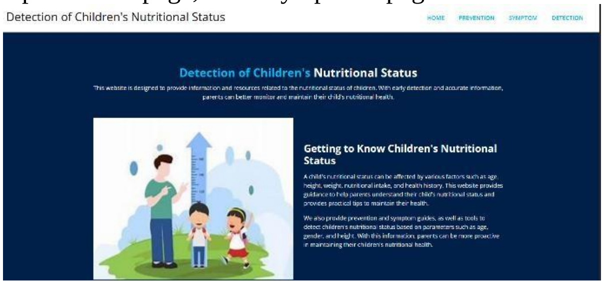
In this study, two models were used to compare performance: Random Forest and TabNet. These models were chosen because of their advantages in handling large datasets and providing accurate predictions. Random Forest was selected for its ability to handle overfitting and manage complex datasets. Using an ensemble learning approach, Random Forest builds multiple decision trees and combines their results, making it more resistant to data fluctuations. TabNet was chosen as a comparative model due to its ability to perform sequential attention, enabling the model to learn representations from features more effectively and hierarchically. TabNet is well-suited for tabular datasets, such as the children's nutritional status dataset used in this study.

E. Training Process

The model training process involves several critical steps, such as model initialization, model training, model performance validation, and the application of SMOTE. In the initial stage, the Random Forest model was initialized with 100 decision trees as the initial parameter. Other parameters were also adjusted for optimization. Meanwhile, TabNet was initialized with appropriate hyperparameters, including batch size, learning rate, and the depth of attention layers, which were optimized to achieve the best results. Next was the model training step, where both models were trained using the training dataset (80% of the data). The "fit" algorithm from scikit-learn was used for Random Forest, while TabNet was trained using a specifically implemented architecture for tabular data. The third step involved validating the model's performance to ensure it was not overfitting; cross-validation was performed on both models. This technique validates the model on several different data subsets. The final step involved applying SMOTE to balance the minority and majority classes in the dataset.

F. Deployment Process

After the models were trained and validated, the next step was deployment into a Flask-based web application. The Random Forest model was then integrated into the web application using the Flask framework. Flask was chosen for its lightweight nature and ease of integration with Python, as well as its compatibility for machine learning model deployment. After that, a user interface design was created to allow users to access the website easily, informatively, and user-friendly. Users can also input data such as the child's age, gender, and height. The model then provides real-time nutritional status predictions. In addition to the prediction feature on the website, it is also equipped with a home page, a prevention page, and a symptoms page.



Detection of Children's Nutritional Status

This website is designed to provide information and resources related to the nutritional status of children. With early detection and accurate information, parents can better monitor and maintain their child's nutritional health.

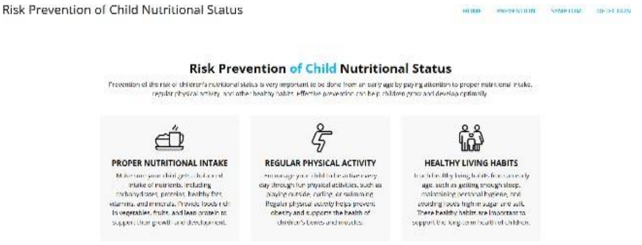
Getting to Know Children's Nutritional Status

A child's nutritional status can be affected by various factors such as age, height, weight, nutritional intake, and health history. This website provides guidance to help parents understand their child's nutritional status and provides practical tips to maintain their health.

We also provide prevention and symptom guides, as well as tools to detect children's nutritional status based on parameters such as age, gender, and height. With this information, parents can be more proactive in maintaining their children's nutritional health.

Figure 2. Home Page Display

The home page is the main view that appears first when users enter the website. It serves as an introduction and provides general information about the purpose and benefits of this toddler nutritional status classification prediction website.



Risk Prevention of Child Nutritional Status

Risk Prevention of Child Nutritional Status

Prevention of the risk of children's nutritional status is very important to be done from an early age by paying attention to proper nutritional intake, regular physical activity, and other healthy habits. Effective prevention can help children grow and develop optimally.

PROPER NUTRITIONAL INTAKE

Make sure your child gets a balanced intake of nutrients, including carbohydrates, proteins, healthy fats, vitamins, and minerals. Provide foods rich in vegetables, fruits, and lean protein to support their growth and development.

REGULAR PHYSICAL ACTIVITY

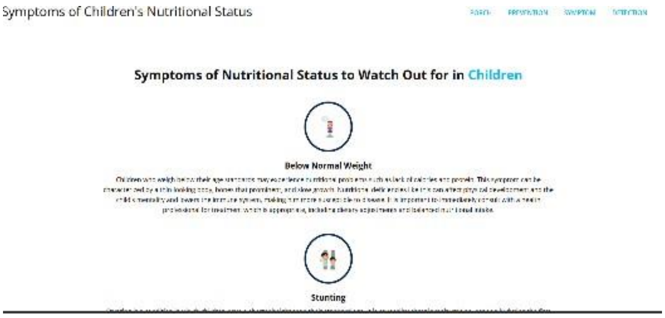
Encourage your child to be active every day through fun physical activities, such as playing outside, cycling, or swimming. Regular physical activity helps prevent obesity and supports the health of children's bones and muscles.

HEALTHY LIVING HABITS

Teach healthy living habits from an early age, such as getting enough sleep, maintaining personal hygiene, and avoiding foods high in sugar and salt. These healthy habits are important to support the long-term health of children.

Figure 3. Prevention Page Display

The prevention page provides information and guidelines on how to prevent the risk of malnutrition in toddlers. This page includes tips on proper nutritional intake, regular physical activity, and a healthy lifestyle.



Symptoms of Children's Nutritional Status

Symptoms of Nutritional Status to Watch Out for in Children

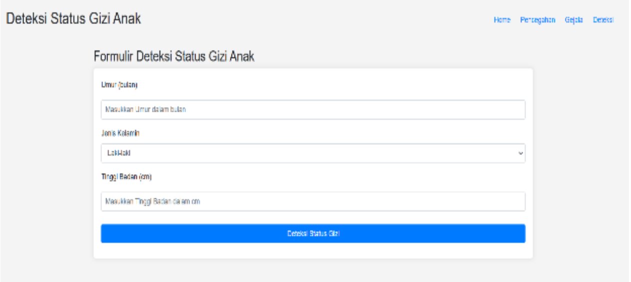
Below Normal Weight

Children who weigh below their age standards may experience nutritional problems such as lack of calories and protein. This symptom can be characterized by a thin looking body, bones that prominent, and slow growth. Nutritional deficiencies like this can affect physical development and the child's mentality and lowers the immune system, making him more susceptible to disease. It is important to immediately consult with a health professional for treatment which is appropriate, including dietary adjustments and balanced nutritional intake.

Stunting

Figure 4. Symptoms Page Display

The symptoms page contains information about various signs that may indicate the risk of malnutrition in toddlers. This page helps identify early symptoms that require medical attention.



Deteksi Status Gizi Anak

Formulir Deteksi Status Gizi Anak

Umur (bulan)

Masukkan Umur dalam bulan

Jenis Kelamin

Laki-laki

Tinggi Badan (cm)

Masukkan Tinggi Badan dalam cm

Figure 5. Detection Page Display

The detection page is a key feature that allows users to predict the nutritional status classification of toddlers. This classification prediction process uses the Random Forest algorithm to analyze data and provide prediction results. This system is expected to enhance the effectiveness of nutritional interventions and health policies by providing more accurate real-time monitoring, enabling quick anticipation and actions against the risks of suboptimal growth, such as height not matching the age and gender, based on established growth standards.

After the deployment process, the web application was tested to ensure the system functions correctly. The test results indicated that the application successfully received inputs, processed data, and provided accurate classification results.

G. Model Evaluation

Model evaluation is a crucial step to measure the model's performance after training. In this study, two models were evaluated: Random Forest and TabNet. The metrics used to evaluate both models were accuracy, precision, recall, F1-Score, confusion matrix, and SHAP analysis. The accuracy evaluation showed that Random Forest achieved an accuracy of 99.91%, demonstrating outstanding performance in predicting children's nutritional status. Additionally, TabNet achieved an accuracy of 96%, slightly lower than Random Forest. However, this result still indicates that TabNet is a competitive model for tabular data classification. Furthermore, for Random Forest, precision and recall for all classes were 100%, resulting in a perfect F1-Score. This indicates that the model can provide highly accurate predictions for all classes, while TabNet, though not as accurate as Random Forest, still demonstrated strong performance with high precision and recall for most classes. The confusion matrix showed that Random Forest made almost no errors in classification, whereas TabNet had a few errors in predicting minority classes like "stunted" and "severely stunted." Next was SHAP analysis to assess the interpretability of the Random Forest model's predictions. SHAP provided insights into how features such as height and age influenced the model's predictions. In Random Forest, height was the most influential feature, whereas TabNet used the attention mechanism to assign different weights to features based on the hierarchical attention sequence [10-15]. With these strong evaluation results, it can be concluded that Random Forest provided superior performance in this study.

III. RESULT AND DISCUSSION

This chapter presents the research findings related to the implementation of the Random Forest algorithm for toddler nutritional status classification. The research results include an evaluation of the model's performance, an analysis of feature importance in prediction outcomes, and the methods applied to address data imbalance. Additionally, this chapter discusses the interpretation of the results obtained, an analysis of the model's limitations, and potential future developments for further research.

The findings are expected to provide a clear understanding of the effectiveness of using the Random Forest algorithm for toddler nutritional status classification and offer insights into how this model can be practically applied in a web-based environment. An in-depth analysis comparing this model with other approaches, such as TabNet, is also conducted to determine the strengths and weaknesses of each model. Finally, the discussion will focus on the implications of these research findings for efforts to prevent nutritional problems in toddlers and the potential for system development in the future.

A. Model Comparison

In this study, a performance comparison was conducted between the Random Forest model and the TabNet model for classifying toddler nutritional status. The models' performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. The evaluation results are presented in a table to provide a clear overview of the strengths of each model.

TABLE I.
MODEL PERFORMANCE

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	99.91	100	100	100
TabNet	96.00	97.50	97.20	97.35

From the table above, it is evident that the Random Forest model has a higher accuracy compared to TabNet, with a difference of 3.91%. This indicates that Random Forest is more effective in classifying toddler nutritional status.

B. Feature Importance

The influence of features on the prediction outcomes was analyzed using SHAP (Shapley Additive Explanations) on the Random Forest model. The results of the feature importance analysis are as follows:

1) *Height*: This feature has the greatest influence with a significantly positive SHAP value, indicating that taller toddlers are more likely to have a better nutritional status. The SHAP value for height is 0.22254059.

2) *Age*: Age shows a negative SHAP value, meaning that younger toddlers tend to have a higher risk of poorer nutritional status. The SHAP value for age is -0.23344736.

3) *Gender*: Gender has a smaller influence compared to height and age but still contributes to the prediction model. The SHAP value for gender is -0.07754895.

C. Data Imbalance Analysis

The dataset used in this study has a class imbalance issue, with the number of samples in the 'normal' class being much higher than in the 'stunted' and 'severely stunted' classes. To address this issue, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied to balance the number of samples in each class.

SMOTE Formula:
$xnew=xi+\lambda\cdot(xj-xi)$
where:

- xnew = New synthetic sample
- xi = Existing minority class sample
- xj = Nearest neighbor of the minority class sample
- λ = Random value between 0 and 1

Applying SMOTE resulted in a more balanced class distribution, which in turn enhanced the model's ability to detect minority classes such as 'stunted' and 'severely stunted’. This improvement can be seen from the increased recall values for these classes after data balancing.

D. System Implementation

The implementation of the web-based nutritional status classification system for toddlers using the Random Forest algorithm was successfully completed. This system allows users to input data such as the toddler's age, gender, and height, and receive real-time nutritional status classification results. The use case diagram is employed to illustrate the various behaviors within the system [11]. The use case diagram for the toddlers' nutritional status classification system is shown in Figure 6.

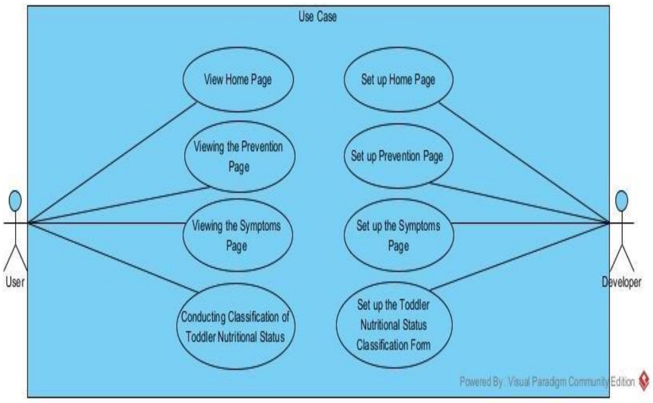

Use Case
View Home Page
Set up Home Page
Viewing the Prevention Page
Set up Prevention Page
Viewing the Symptoms Page
Set up the Symptoms Page
Conducting Classification of Toddler Nutritional Status
Set up the Toddler Nutritional Status Classification Form
User
Developer
Powered By: Visual Paradigm Community Edition

Figure 6. Use Case Diagram

The use case diagram shown in Figure 6 illustrates that users can perform several activities such as viewing the home page, viewing the prevention page, viewing the symptoms page, and making predictions for toddlers' nutritional status classification by inputting necessary data such as age, gender, and height. Additionally, developers can perform activities such as setting up the home page, setting up the prevention page, setting up the symptoms page, and setting up the nutritional status classification form.

E. Result Analysis

The evaluation results indicate that the Random Forest model performs better than the TabNet model in classifying the nutritional status of toddlers. The strength of Random Forest lies in its ability to handle large and complex datasets and overcome overfitting issues through its ensemble method, which combines the results of multiple decision trees. The TabNet model, despite its advantage in processing tabular data with its attention mechanism, shows lower performance compared to Random Forest. This may be due to the characteristics of the data not being well-suited to the hierarchical attention approach used by TabNet.

F. Limitations and Potential Developments

Some limitations in this study include:

1) Feature Limitation: This study only used three main features, namely age, gender, and height. Adding other features such as health history, diet, and socioeconomic status could improve the model's accuracy and generalization capability.

2) Data Imbalance: Although data balancing was performed using SMOTE, there is still the possibility of bias in the model towards the majority class.

Potential future developments include:

- Adding Features: Adding more relevant features to improve the model's performance, such as environmental and genetic factors.
- Hybrid Model: Combining several models such as Random Forest with deep learning models to leverage the strengths of each approach.
- Cross-Region Evaluation: Using data from various regions to ensure that the model can be generalized to a wider population.

In conclusion, the results of this study demonstrate that the use of the Random Forest algorithm in the web-based nutritional status classification system for toddlers can provide high accuracy and effectively handle data imbalance. Further developments could improve the model's performance and extend its application in various contexts related to maternal and child health.

V. Conclusion

This study successfully implemented the Random Forest algorithm in a web-based system for classifying toddler nutritional status, achieving high performance with 99.91% accuracy, 100% precision, recall, and F1-score. The model effectively handled class imbalances using the SMOTE technique and demonstrated accurate classification of both majority and minority classes, including 'stunted' and 'severely stunted'. Key features such as age, gender, and height significantly contributed to the model's success. The Flask-based application allows real-time classification, providing healthcare professionals and policymakers with a valuable tool for timely and precise nutritional interventions. Future research could enhance the model by incorporating additional variables like socioeconomic status and dietary patterns, broadening its applicability for better-informed health policies to combat malnutrition and stunting in Indonesia.

References

[1] S. Anwar, E. Winarti, and S. Sunardi, "Systematic Review of Risk Factors, Causes, and Impacts of Stunting in Children," Journal of Health Sciences, vol. 11, no. 1, pp. 88-94, 2022.

[2] O. Martony, "Stunting in Indonesia: Challenges and Solutions in the Modern Era," Journal of Telenursing (JOTING), vol. 5, no. 2, pp. 1734-1745, 2023.

[3] M. Mukhsin, "The Role of Information and Communication Technology in Implementing Village Information Systems for the Publication of Village Information in the Globalization Era," Teknokom, vol. 3, no. 1, pp. 7-15, 2020.

[4] M. S. Haris, M. Anshori, and A. Khudori, "Prediction of Stunting Prevalence in East Java Province with Random Forest Algorithm," Journal of Informatics Engineering (Jutif), vol. 4, no. 1, pp. 11-13, 2023.

[5] M. R. A. Ariyadi, S. Lestanti, and S. Kirom, "Classification of Stunted Toddlers Using Random Forest Classifier in Blitar District," JATI (Journal of Informatics Engineering Students), vol. 7, no. 6, pp. 3846-3851, 2023.

[6] P. Handayani, A. F. Charis, and H. Harliana, "Machine Learning Classification of Toddler Nutritional Status Using Random Forest Algorithm," KLIK: Scientific Review of Informatics and Computer, vol. 4, no. 6, pp. 3064-3072, 2024.

[7] E. N. Candra, I. Cholissodin, and R. C. Wihandika, "Classification of Toddler Nutritional Status Using Random Forest Optimization Method with Genetic Algorithm (Case Study: Cakru Public Health Center)," Journal of Information Technology and Computer Science Development, vol. 6, no. 5, pp. 2188-2197, 2022.

[8] O. Adiputra and E. Setiawan, "Classification of Malicious URLs Using Improved Random Forest and Web-Based Random Forest Algorithm," Sains dan Informatika: Research of Science And Informatic, vol. 9, no. 1, pp. 8-14, 2023.

[9] Mansourifar, Hadi; SHI, Weidong. Deep synthetic minority over-sampling technique. arXiv preprint arXiv:2003.09788, 2020.

[10] De Zarzà, Irene; De Curtò, Joachim; Calafate, Carlos T. Area Estimation Of Forest Fires using TabNet with Transformers. Procedia Computer Science, 2023, 225: 553-563.

[11] M. N. Arifin and D. Siahaan, "Structural and Semantic Similarity Measurement of UML Use Case Diagram," Lontar Komputer: Scientific Journal of Information Technology, vol. 11, no. 2, p. 88, 2020.

[12] R. Gustriansyah, N. Suhandi, S. Puspasari, and A. Sanmorino, “Machine Learning Method to Predict the Toddlers’ Nutritional Status”, INFOTEL, vol. 16, no. 1, pp. 32-43, Jan. 2024.

[13] I. Rahmi, Y. Wulandari, H. Yozza, and M. Syafwan, “Classification Of Toddler’s Nutritional Status Using The Rough Set Algorithm”, Barekeng: J. Math. & App., vol. 17, no. 3, pp. 1483-1494, Sep. 2023.

[14] M. Ula, A. F. Ulva, M. Mauliza, M. A. Ali, and Y. R. Said, “Application Of Machine Learning In Determining The Classification Of Children’s Nutrition With Decision Tree”, J. Tek. Inform. (JUTIF), vol. 3, no. 5, pp. 1457-1465, Sep. 2022.

[15] M. G. Daffa and P. H. Gunawan, "Stunting Classification Analysis for Toddlers in Bojongsoang: A Data-Driven Approach," 2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT), Bandung, Indonesia, 2024, pp. 42-46, doi: 10.1109/ICoSEIT60086.2024.10497515.