Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.14264 p-ISSN : 2541-044X

Parameter Testing on Random Forest Algorithm for Stunting Prediction

Ahmad Hasan Mubarok[1)*], Pujiono[2)], Dicky Setiawan[3)], Duta Firdaus Wicaksono[4)], Eti Rimawati[5)]
[1,2,3,4,5)]Universitas Dian Nuswantoro, Indonesia
[1)]112202106725@mhs.dinus.ac.id, [2)]pujiono@dsn.dinus.ac.id, [3)]dickysstwn@gmail.com,
[4)]1112202006480@mhs.dinus.ac.id, [5)]eti.rimawati@dsn.dinus.ac.id

Submitted :Nov 26, 2024 | **Accepted** : Dec 2, 2024 | **Published** : Jan 8, 2025

Abstract: Stunting is a significant public health problem, especially in developing countries like Indonesia. It is often caused by chronic malnutrition in the first 1,000 days of life, which can impact a child's physical growth and cognitive development. To find risk factors and find effective solutions, data analysis was conducted by utilising machine learning to predict stunting. This research uses the Random Forest algorithm with a focus on setting parameters such as n_estimators, max_depth, and the number of features to optimize model efficiency and accuracy. Using the 2023 Indonesian Health Survey data consisting of 25,800 data, this study managed to get the highest accuracy of 91.65% by a combination of Random Forest with parameter settings n_estimators 200, max_depth 30, and Synthetic Minority Oversampling Technique (SMOTE). Despite the high accuracy results, there are limitations such as potential noise coming from synthetic data from SMOTE and the limited number of features analysed. It is hoped that this research can contribute to the field of machine learning model development that is practically used to predict stunting and support the government's efforts to reduce the stunting prevalence rate to 14% as targeted. This model also provides strategic insights for policy makers to design more effective data-driven interventions, which can help in decision making.

Keywords: Machine Learning; Random Forest; n_estimators; max_depth; SMOTE; Stunting

INTRODUCTION

Stunting is a serious public health problem today, especially in developing countries such as Indonesia. This condition is characterised by the height of children under five who are below minus two standard deviations from the median child growth standards according to the World Health Organisation (WHO) (Fitriani & Darmawi, 2022). The cause is none other than chronic malnutrition that occurs from pregnancy to the age of two, known as the first 1000 days of life. Stunted physical growth and impaired cognitive development have a devastating impact on the sufferer. Worse still, these impacts can affect economic productivity and increase the risk of chronic disease when the sufferer grows up (Ritonga et al., 2024). The Indonesian Nutrition Status Survey (SSGI) in 2022, stated that the prevalence of stunting in Indonesia reached 21.6%. When compared to previous years, this figure shows a decrease, but the government has a target of reducing the prevalence of stunting to 14% by 2024 so significant efforts are still needed to achieve this target (RI, 2022).

In addressing stunting, accurate data collection and analysis is essential to understand risk factors and help plan effective interventions. Machine learning and Artificial intelligence (AI) offer innovative approaches to health data analysis. Machine learning enables the discovery of hidden patterns in large-scale data with high complexity, which may be difficult to find with conventional statistical methods (Royhan Zaki Ramadhana & Muhammad Irwan Padli Nasution, 2024). From several widely used machine learning algorithms, Random Forest was chosen as a research testing tool on the basis of its ability to handle large, heterogeneous datasets full of complex variables (Melvin & Soraya, 2023).

Efforts to produce model accuracy and efficiency in Random Forest, of course, cannot be separated from the role of hyperparamter tuning, such as the number of trees, maximum depth, and the number of features selected for each separator (max features). Therefore, this research will conduct a more in-depth study of the effect of these parameter settings in the development of machine learning models for stunting prediction. On the other hand, this research also seeks to identify the main variables that have a strong influence on the risk of stunting, so as to provide insight and guidance in decision making..

*name of corresponding author

Through this research, we hope to contribute to the development of machine learning models for stunting prediction that are not only accurate but also practical to be applied in various contexts. Strategically, this research can also be one of the efforts in supporting the government to reduce the prevalence of stunting and achieve the national health targets set.

LITERATURE REVIEW

One of the powerful and popular machine learning algorithms is random forest, this algorithm belongs to the ensemble learning category, combining the power of several simple models to produce more accurate and reliable predictions (Andriyani et al., 2024). Previous research shows the effectiveness of random forest in predicting various symptoms or diseases. In the study (Sitanggang & Sitompul, 2024), random forest parameters were tested by comparing hyperparameter tuning: random search and grid search for early detection of survival of heart failure patients which showed an increase in accuracy performance from the original 80% to 83.33% and 85%. Then in research (Kurniawan et al., 2024), also testing the parameters of the random forest algorithm with a total of 10 trees for lung disease prediction showed a high accuracy of 94.7%. Regarding the prediction of stunting disease itself, there are various studies that also show that random forest is superior to other algorithms, such as in research (Putri et al., 2024)al., 2024) which involves a comparison of machine learning algorithms namely Naïve Bayes, KNN, and random forest, successfully getting an accuracy of 87.75% through the random forest algorithm. The same discussion is also in research (Marsya Finda & Wahyu Utomo, 2024), by applying Random Forest to classify stunting data resulting in good accuracy of 98.25%. Random Forest can also be integrated with feature selection such as ANOVA and SMOTE techniques to predict stunting, as in the research (Dhani et al., 2024) which produced almost perfect accuracy at 99.77%. However, most of these studies emphasise the prediction results without examining more deeply the effect of parameter settings in the random forest algorithm on model performance. Other relevant research shows that the parameter factor in the random forest algorithm has a major influence on model performance. For example, (Arisusanto et al., 2023) conducted parameter testing by comparing the max_depth value which has shown various differences in accuracy. On the other hand, research in the case of stunting prediction rarely explores the optimisation of parameters such as the number of trees, tree depth, or features selected at each node. This shows a gap in utilising the potential of this algorithm to improve prediction accuracy. Therefore, this research aims to fill this gap by testing and optimising the parameters of the random forest algorithm to improve the accuracy and reliability of stunting prediction.

METHOD

In testing the parameters of the random forest algorithm in this study, of course there are several stages that should be passed, including in the following figure 1.

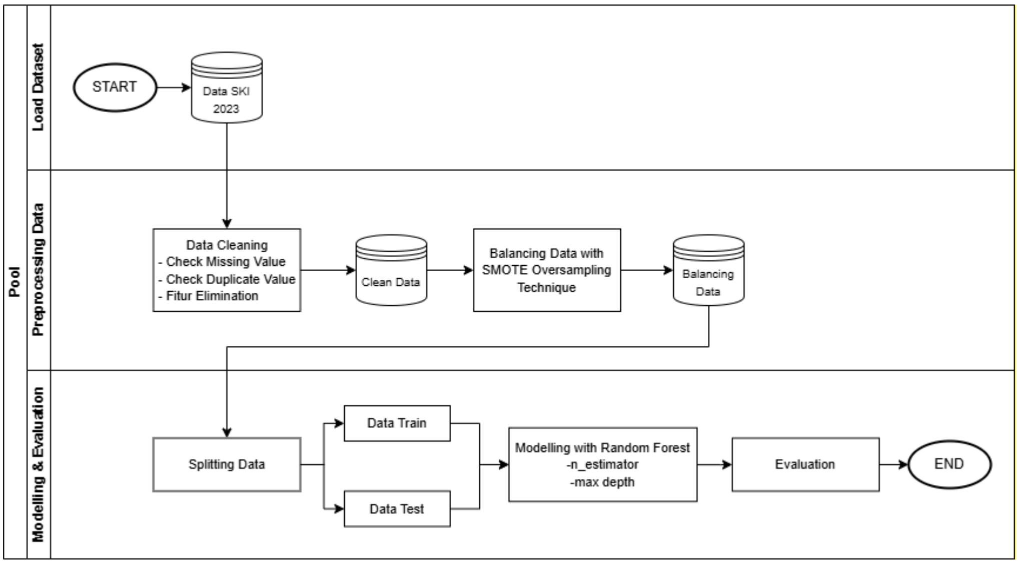

Fig. 1 Method

The first step, of course, is to collect relevant data to support this research. Then it is continued with data pre-processing starting from checking missing values, duplicate values, and selecting the features used. The existing data shows an imbalance in the number between classes, so after the data is confirmed to be clean, data balancing is carried out with SMOTE oversampling. Data that has been clean and balanced, then splitting is carried out until finally modelling is carried out using the selected algorithm, namely random forest with various types of parameter

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.14264 p-ISSN : 2541-044X

testing. Finally, of course, there needs to be an evaluation of parameter testing to see the comparison of results, which is the best and effective.

Data Source

This study raises the Indonesian Health Survey (IHS) data in 2023 which has been collected from the results of a survey conducted on the Indonesian people as a support in this study. The data has 25,800 rows with 67 features that have covered important variables for stunting prediction. The variables collected include anthropometric data such as child height and weight, gender, age, parents' socioeconomic conditions, and others. This dataset was chosen due to its relevance to the research objectives as well as the availability of sufficient data for machine learning-based analyses. However, this data has challenges such as a large number of features and class imbalance, which requires certain customisation methods to make the analysis results more representative.

Data Pre-Processing

The first stage in this research is data pre-processing. This step includes checking for missing values, where missing values are incomplete and missing datasets (Setiawan, Suryawijaya, et al., 2024), and then resolved by deleting the missing data because the number is relatively small. Next, the data is checked to identify the deleted duplicate values to ensure the integrity of the dataset. This process is followed by feature selection to filter out variables that are most relevant to the prediction target. According to (Setiawan, Nugraha, et al., 2024), ensuring that the data is clean is important in data processing. At this stage, several features used in this study were selected based on functions and variables that are considered important, including the following.

Table 1 Selected Feature

Features	Description	Features	Description
B4K4	Gender	B4K8_ibu	Mother's Highest Education
I04	Gestational Age at Birth	B4K9_ibu	Employment status of mother
I05A	Weight at Birth	B4K10_ibu	Current Pregnancy Status
I07	Length at Birth	B4K11_ibu	Health Insurance Ownership
I10	Head Circumference at Birth	B4K8_KK	Father's Highest Education
J01B	Health Condition of Toddlers at Measurement	B4K9_KK	Father's occupation
J01C	Weight	stunting	Stunting
J02B	Height	underweight	Underweight
J02C	Position When Measured	wasting	Wasting
B4K5_ibu	Married Status	overweight	Overweight

Balancing Data with SMOTE

Data imbalance is a condition where the distribution of data in a data class is unbalanced, resulting in the amount of data in one class being too much or too little compared to other data classes (Setiawan, Nugraha, et al., 2024). Overcoming class imbalance is the next focus, considering that the dataset has an uneven distribution between stunted and non-stunted children. Given the small amount of data, a random oversampling technique is used in this study, which replicates minority class data to balance with majority class data (I. K. Ananda et al., 2024). One of the oversampling methods is the Synthetic Minority Oversampling Technique (SMOTE). SMOTE aims to improve the performance of classification methods by modifying unbalanced datasets and creating new synthetic data derived from minority classes (Wicaksono et al., 2024). This process results in a more balanced class distribution and is essential to prevent bias towards the majority class in predicting stunting cases.

Modelling with Random Forest Algorithm

The cleaned and balanced dataset is then divided into training data and data to ensure the class distribution remains proportional. The prediction model is built using the random forest algorithm, which has been recognised as one of the most widely used machine learning methods in dealing with large and complex data problems (Purwati & Pristyanto, 2024).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14264
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Key parameters such as number of trees (n_estimators), tree depth (max_depth), number of leaves, and number of nodes were tested through a series of experiments.

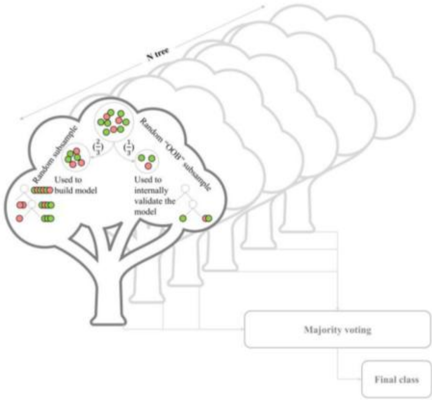

Fig. 2 Random Forest Model (Sheykhmousa et al., 2020)

In the same study (Sheykhmousa et al., 2020), it was revealed that to run a random forest model, two parameters must be set: the number of trees and the number of randomly selected features. These parameters are adjusted to find the optimal configuration that provides high accuracy and computational time efficiency.

Evaluation

This evaluation aims to determine the performance of the prediction model (Anugrah et al., 2024). To provide an in-depth picture of the model's ability to predict stunting, testing is carried out using confusion matrix and various metrics such as accuracy, precision, recall, and F1 score (Wicaksono et al., 2024).

Table 2 Confusion Matrix (Setiawan, Nugraha, et al., 2024)

Confusion Matrix		Predicted Class	
		P	N
Actual Class	P	TP (True Positive)	FN (False Negative)
	N	FP (False Positive)	TN (True Negative)

The accuracy matrix shows the number of classes accurately predicted by the model (Setiawan, Nugraha, et al., 2024). The equation is as follows:

$$accuracy = \frac{TP+TN}{\mathrm{TP+TN+FP+FN}}$$ (Wicaksono et al., 2024)

The precision metric shows the ratio of correctly predicted classes to the total number of positive classes (Setiawan, Nugraha, et al., 2024). The equation is as follows:

$$precission = \frac{\mathrm{TP}}{TP+FP}$$ (Wicaksono et al., 2024)

Recall is the model's ability to recover all true positive instances (true positive or false negative) from all actual positive instances. In other words, recall shows how well the model's ability to predict positive (actual) classes (Setiawan, Nugraha, et al., 2024). The equation for recall is as follows:

$$\mathrm{recall} = \frac{\mathrm{TP}}{TP+FN}$$ (Wicaksono et al., 2024)

The F1 score calculates the harmonic mean of the two by assessing the positive class performance to cover the natural shortcomings of precision and recall (Setiawan, Nugraha, et al., 2024). The F1 Score equation is as follows:

$$f1 - \mathrm{score} = 2 * \frac{\mathrm{precision*recall}}{precision+recall} = 2 * \frac{\mathrm{precision*recall}}{precision+recall}$$ (Wicaksono et al., 2024)

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14264
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In addition, the duration of training and testing was recorded to assess the efficiency of computing time, which is an important consideration in the application of the model. This research also compares the performance of the model based on different parameter configurations to provide the best recommendation for the development of a random forest-based prediction system in predicting stunting cases.

RESULT

Dataset

Overall, the dataset consists of 67 features, including target features that represent stunting status. However, this dataset faces the challenge of an unbalanced distribution of target feature classes, as shown in Figure 3 and Table 2 below.

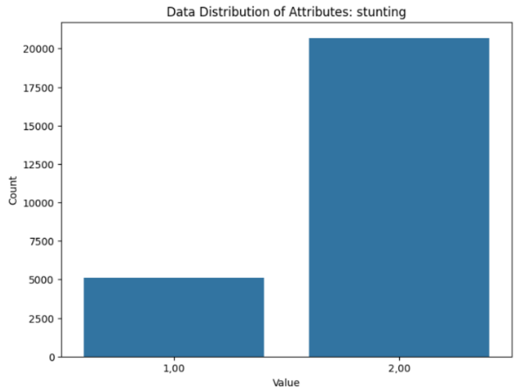

Fig 3 Distribution of stunting

Table 3 Number of each stunting class

Class	Count
1 (Stunted)	5132
2 (Non Stunted)	20688

Based on the graph analysis in Figure 3, there is a significant difference in the amount of data between class 1 and class 2 in the dataset. Class 1 has 5132 data, while class 2 has 20688 data. This imbalance can affect classification performance, as the algorithm tends to favour the class with more data during the training process (Silviana et al., 2024). To overcome this problem, special handling is required. One approach used is the oversampling method combined with the Random Forest Classifier algorithm. This step is expected to improve the class distribution in the dataset while increasing the model's ability to predict both classes more equally..

Data Pre-Processing

The first step is to detect missing values in the dataset, followed by checking for duplicate data. Based on the results of this process, it is confirmed that the dataset does not have duplicate data but there are missing values. One way to handle missing values is to delete the missing values in the dataset. Deleting missing values is more effective because the amount of data that has missing values is small (Dessiaming et al., 2022). The next thing to do is to remove features that are considered not related to stunting. As a result, 20 features were selected as stunting indicators with a total of 22,399 rows of data as shown in the following table. Based on the steps that have been taken, it is confirmed that the dataset is clean and ready for further analysis. With clean data, the analysis results are expected to be more accurate, reliable, and free from potential bias that can affect the research results..

Balancing Data with SMOTE

Before balancing the data distribution, what must be done is to convert categorical data into numerical form with a scale of 0 to 1. This will be very helpful when the data is used for analysis that requires statistical modelling or mathematical processes (Febrianty et al., 2023). After all the data is in numerical form, the data is ready to be processed at the oversampling stage. In this research, the oversampling technique used is the Synthetic Minority Oversampling Technique (SMOTE). The results of the application of the SMOTE technique can be seen in Figure 4 below:

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14264
e-ISSN : 2541-2019
p-ISSN : 2541-044X

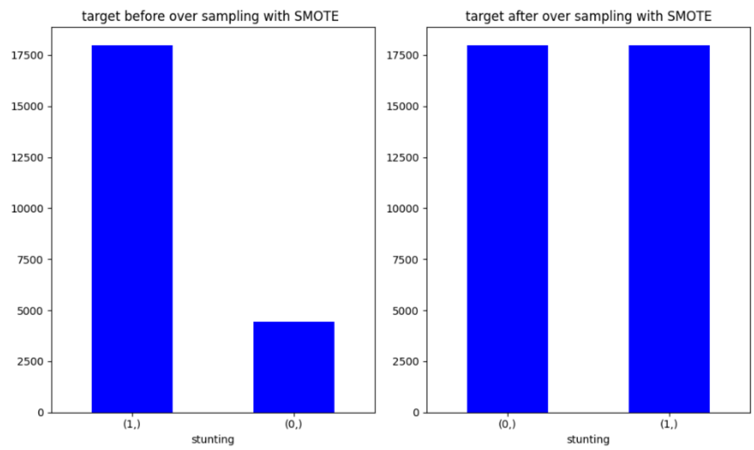

Fig 4 Balancing Data with SMOTE

Based on the analysis in Figure 4, it can be seen that SMOTE effectively balances the data. SMOTE enriches the dataset by creating synthetic observations similar to the minority data, but with slight variations. The large difference in the amount of data between the majority and minority classes has now been fairly resolved. Since both classes are balanced, there is no bias caused by class imbalance which makes the machine learning process more balanced and the model can produce more precise predictions. (D. Ananda & Suryono, 2024(Fi 202

.

Modelling with Random Forest Algorithm

The modelling stage is the main step in this research. Random Forest Classifier will be used as a machine learning algorithm in the modelling process. The selection of the Random Forest model in this study is based on the results of previous trials which show that this model has the highest accuracy rate compared to other models. This is shown more clearly through the visualisation in the following figure:

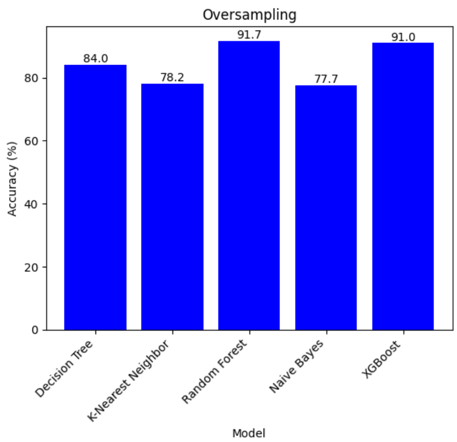

Fig 5 Model Testing

Based on the analysis of Figure 5, it can be seen that the accuracy of the Random Forest model is the highest of the other models, which is 91.7%. As in research (Putri et al., 2024) which involves a comparison of machine learning algorithms namely Naïve Bayes, KNN, and random forest, managed to get an accuracy of 87.75% through the random forest algorithm.

The modelling process begins by dividing the data into two subsets: training data (80%) and testing data (20%). The training data was used to train the Random Forest algorithm model and to learn the patterns present in the data analysed in this study. Then experiments were conducted to test the main parameters in Random Forest such as the number of trees (n_estimators), tree depth (max_depth), number of leaves, and number of nodes.

Evaluation

An evaluation of the model developed in this study was conducted to assess the performance of the Random Forest Classifier in predicting stunting status. The model was evaluated using several performance metrics, including accuracy, precision, recall, and F1-score. These metrics were chosen to provide a comprehensive overview of the model's ability to classify both stunting and non-stunting classes, as shown in the following table.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14264
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 4 Model Evaluation with Max Depth 10

n_estimator	Accuracy	Micro F1-Score	Macro Precision	Macro Recall	Macro F1-Score	Train / Test Duration	Number of Leaves	Number of Nodes
50	0.8888	0.8888	0.8921	0.8888	0.8886	14.87 seconds	469	937
100	0.8888	0.8888	0.8916	0.8888	0.8886	30.32 seconds	387	773
200	0.8886	0.8886	0.8919	0.8886	0.8883	56.71 seconds	364	727

Table 5 Model Evaluation with Max Depth 15

n_estimator	Accuracy	Micro F1-Score	Macro Precision	Macro Recall	Macro F1-Score	Train / Test Duration	Number of Leaves	Number of Nodes
50	0.9064	0.9064	0.9085	0.9064	0.9063	20.15 seconds	1979	3957
100	0.9072	0.9072	0.9092	0.9072	0.9071	36.36 seconds	1801	3601
200	0.9069	0.9069	0.9090	0.9069	0.9068	70.42 seconds	1524	3047

Table 6 Model Evaluation with Max Depth 20

n_estimator	Accuracy	Micro F1-Score	Macro Precision	Macro Recall	Macro F1-Score	Train / Test Duration	Number of Leaves	Number of Nodes
50	0.9129	0.9129	0.9144	0.9129	0.9128	20.32 seconds	2650	5299
100	0.9128	0.9128	0.9146	0.9128	0.9127	39.61 seconds	3137	6273
200	0.9142	0.9142	0.9159	0.9142	0.9141	75.20 seconds	3379	6757

Table 7 Model Evaluation with Max Depth 25

n_estimator	Accuracy	Micro F1-Score	Macro Precision	Macro Recall	Macro F1-Score	Train / Test Duration	Number of Leaves	Number of Nodes
50	0.9139	0.9139	0.9150	0.9139	0.9138	18.83 seconds	3490	6979
100	0.9130	0.9130	0.9143	0.9130	0.9130	36.97 seconds	3250	6499
200	0.9140	0.9140	0.9155	0.9140	0.9139	75.60 seconds	3234	6467

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025
DOI : https://doi.org/10.33395/sinkron.v9i1.14264
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 8 Model Evaluation with Max Depth 30

n_estimator	Accuracy	Micro F1-Score	Macro Precision	Macro Recall	Macro F1-Score	Train / Test Duration	Number of Leaves	Number of Nodes
50	0.9139	0.9139	0.9150	0.9139	0.9138	19.61 seconds	3717	7433
100	0.9136	0.9136	0.9150	0.9136	0.9135	37.82 seconds	3470	6939
200	0.9165	0.9165	0.9181	0.9165	0.9164	75.74 seconds	3515	7029

Based on the evaluation results table, it can be seen that overall the micro F1 score and macro F1 score results are almost the same in each max depth. This happens because the number of data classes between the two classes is balanced. The n_estimators parameter affects the training time. On the other hand, the depth of the tree (max_depth) also directly affects the performance of the model. The highest accuracy results were achieved with n_estimators 200 and max_depth 30, where the model achieved an accuracy of 91.65%. The prediction results with the highest accuracy can be seen in the following figure:

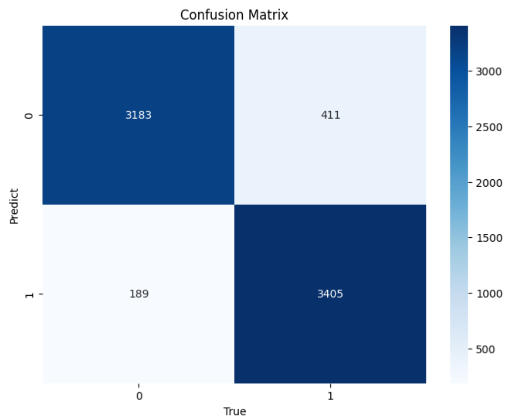

Fig 6 Confusion Matrix Best Accuracy

Based on the results of confusion matrix analysis, it can be seen that the number of correct predictions for class 0 (Stunted) is 3183, and the correct predictions for class 1 (Non Stunting) are 3405. The total data used is 7188. From these results it can be concluded that the model can predict well.

DISCUSSIONS

In this study, we have developed a Random Forest Classifier model to predict stunting status using the Indonesian Health Survey (IHS) dataset of Central Java province in 2023. One of the important findings of this research is that the use of Synthetic Minority Oversampling Technique (SMOTE) to handle class imbalance proved to be effective. By applying SMOTE, the class distribution between stunted and non-stunted children becomes more balanced. Handling class balance is very important because it improves the model's ability to recognise patterns in the data (I. K. Ananda et al., 2024). In addition, parameter settings, such as the number of trees (n_estimators) and tree depth (max_depth), have a significant influence on model performance. The n_estimators parameter affects the training time. At n_estimators of 100, the training time is about 37.82 seconds, while with n_estimators of 200, the training time increases to 75.74 seconds.

On the other hand, tree depth (max_depth) also has a direct effect on the performance of the model. With a max_depth of 10, the model achieved an accuracy of 88.88%. When max_depth was increased to 15, the accuracy increased to 90.72%. At max_depth 20, the accuracy reached 91.42%, and with max_depth 30, the accuracy reached 91.65%. The deeper the tree, the higher the accuracy, number of nodes, and number of leaves. The evaluation results showed that the highest accuracy was achieved with n_estimators 200 and max_depth 30, where the model achieved an accuracy of 91.65%.

In a study (Ramadani Akbar Ariyadi et al., 2023), that classified stunting in toddlers using Random Forest, the highest accuracy was 90.1%. The study recommended parameter tuning to improve model performance. In

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

comparison, this study achieved an accuracy of 91.65% by utilizing more optimal parameter settings. This shows that the right combination of parameter settings can significantly improve model performance. (Muhamad Malik Matin, 2023).

However, although the results obtained are quite good, there are some limitations that need to be noted. Firstly, although SMOTE was successful in class balancing, the generation of synthetic data may introduce noise that may affect the quality of the predictions. Secondly, this model only uses 20 features selected as stunting indicators, so it is possible that other relevant factors are not represented in the analysis. Therefore, further research is recommended to explore the use of more features and more sophisticated feature selection techniques.

CONCLUSION

This study successfully developed a Random Forest Classifier model to predict stunting status using the Indonesian Health Survey (IHS) dataset for Central Java province in 2023. The use of the Synthetic Minority Oversampling Technique (SMOTE) proved effective in handling class imbalance, which is important for improving the model's ability to recognize patterns in the data. The evaluation results show that the parameter settings, particularly the number of trees (n_estimators) and tree depth (max_depth), have a significant influence on the model performance. The highest accuracy was achieved with the combination of n_estimators 200 and max_depth 30, which resulted in an accuracy of 91.65%. In addition, the training time increases as the number of trees increases, at n_estimators of 100, the training time is recorded at around 37.82 seconds, while with n_estimators of 200, the training time increases to 75.74 seconds.

Although the results obtained show good performance, there are some limitations that need to be considered, such as potential noise from synthetic data and the limited number of features used. Therefore, further research is recommended to explore the use of more features and more sophisticated feature selection techniques, as well as considering other methods to handle class imbalance. Overall, the results of this study are expected to serve as a reference for researchers and practitioners in designing more effective interventions to address the problem of stunting in Indonesia, as well as supporting the government's efforts in achieving the stunting prevalence reduction target that has been set.

REFERENCES

Ananda, D., & Suryono, R. R. (2024). Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *8*(2), 748. https://doi.org/10.30865/mib.v8i2.7517

Ananda, I. K., Fanani, A. Z., Setiawan, D., & Wicaksono, D. F. (2024). Penerapan Random Oversampling dan Algoritma Boosting untuk Memprediksi Kualitas Buah Jeruk. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 282–289.

Andriyani, W., Natsir, F., Lubis, H., Tyas, S. H. Y., Meidelfi, D., Faizah, S., Nurlaida, N., Kurniawan, H., Wahyuningtyas, I., & Hasan, F. N. (2024). *PERANGKAT LUNAK DATA MINING*. Penerbit Widina.

Anugrah, M. I., Zeniarja, J., & Setiawan, D. S. (2024). Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 290–299.

Arisusanto, A., Suarna, N., & Dwilestari, G. (2023). Analisa Klasifikasi Data Harga Handphone Menggunakan Algoritma Random Forest Dengan Optimize Parameter Grid. *Jurnal Teknologi Ilmu Komputer*, *1*(2), 43–47.

Dessiaming, T. Z., Anraeni, S., & Pomalingo, S. (2022). COLLEGE ACADEMIC DATA ANALYSIS USING DATA VISUALIZATION. *Jurnal Teknik Informatika (Jutif)*, *3*(5), 1203–1212. https://doi.org/10.20884/1.jutif.2022.3.5.310

Dhani, A. A., Siswa, T. A. Y., & Pranoto, W. J. (2024). Perbaikan Akurasi Random Forest Dengan ANOVA Dan SMOTE Pada Klasifikasi Data Stunting. *Teknika*, *13*(2), 264–272. https://doi.org/10.34148/teknika.v13i2.875

Febrianty, E., Awalina, L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions. *Jurnal Teknologi Dan Informasi (JATI)*, *13*. https://doi.org/10.34010/jati.v13i2

Fitriani, F., & Darmawi, D. (2022). Hubungan Pengetahuan Dan Sikap Ibu Dengan Kejadian Stunting Pada Balita Di Desa Arongan Kecamatan Kuala Pesisir Kabupaten Nagan Raya. *Jurnal Biology Education*, *10*(1), 23–32. https://doi.org/10.32672/jbe.v10i1.4114

Kurniawan, D., Wahyudi, M., Pujiastuti, L., & Sumanto, S. (2024). Deteksi dan Prediksi Cerdas Penyakit Paru-Paru dengan Algoritma Random Fores. *Indonesian Journal Computer Science*, *3*(1), 51–56.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Number 1, January 2025 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.14264 p-ISSN : 2541-044X

Marsya Finda, S., & Wahyu Utomo, D. (2024). Klasifikasi Stunting Balita menggunakan Metode Ensemble Learning dan Random Forest. *Jl. Imam Bonjol No*, *15*(02), 287–295. https://doi.org/10.35970/infotekmesin.v15i2.2326

Melvin, J., & Soraya, A. (2023). *Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit*. *2*(2).

Muhamad Malik Matin, I. (2023). Hyperparameter Tuning Menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware. *Multinetics*, *9*(1), 43–50. https://doi.org/10.32722/multinetics.v9i1.5578

Purwati, S. E., & Pristyanto, Y. (2024). Model Random Forest and Support Vector Machine for Flood Classification in Indonesia. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *8*(4), 2261–2268.

Putri, I. P., Terttiaavini, T., & Arminarahmah, N. (2024). Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 257–265. https://doi.org/10.57152/malcom.v4i1.1078

Ramadani Akbar Ariyadi, M., Lestanti, S., Kirom, S., Balitar Jl Majapahit No, I., Sananwetan, K., Blitar, K., & Timur, J. (2023). KLASIFIKASI BALITA STUNTING MENGGUNAKAN RANDOM FOREST CLASSIFIER DI KABUPATEN BLITAR. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 6).

RI, K. K. (2022). Panduan Hari Gizi Nasional ke 64 Tahun 2024. In *Экономика Региона* (p. 32).

Ritonga, P. T., Sembiring, R. N. S., & Siburian, U. D. (2024). Pentingnya Pemantauan Status Gizi Pada Ibu Hamil Dan 1000 Hari Pertama Kehidupan Dalam Pencegahan Stunting di Desa Sangkaran Kabupaten Tapanuli Utara. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (PKM)*, *7*(7), 2822–2836.

Royhan Zaki Ramadhana, & Muhammad Irwan Padli Nasution. (2024). Analisis Dampak Penerapan Teknologi AI pada Pengambilan Keputusan Strategis dalam Sistem Informasi Manajemen. *Jurnal Ilmiah Research and Development Student*, *2*(1), 161–168. https://doi.org/10.59024/jis.v2i1.579

Setiawan, D., Nugraha, A., & Luthfiarta, A. (2024). Komparasi Teknik Feature Selection Dalam Klasifikasi Serangan IoT Menggunakan Algoritma Decision Tree. *Jurnal Media Informatika Budidarma*, *8*(1), 83–93.

Setiawan, D., Suryawijaya, T. W. E., Anugrah, M. I., Pearl, J., & Chasanah, A. N. (2024). Optimizing Banking Stock Price Prediction: Deep Learning Based Approach. *International Journal Of Accounting, Management, And Economics Research*, *2*(1), 111–125.

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325.

Silviana, S., Astuti, R., & Basysyar, F. M. (2024). PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) PADA ULASAN PENGUNJUNG WISATA KABUPATEN KUNINGAN. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).

Sitanggang, B. F., & Sitompul, P. (2024). Deteksi Awal Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Machine Learning Metode Random Forest. *Innovative: Journal Of Social Science Research*, *4*(2), 3347–3357.

Wicaksono, D. F., Basuki, R. S., & Setiawan, D. (2024). Peningkatan Performa Model Machine Learning XGBoost Classifier melalui Teknik Oversampling dalam Prediksi Penyakit AIDS. *Jurnal Media Informatika Budidarma*, *8*(2), 736–747.