Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13698
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Analysis of Machine Learning Algorithm Performance in Predicting Stunting in Toddlers

Nur'aini Indah Syahfitri[1)*], Angga Putra Juledi[2)], Rahma Muti'ah[3)]
[1,2,3)]Universitas Labuhanbatu, Indonesia
[1)]nurainiindahsyahfitri342@gmail.com, [2)]anggapj19@gmail.com, [3)]rmuthea5@gmail.com

Submitted : May 23, 2024 | **Accepted** : Jun 19, 2024 | **Published** : Jul 1, 2024

Abstract: Stunting is a condition where the growth of children and toddlers is stunted, which causes children to be shorter than they should be. In the long term, stunting can reduce reproductive health, study concentration, and work productivity, thereby causing significant state losses. The prevalence of stunting in Indonesia, which is still above 20 percent, shows that there are still chronic nutritional problems among toddlers. To prevent this from happening, identification as early as possible can be done using machine learning for predictions. The aim of this research is to conduct a comparative analysis of the performance of machine learning algorithms for predicting stunting in toddlers. Random Forest, K-Nearest Neighbors, and Extreme Gradient Boosting are the algorithms that are compared for their performance. The performance of each algorithm is measured using evaluation matrices such as accuracy, precision, recall, and f1-score. The research method starts with data collection, data preprocessing, data splitting, application of machine learning algorithms, evaluation of algorithm performance, and comparison of results. The performance evaluation matrix measurement results show that Random Forest has an accuracy of 99.95%, precision of 99.89%, recall of 99.94%, and f1-score of 99.91%. K-Nearest Neighbors has an accuracy of 99.93%, precision of 99.87%, recall of 99.88%, and f1-score of 99.88%. Meanwhile, Extreme Gradient Boosting has an accuracy of 99.36%, precision of 98.86%, recall of 98.95%, and f1-score of 98.90%. From the results of all performance evaluation matrices, it can be concluded that the random forest algorithm is the best algorithm for predicting stunting in toddlers.

Keywords: K-Nearest Neighbor; Machine Learning; Random Forest; Stunting; XGBoost.

INTRODUCTION

Toddler age is a time when children's growth and development require adequate quantity and quality of nutritional intake (Purwanti et al., 2020). Lack of nutritional intake over a long period of time can cause chronic nutritional problems, which can lead to stunting (Suratri et al., 2023). Stunting is a condition where the growth of children and toddlers is hampered, which causes children to be shorter than they should be (Riyan & Nendi, 2024). The condition of stunting in toddlers is characterized by a length or height of more than minus two standard deviations from the WHO average child growth standard (Kusrini & Laksono, 2020; Li et al., 2019; World Health Organization (WHO), 2022). Stunting in toddlers can be caused by various factors, including socio-economic conditions, maternal nutrition during pregnancy, infectious diseases, and insufficient food availability to meet the baby's nutritional

*Nur'aini Indah Syahfitri

intake. These factors can cause difficulties for babies to achieve ideal physical and cognitive development (Laksono, Sukoco, Rachmawati, & Wulandari, 2022). Stunting has short-term and long-term effects. As a short-term consequence, there is an increase in morbidity and mortality rates, developmental disorders such as cognitive, motor, and language disorders, as well as an increase in financial responsibility for the care and treatment of sick children (Das, Salam, Saeed, Bilal, & Bhutta, 2018). In the long term, it can reduce reproductive health, study concentration, and work productivity, so it will have a significant impact on state losses (Pangestuti, Khomsan, & Ekayanti, 2023; Suratri et al., 2023).

In Indonesia, the problem of stunting in children under five is still a priority program for the government as outlined in the National Medium Term Development Plan (Setianingsih, Musyarofah, PH, & Indrayati, 2022). he prevalence of stunting in Indonesia is still high, ranging from 30–39% according to WHO public health limit values (Anastasia et al., 2023). The prevalence of stunting in Indonesia tends to change from year to year. Increased from 2010 to 2013, then decreased from 2014 to 2018. The results of the Indonesian Nutrition Status Survey (SSGI) show a decrease in prevalence of 3.3% in 2021 to 24.4% (Nasional, 2023) and in 2022 to 21.6% (Rokom, 2023). Meanwhile, the national prevalence of stunting in 2023 will be 21.5% (Kesehatan, 2023). This data indicates that the prevalence of stunting in Indonesia is still at more than 20 percent, indicating that there are still chronic nutritional problems among children under five (Wardani, Sukandar, Baliwati, & Riyadi, 2021). Meanwhile, the prevalence target to be achieved by 2024 is 14% (Paudpedia, 2023). On the other hand, according to the World Bank, Indonesia has failed to reduce the level of stunting compared to upper-middle-income countries and other countries in the region (Bank, 2020).

To prevent the long-term effects of stunting on toddlers, preventive measures must be taken. This prevention can be done by identifying it as early as possible through fast and effective treatment so that the condition does not get worse (Aryuni et al., 2023). The use of machine learning in data processing is really needed by health experts to obtain information automatically to solve problems with more accurate results (Byna, 2020). Much knowledge has been produced from the use of machine learning in many cases and datasets in the health sector (Ahmed, Al-Hamadani, & Satam, 2022). Machine learning technology is very important to use to detect, diagnose, and classify various types of disease automatically through various methods (Mambang, Marleny, & Zulfadhilah, 2023).

Machine learning offers many methods that can be applied to predict stunting conditions in toddlers (Bitew, Sparks, & Nyarko, 2022; Daffa & Gunawan, 2024). These various methods have been applied in previous studies. In research comparing five types of machine learning algorithms, the best model was obtained, namely the gradient boosting algorithm, with test accuracy results of 79.33% (Ndagijimana, Kabano, Masabo, & Ntaganda, 2023). Tests using the bagging method and random forest algorithm produced an accuracy of 91.98% (Juwariyem, Sriyanto, Lestari, & Chairani, 2024). In research conducted by (Reza & Rohman, 2024), the random forest algorithm had relatively high accuracy, amounting to 90.7%. The results of research by (Sylivester, 2020) show that random forest is the best classifier for predicting stunting in toddlers, with an accuracy score of 83% and an under-the-curve score of 92%. Based on the results of research (Lonang & Normawati, 2022), the average accuracy produced by the k-nearest neighbor algorithm is 91.90%. In research conducted by (Lonang, Yudhana, & Biddinika, 2023), the K-Nearest Neighbor model was superior with accuracy results of 94.85% and an area under curve value of 0.924 compared to five other machine learning algorithms. The classification model produced by the Extreme Gradient Boosted Trees (XGBoost) algorithm succeeded in identifying stunting toddlers with an accuracy rate of 95.9% (Azis, 2023). Research results in predicting toddler stunting show that the XGBoost algorithm has the best performance with an accuracy score of 72.8% (Shen, Zhao, & Jiang, 2023).

In the previous section, various types of machine learning algorithms used by researchers were described for predicting stunting in toddlers. From this description, the focus of this research is to compare the performance of the random forest, k-nearest neighbor, and xgboost algorithms in predicting stunting in toddlers. This research is important to do to find out which algorithm is the best of the three algorithms applied in predicting stunting. The best algorithm can be used as a model to predict stunting cases in Indonesia, so that stunting prevention can be done early and will have an impact on reducing the stunting prevalence rate in Indonesia. The difference between this research and previous research lies in the independent and dependent variables used. This research uses three independent variables, namely age, gender, and length or height of toddlers. This variable was chosen because it has a very significant contribution to predicting stunting in toddlers (Azis, 2023). Meanwhile, the dependent variable used is stunting status in toddlers based on the height achieved at a certain age. The categories of stunting status in toddlers based on height, according to the WHO (Laksono et al., 2022) and the Ministry of Health (Kesehatan, 2023), are severally stunted, stunted, and normal. Meanwhile, previous research (Juwariyem et al., 2024; Lonang & Normawati, 2022; Lonang et al., 2023; Reza & Rohman, 2024; Shen et al., 2023; Sylivester, 2020) only used two categories of stunting status, namely stunted and not stunted.

This research aims to conduct a comparative analysis of the performance of machine learning algorithms in predicting stunting in toddlers. The algorithms whose performance was compared were Random Forest, K-Nearest Neighbors, and XGBoost. The random forest algorithm was chosen because it can reduce overfitting in the data, so it can effectively predict stunting (Darnila, Maryana, Mawardi, Sinambela, & Pahendra, 2022). K-Nearest Neighbor was chosen because it has consistent and balanced performance, making it suitable for sensitive classification tasks such as stunting problems (Lonang et al., 2023). Meanwhile, the XGBoost algorithm can handle complex prediction problems and process very large amounts of data. Thus, this method can help find important patterns or components that cause stunting in toddlers more efficiently and accurately (Azis, 2023). To measure the performance of each algorithm, a performance evaluation matrix such as accuracy, precision, recall, and f1-score is used. From the evaluation matrix scores obtained for each algorithm, it can be determined which algorithm is the best at predicting stunting in toddlers.

METHOD

This section discusses the research stages used to predict stunting in toddlers by applying the random forest, k-nearest neighbor, and xgboost algorithms. Figure 1 shows the research stages starting from dataset collection, data preprocessing, data separation, application of machine learning prediction methods, evaluation of algorithm performance, and comparison of results. The sources used in selecting this method came from research conducted by (Lonang et al., 2023) and (Gustriansyah, Suhandi, Puspasari, & Sanmorino, 2024).

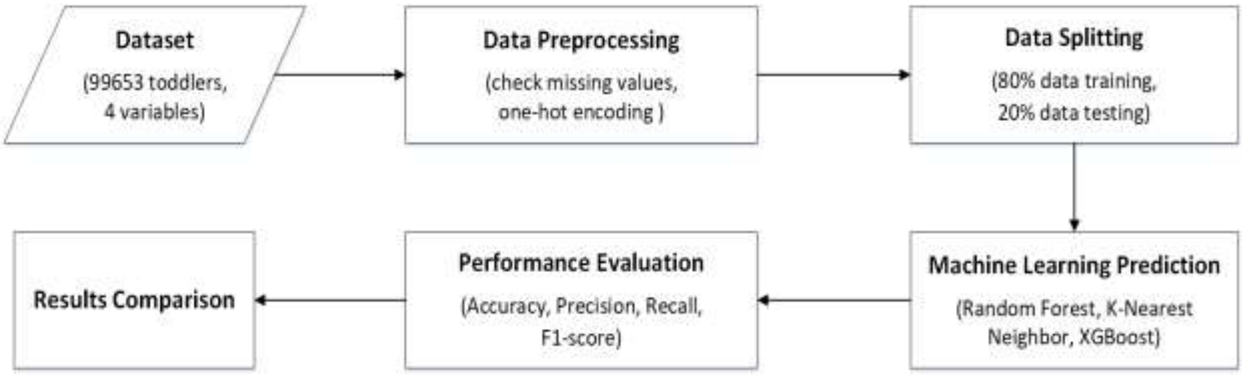


Fig 1. Research stages

Dataset

The dataset used in this research was sourced from the Kaggle website (Pradana, 2024). This dataset is based on the WHO z-score formula, which focuses on detecting stunting in toddlers. This dataset consists of 121,000 rows of data with four variables, namely: age, gender, height, and stunting status of toddlers. In this study, only 99,653 rows of data were used, after subtracting data on the age of toddlers who were more than 59 months old and data on nutritional status in the high category.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13698
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1. Sample of Dataset

No	Age (month)	Gender	Height (cm)	Status
1	0	Male	46.86	Normal
2	6	Female	59.60	Stunted
3	12	Male	70.70	Stunted
4	18	Female	68.10	Severely stunted
5	24	Female	79.00	Stunted
6	30	Male	91.90	Normal
7	36	Female	85.40	Stunted
8	42	Female	85.50	Severely stunted
9	48	Male	92.60	Stunted
10	50	Male	92.50	Stunted
11	59	Female	93.90	Severely stunted

Table 1 shows a sample research dataset. The age column shows the age of the toddler in months, with an age range of 0 to 59 months. To determine a child's growth phase and compare it with healthy growth standards, age is important. There are two categories in the gender column, namely male and female. In assessing growth patterns and the risk of stunting, gender is an important factor. Height is measured in centimeters. The height column is an important indicator for assessing the physical growth of toddlers. With this data, researchers can find out whether the child's growth is in accordance with age standards. The status column is categorized into three statuses: "severely stunted," "stunted," and "normal". Severely stunted indicates a very serious stunting condition (< -3.0 SD), stunted indicates a stunting condition (-3.0 SD to -2.0 SD), while normal status indicates a good nutritional condition (≥ -2 SD) (Kesehatan, 2023; Laksono et al., 2022).

Data Preprocessing

The check missing values stage is important to do to clean the dataset from "null" data. If there are missing values in the dataset, then the data cleaning stage is carried out. The one-hot encoding stage will help in processing categorical variables that do not have ordinal relationships, where differences between values have no meaning. In this case, categorical variables such as "gender" will be converted into binary form (0 and 1).

Data Splitting

At this stage, the dataset will be divided into training data and testing data. This research applies a division of training data of 80% and testing data of 20%. This process involves using functions like train_test_split from the scikit-learn library in Python to split the data in a random and proportional manner between the training and test sets.

Machine Learning Prediction

At this stage, the three machine learning algorithms will be implemented to make predictions. By using a combination of individual predictions provided by each tree in its group, the Random Forest algorithm can produce predictions. Random Forest can handle various types of data, including numeric

*Nur'aini Indah Syahfitri

and categorical data, in addition to having the ability to handle large and complex datasets. K-Nearest Neighbor, in this case, predicts the sample class based on the majority of its nearest neighbor classes. This algorithm works by calculating the distance between a sample and its nearest neighbors, with k = the number of previous neighbors. In this context, XGBoost is used to predict data classes using its features. With the ability to overcome overfitting, optimize loss functions, and produce good results in various types of datasets, this algorithm is very effective.

Performance Evaluation

To evaluate the performance of the prediction model, measurements were carried out using the accuracy, precision, recall, and f1-score evaluation matrices. Accuracy is used to measure how close the predicted results are to the actual value. Precision describes the accuracy between the requested data and the prediction results produced by the model. Recall is used to provide an overview of the model's success in retrieving information. The F1-score describes the comparison of weighted average precision and recall.

Results Comparison

The final process of this research is to compare the accuracy results of each machine learning algorithm used. From these results, it can be determined which algorithm is the best for making predictions.

RESULT

In this research, there are three types of machine learning prediction algorithms used, namely Random Forest, K-Nearest Neighbors, and Extreme Gradient Boosted Trees (XGBoost). Each prediction model is evaluated using a confusion matrix to measure metrics such as accuracy, precision, recall, and f1-score. After that, the performance of each algorithm is compared to find out which one gives the best results. The programming language used from the data preprocessing stage to the comparison of results is Python, using the Jupyter Notebook interface on Google Colab.

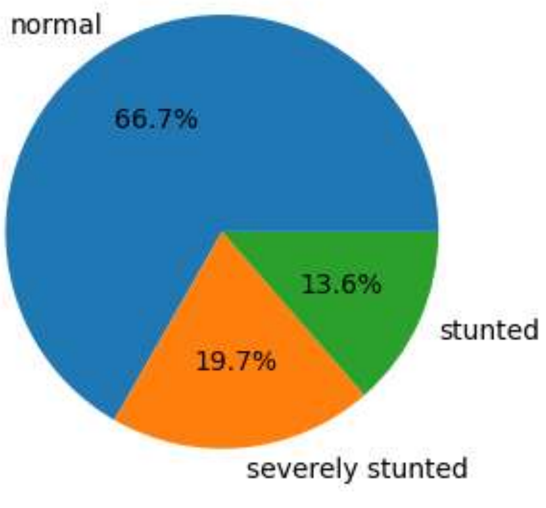


Fig 2. Distribution of stunting based on dataset

Figure 2 shows the distribution of stunting in the dataset, which is categorized based on stunting status in toddlers. From a dataset of 99,653 records, 19,632 (19.7%) toddlers were categorized as severely stunted. Then, as many as 13,539 (13.6%) toddlers were stunted, and 66,482 (66.7%) toddlers were categorized as normal. From this dataset, more toddlers are in normal condition. However, 33.3% of toddlers still experience stunting.

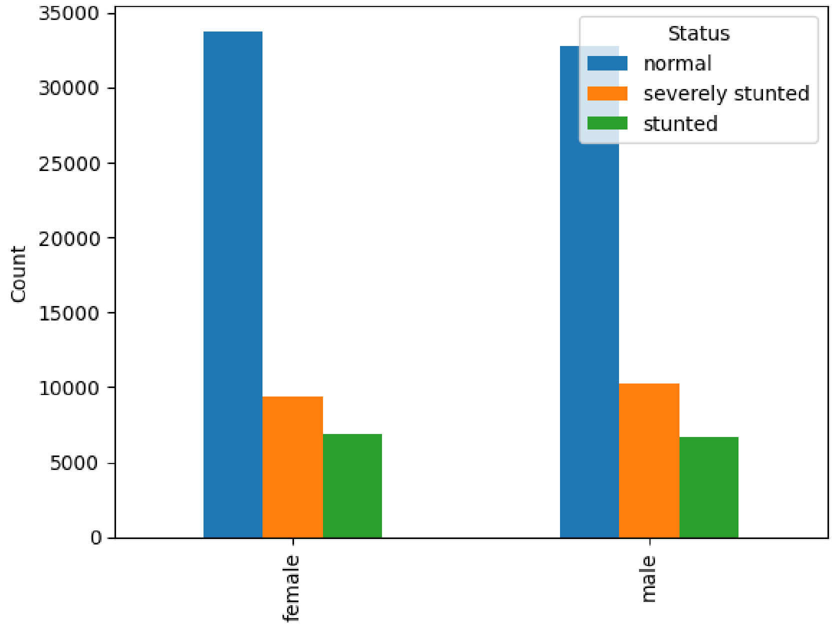


Fig 3. Stunting based on gender

Figure 3 shows the distribution of stunting in the dataset based on the gender of the toddler. From the graph, it can be seen that 9377 girls under five were severely stunted, 6856 were stunted, and 33725 were in normal condition. Meanwhile, 10,255 male toddlers were severely stunted, 6,681 were stunted, and 32,747 were in normal condition. From this graph, there are more male toddlers who experience stunting compared to female toddlers.

From the results of checking for missing values in the dataset, no missing data was found. This section is very important to ensure whether the data is suitable for use in a prediction model. Based on these results, there is no need to carry out a data cleaning process. Meanwhile, the results of the one-hot encoding stage are shown in Figure 4.

	Age	Gender	Height
41441	27	male	88.90
74202	45	female	86.40
1550	1	male	55.82
72486	44	female	87.10
73564	45	female	104.50

before One-hot encoding

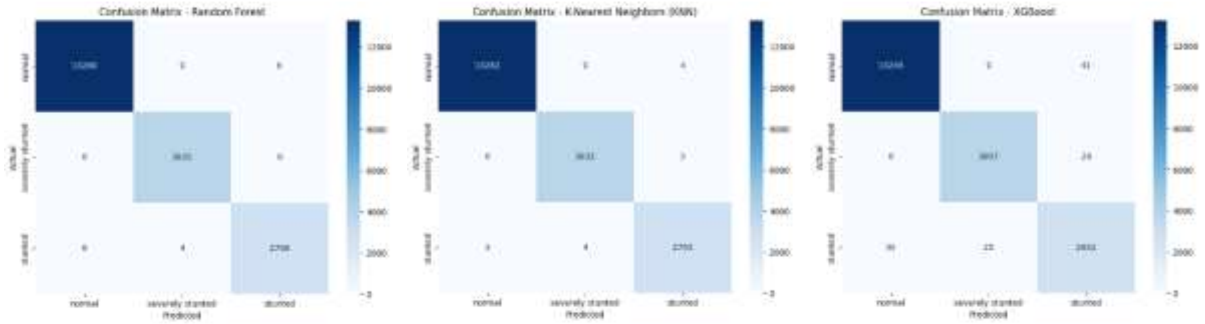
	Age	Gender	Height
41441	27	1.0	88.90
74202	45	0.0	86.40
1550	1	1.0	55.82
72486	44	0.0	87.10
73564	45	0.0	104.50

after One-hot encoding

Fig 4. One-hot encoding results

Figure 4 shows the results of the one-hot encoding process. The gender variable, which is categorical, has been converted into binary form 0 and 1, where 1 is for male and 0 for female. In machine learning analysis or modeling, one-hot encoding can assist in the processing and utilization of categorical variable data such as "gender." In the data splitting process, which applies a data division of 70% for training data and 30% for testing data, 79,722 records are produced for training data and 19,931 records for testing data.

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



(a) Random Forest (b) K-Nearest Neighbors (c) XGBoost

Fig 5. Confusion Matrix

Figure 5 (a) shows the confusion matrix results from the random forest algorithm. In the normal class, it is generated: true positives were 13280, false positives were 6, false negatives were 0, and true negatives were 6645. For the severely stunted class, the results were: true positives were 3935, false positives were 0, false negatives were 4, and true negatives were 15992. Meanwhile, for the stunted class, the results were: 2706 true positives, 4 false positives, 6 false negatives, and 17215 true negatives.

Figure 5 (b) shows the confusion matrix results from the k-nearest neighbors algorithm. In the normal class, it is generated: there were 13282 true positives, 4 false positives, 3 false negatives, and 6642 true negatives. For the severely stunted class, the results were: 3932 true positives, 3 false positives, 4 false negatives, and 15992 true negatives. Meanwhile, for the stunted class, the results were: 2703 true positives, 7 false positives, 7 false negatives, and 17214 true negatives.

Figure 5 (c) shows the confusion matrix results from the XGBoost algorithm. In the normal class, it is generated; there were 13,245 true positives, 41 false positives, 33 false negatives, and 6612 true negatives. For the severely stunted class, the results were: 3907 true positives, 28 false positives, 25 false negatives, and 15971 true negatives. Meanwhile, for the stunted class, the results were: 2652 true positives, 58 false positives, 69 false negatives, and 17152 true negatives.

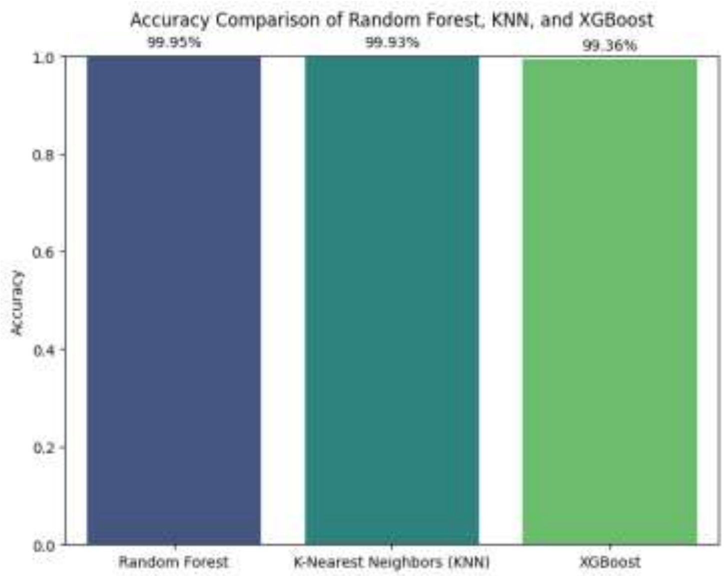


Fig 6. Accuracy Comparison

Figure 6 shows a graph comparing the accuracy results of three machine learning algorithms after testing. The random forest algorithm has the highest accuracy compared to the other two algorithms, namely 99.95%. Followed by the k-nearest neighbors algorithm with an accuracy rate of 99.93%. Meanwhile, the XGBoost algorithm is in last place with an accuracy rate of 99.36%.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, July 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13698
e-ISSN : 2541-2019
p-ISSN : 2541-044X

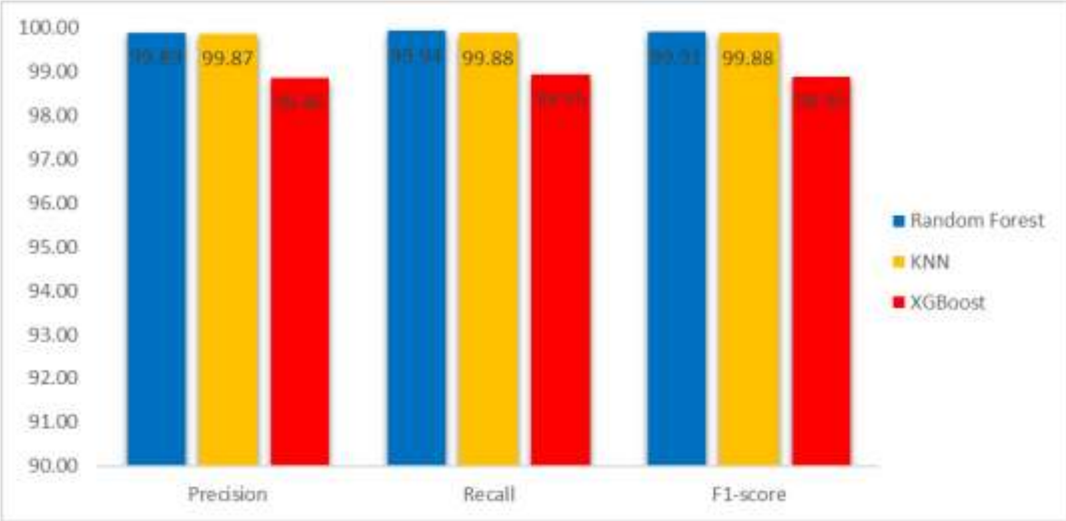


Fig 7. Precision, Recall, and F1-score

Figure 7 shows the results of the performance evaluation matrix based on the precision, recall, and f1-score obtained from the random forest, k-nearest neighbors, and xgboost algorithms. The random forest algorithm has a precision score of 99.89%, recall of 99.94%, and f1-score of 99.91%. The k-nearest neighbors algorithm has a precision score of 99.87%, recall of 99.88%, and f1-score of 99.88%. Meanwhile, the xgboost algorithm has a precision score of 98.86%, recall of 98.95%, and f1-score of 98.90%. From this graph, the random forest algorithm has a higher precision, recall, and f1-score, followed by the k-nearest neighbors algorithm and XGBoost.

DISCUSSIONS

The measurement results based on the accuracy matrix show that the random forest algorithm is superior to other algorithms, with an accuracy score of 99.95%. This accuracy score shows that the random forest model is very accurate in predicting stunting in toddlers. This score also shows that as many as 19921 actual datasets were predicted correctly from a total of 19931 test datasets. Meanwhile, the accuracy error rate was only 0.05%, that is, as many as 10 actual data points were predicted incorrectly. These results are in accordance with research (Darnila et al., 2022) showing that the random forest algorithm is effective in predicting stunting. The k-nearest neighbors algorithm obtained an accuracy score of 99.93%. The scores obtained for the other matrices show balanced results, where the percentage differences are relatively small. These results are in accordance with research (Lonang et al., 2023) showing that k-nearest neighbors have consistent performance. An accuracy score of 99.93% on the k-nearest neighbors algorithm shows that 19917 actual data were predicted correctly from a total of 19931 test datasets, with an accuracy error rate of 0.07%; that is, 14 actual data were predicted incorrectly. Meanwhile, the XGBoost algorithm obtained an accuracy score of 99.36%. The accuracy score shows that 19,804 actual data were predicted correctly from a total of 19,931 test datasets, with an accuracy error rate of 0.64%; that is, 127 actual data were predicted incorrectly. Whether the XGBoost algorithm can find patterns in the causes of stunting in toddlers, according to research conducted by (Azis, 2023), has not been proven in this study.

Overall, based on the results of the performance evaluation matrix, a score above 90% was obtained for the entire algorithm. This indicates that the three algorithms work well in predicting stunting in toddlers. This research has limitations, including that it did not test the independent variable features that influence the results of stunting status in toddlers, as was done by (Lonang & Normawati, 2022) by applying feature selection to eliminate features that have correlation values. that are very small.

CONCLUSION

From the tests that have been carried out, the results of the accuracy, precision, recall, and f1-score evaluation matrix measurements have been obtained for each random forest using the k-nearest neighbor algorithm. The accuracy results obtained were 99.95% for random forest, 99.93% for k-nearest neighbors, and 99.36% for XGBoost. Based on the accuracy results, the random forest algorithm is more accurate than other algorithms. The precision results obtained show numbers of 99.89% for random forest, 99.87% for k-nearest neighbors, and 98.86% for XGBoost. From these results, the random forest algorithm is more precise than other algorithms. Meanwhile, the recall results show numbers of 99.94% for random forest, 99.88% for k-nearest neighbors, and 98.95% for XGBoost. From these results, the random forest algorithm is more successful in retrieving information. And the f1-score results show the numbers: 99.91% for random forest, 99.88% for k-nearest neighbors, and 98.90% for XGBoost. From these results, the random forest algorithm has a harmonious average between precision and recall results. From the results of all performance evaluation matrices, it can be concluded that the random forest algorithm is the best algorithm for predicting stunting in toddlers. For further research, it is recommended to add other features that influence stunting, then carry out a feature selection process to see the correlation.

REFERENCES

Ahmed, A. H., Al-Hamadani, M. N. A., & Satam, I. A. (2022). Prediction of COVID-19 disease severity using machine learning techniques. *Bulletin of Electrical Engineering and Informatics*, *11*(2), 1069–1074. https://doi.org/10.11591/eei.v11i2.3272

Anastasia, H., Hadju, V., Hartono, R., Samarang, Manjilala, Sirajuddin, … Atmarita. (2023). Determinants of stunting in children under five years old in South Sulawesi and West Sulawesi Province: 2013 and 2018 Indonesian Basic Health Survey. *PLOS ONE*, *18*(5), e0281962. Retrieved from https://doi.org/10.1371/journal.pone.0281962

Aryuni, M., Miranda, E., Kumbangsila, M., Richard, Zakiyyah, A. Y., Sano, A. V. D., & Bhatti, F. M. (2023). Comparison of Nutritional Status Prediction Models of Children Under 5 Years of Age Using Supervised Machine Learning. In T. Triwiyanto, A. Rizal, & W. Caesarendra (Eds.), *Proceeding of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics* (pp. 265–277). Singapore: Springer Nature Singapore.

Azis, M. N. (2023). *Klasifikasi Balita Stunting berdasarkan Karakteristik Balita dan Keluarga menggunakan Extreme Gradient Boosted Trees Algorithm* (Universitas Jember). Universitas Jember. Retrieved from https://repository.unej.ac.id/handle/123456789/119017

Bank, W. (2020). Spending Better to Reduce Stunting in Indonesia. In *Public Expenditure Review*. World Bank. https://doi.org/doi:10.1596/34196

Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, *25*(2), 269–280. https://doi.org/10.1017/S1368980021004262

Byna, A. (2020). Comparative Analysis of Machine Learning Algorithms for classification about Stunting Genesis. *Proceedings of the First National Seminar Universitas Sari Mulia, NS-UNISM 2019*. European Alliance for Innovation (EAI). https://doi.org/10.4108/eai.23-11-2019.2298349

Daffa, M. G., & Gunawan, P. H. (2024). Stunting Classification Analysis for Toddlers in Bojongsoang: A Data-Driven Approach. *2024 2nd International Conference on Software Engineering and Information Technology (ICoSEIT)*, 42–46. https://doi.org/10.1109/ICoSEIT60086.2024.10497515

Darnila, E., Maryana, M., Mawardi, K., Sinambela, M., & Pahendra, I. (2022). Supervised Models to Predict the Stunting in East Aceh. *International Journal of Engineering, Science and Information Technology (IJESTY)*, *2*(3), 33–39. https://doi.org/10.52088/ijesty.v2i3.280

*Nur'aini Indah Syahfitri

Das, J. K., Salam, R. A., Saeed, M., Bilal, H., & Bhutta, Z. A. (2018). PROTOCOL: Effectiveness of interventions to manage acute malnutrition in children under five years of age in low- and middle-income countries: a systematic review. *Campbell Systematic Reviews*, *14*(1), 1–26. https://doi.org/10.1002/CL2.193

Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine Learning Method to Predict the Toddlers' Nutritional Status. *Jurnal Infotel*, *16*(1), 32–43. https://doi.org/10.20895/infotel.v15i4.988

Juwariyem, Sriyanto, Lestari, S., & Chairani. (2024). Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, *8*(2), 947–955.

Kesehatan, B. K. P. (2023). *Survei Kesehatan Indonesia (SKI) Dalam Angka: Data Akurat Kebijakan Tepat*. Jakarta. Retrieved from https://drive.google.com/file/d/1SAomJxUTXwlSzsRrGJfRPxzV3ZzypaRU/view

Kusrini, I., & Laksono, A. D. (2020). Regional Disparities of Stunted Toddler in Indonesia. *Indian Journal of Forensic Medicine & Toxicology*, *14*(3), 1952–1956. https://doi.org/doi.org/10.37506/ijfmt.v14i3.10706

Laksono, A. D., Sukoco, N. E. W., Rachmawati, T., & Wulandari, R. D. (2022). Factors Related to Stunting Incidence in Toddlers with Working Mothers in Indonesia. *International Journal of Environmental Research and Public Health*, *19*(17). https://doi.org/10.3390/ijerph191710654

Li, C., Zeng, L., Wang, D., Allen, S., Jaffar, S., Zhou, J., … Yan, H. (2019). Growth in early life and physical and intellectual development at school age: a cohort study. *British Journal of Nutrition*, *121*(8), 866–876. https://doi.org/DOI: 10.1017/S0007114519000060

Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, *6*(1), 49–56. https://doi.org/10.30865/mib.v6i1.3312

Lonang, S., Yudhana, A., & Biddinika, M. K. (2023). Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting. *Media Informatika Budidarma*, *7*(4), 2109–2117. https://doi.org/10.30865/mib.v7i4.6553

Mambang, Marleny, F. D., & Zulfadhilah, M. (2023). Prediction of linear model on stunting prevalence with machine learning approach. *Bulletin of Electrical Engineering and Informatics*, *12*(1), 483–492. https://doi.org/10.11591/eei.v12i1.4028

Nasional, B. R. dan I. (2023). BRIN Ungkap Prevalensi Stunting di Indonesia Cenderung Fluktuatif. Retrieved from brin.go.id website: https://www.brin.go.id/news/116962/brin-ungkap-prevalensi-stunting-di-indonesia-cenderung-fluktuatif

Ndagijimana, S., Kabano, I. H., Masabo, E., & Ntaganda, J. M. (2023). Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques. *Journal of Preventive Medicine & Public Health*, *56*(1), 41–49. https://doi.org/doi.org/10.3961/jpmph.22.388

Pangestuti, M., Khomsan, A., & Ekayanti, I. (2023). Determinants of stunting in children aged 6-24 months in rural areas: Case Control Study. *AcTion: Aceh Nutrition Journal*, *8*(3), 318–330. https://doi.org/dx.doi.org/10.30867/action.v8i3.918

*Nur'aini Indah Syahfitri

Paudpedia. (2023). 149 Juta Anak di Dunia Alami Stunting Sebanyak 6,3 Juta di Indonesia, Wapres Minta Keluarga Prioritaskan Kebutuhan Gizi. Retrieved from kemdikbud.go.id website: https://paudpedia.kemdikbud.go.id/berita/149-juta-anak-di-dunia-alami-stunting-sebanyak-63-juta-di-indonesia-wapres-minta-keluarga-prioritaskan-kebutuhan-gizi?do=MTY2NC01YjRhOGZkNA==&ix=MTEtYmJkNjQ3YzA=

Pradana, R. P. (2024). The Stunting Toddler (Balita) with 121K Dataset. Retrieved from Kaggle website: https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows

Purwanti, R., Diananingrum, I., Azni, H., Savitri, R. A., Rahmarani, H., & Febrianah, N. (2020). Program Pendampingan Keluarga Balita Gizi Kurang di Wilayah PUSKESMAS Karanganyar Kota Semarang. *Wikrama Parahita*, *4*(2). https://doi.org/http://dx.doi.org/10.30656/jpmwp.v4i2.1976

Reza, A. A., & Rohman, M. S. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *JITE (Journal of Informatics and Telecommunication Engineering)*, *7*(2), 534–544. https://doi.org/10.31289/jite.v7i2.10628

Riyan, A. B., & Nendi, I. (2024). K-Means Algorithm for District/City Classification in West Java Based on Stunting Toddler Data. *Journal Transnational Universal Studies*, *02*(2), 60–65.

Rokom. (2023). Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%. Retrieved from Kemenkes website: https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230125/3142280/prevalensi-stunting-di-indonesia-turun-ke-216-dari-244/

Setianingsih, Musyarofah, S., PH, L., & Indrayati, N. (2022). Tingkat Pengetahuan Kader dalam Upaya Pencegahan Stunting. *Jurnal Ilmu Keperawatan Jiwa*, *5*(3), 447–454.

Shen, H., Zhao, H., & Jiang, Y. (2023). Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children*, Vol. 10. https://doi.org/10.3390/children10101638

Suratri, M. A. L., Putro, G., Rachmat, B., Nurhayati, Ristrini, Pracoyo, N. E., … Raharni. (2023). Risk Factors for Stunting among Children under Five Years in the Province of East Nusa Tenggara (NTT), Indonesia. *International Journal of Environmental Research and Public Health*, *20*(2). https://doi.org/10.3390/ijerph20021640

Sylivester, L. L. (2020). *Predicting stunting status among children under five years: The case study of Tanzania* (University of Rwanda). University of Rwanda. Retrieved from http://dr.ur.ac.rw/handle/123456789/1389

Wardani, Z., Sukandar, D., Baliwati, Y. F., & Riyadi, H. (2021). Sebuah Alternatif: Indeks Stunting Sebagai Evaluasi Kebijakan Intervensi Balita Stunting di Indonesia. *Gizi Indonesia: Journal of The Indonesian Nutrition Association*, *44*(1), 21–30. https://doi.org/10.36457/gizindo.v44i1.535

World Health Organization (WHO). (2022). Stunting prevalence among children under 5 years of age (%) (model-based estimates). Retrieved from The Global Health Observatory website: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-stunting-prevalence