Performance Comparison of Random Forest, SVM, and XGBoost Algorithms with SMOTE for Stunting Prediction

Maulana As'an Hamid [1]*, Egia Rosi Subhiyakto [2]**
* Teknik Informatika, Universitas Dian Nuswantoro, Semarang
** Research Center for Intelligent Distributed Surveilance and Security (IDSS), Universitas Dian Nuswantoro, Semarang
111202113819@mhs.dinus.ac.id [1], egia@dsn.dinus.ac.id [2]

Article Info

Article history:

Received 2025-06-02
Revised 2025-06-21
Accepted 2025-07-03

Keyword:

Comparison,
Machine Learning,
Prediction,
Stunting,
SMOTE.

ABSTRACT

Stunting is a growth and development disorder caused by malnutrition, recurrent infections, and lack of psychosocial stimulation in which a child's length or height is shorter than the growth standard for their age. With a prevalence of 21.5% in Indonesia by 2023, stunting is a global health problem that requires technology-based detection approaches for early intervention. This study evaluates the performance of three machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost) in predicting childhood stunting, and applying the SMOTE technique to handle data imbalance. The evaluation results show that XGBoost with SMOTE achieves the best performance with 87.83% accuracy, 85.75% precision, 91.59% recall, and 88.57% F1-score, superior to RF (84.56%) and SVM (68.59%). These results show that the combination of XGBoost and SMOTE is the best solution for an accurate and effective machine learning-based stunting detection system, so it can be used in public health programs to accelerate stunting risk identification.

I. INTRODUCTION

Stunting is a disorder of child growth and development caused by malnutrition, repeated infections, and lack of psychosocial stimulation [1]. A child is diagnosed with stunting when their length or height is shorter than the growth standard for their age [2].

Apart from inadequate nutrition received by children, several other direct and indirect factors contribute to stunting, including exclusive breastfeeding, birth weight, parental education and occupation, and family economic status [3]. Children who experience stunting typically suffer impacts on their immune system, making them susceptible to disease, having non-ideal body shapes, and becoming less productive in adulthood [4][5].

In Indonesia, the prevalence of stunting stands at 21.5% in 2023, which has not yet met the government's target of reducing it to 14% by the end of 2024 [6][7]. This target indicates that reducing stunting rates in Indonesia has become a priority, although many challenges remain to achieve this goal. A comprehensive approach is needed to adequately address stunting, including accurately identifying children at risk of stunting for faster intervention [8]. However, detecting stunting using manual methods, such as height and weight measurements at community health centers or clinics, requires considerable time and medical personnel. Therefore, a technology-based approach, such as machine learning, is expected to provide faster, more efficient, and more accurate solutions in predicting stunting risk in children.

With technological advances, machine learning algorithms are becoming increasingly popular in healthcare. Many studies have used machine learning to address various health problems. In a study conducted by Abdul Mizwar comparing various algorithms to predict HIV/AIDS, the results showed that Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) algorithms had the highest accuracy at 98% [9]. Additionally, in a study by Herliyani Hasanah comparing Support Vector Machines (SVM) and C4.5 algorithms for predicting heart disease, SVM excelled with an accuracy rate of 87%. [10].

In the context of stunting, several studies have used machine learning to predict stunting risk in children. For example, a study by Indah Pratiwi Putri comparing Naïve Bayes, K-Nearest Neighbor (K-NN), and RF algorithms for predicting stunting showed that RF had the highest accuracy at 87.75% [11]. Another study by Tri Sugihartono compared XGBoost, RF, SVM, and K-NN algorithms for detecting stunting adn showed that XGBoost performed best with an accuracy of 85.74%. [12]. Additionally, research by Nada Rizki Febriyanti compared machine learning algorithms for predicting stunting, comparing SVM, RF, and Logistic Regression algorithms, with SVM performing best with an accuracy rate of 92% [13].

This demonstrates that machine learning algorithms such as RF, SVM, and XGBoost are highly effective in healthcare, especially in predicting stunting. Although these three algorithms have proven effective, data imbalance is one of the biggest challenges to address. This occurs because the number of samples for minority classes is often smaller than that for majority classes. To overcome this problem, the SMOTE (Synthetic Minority Over-sampling Technique) is often used; this technique aims to increase the number of minority class samples by generating synthetic data based on existing class distributions [14].

This research aims to compare the performance of three machine learning algorithms: RF, SVM, and XGBoost in predicting stunting in children. Additionally, this study aims to evaluate the SMOTE technique in handling data imbalance problems and to determine whether using SMOTE can improve stunting prediction performance compared to models that do not use SMOTE. The data used in this study include age, birth weight and height, weight and height at examination, breastfeeding status, and stunting status of the child. It is expected that the results of this study will provide a better understanding of how effective each algorithm is in predicting stunting, as well as how the SMOTE technique can be used to improve model accuracy on imbalanced datasets.

The benefit of this research is that it helps health systems detect stunting more quickly and accurately. Using machine learning, it is hoped that children at risk of stunting can be more precisely identified so that necessary interventions can be carried out earlier. This research can also contribute to developing technology-based prediction methods in healthcare, particularly to address imbalanced data problems that frequently occur in health research. This research is expected to serve as a basis for further research and for implementing better prediction systems in the field.

II. METHODS

This research was conducted through several systematic stages. The first stage of this research is data collection. This is followed by the data pre-processing stage, where data cleaning, normalization, and balancing are performed. After that, the data that has gone through the pre-processing stage will be divided into training and testing sets. After being divided, the data will enter the modelling stage, which is carried out using three algorithms: Random Forest (RF), Support Vector Machines (SVM), and eXtreme Gradient Boosting (XGBoost). In the final stage, the modelling results will be evaluated using a confusion matrix and accuracy, precision, recall, and F1-Score metrics to determine the best algorithm for predicting stunting. All stages conducted in this research can be seen in Figure 1.

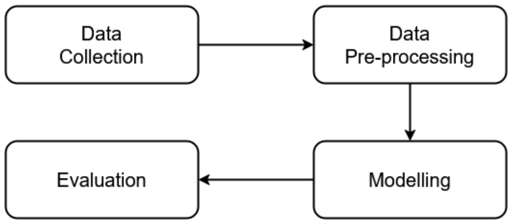


Figure 1. Research Stages

A. Data Collection

In this study, the data used comes from public data on the Kaggle website titled "factor stunting" by Harnelia, which was uploaded one year ago. This dataset contains data on children totalling 10,000 rows and 8 columns (7 features and 1 target). The features in the dataset include: child’s gender (Gender), child’s age at examination (Age), child’s weight at birth (Birth Weight), child’s height at birth (Birth Length), child’s weight at examination (Body Weight), child’s height at examination (Body Length), breastfeeding status (Breastfeeding), and the target in this dataset is stunting status (Stunting).

B. Data Pre-processing

The data pre-processing stage transforms raw data and cleans it to prepare it for use in modelling. The first step in this stage is the selection of features that contribute little to model performance. After that, the dataset is cleaned of missing, duplicate, and outlier data, so the data is more optimal during modelling. Then, data balancing is carried out using the SMOTE technique to balance the number of data points in each target class. The final step is to normalize the data using a Standard Scaler to change the scale of features so that all features have the same scale.

C. Modelling

After data pre-processing, the data is divided into training and testing data for modelling using machine learning algorithms. In this study, three algorithms are used for modelling, including Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). In this stage, data is trained to recognize patterns so that it can recognize relationships that exist in the data [15]. Then, the trained model will be evaluated.

D. Evaluation

After that, the final stage in this research is to evaluate the three algorithms using a confusion matrix. A confusion matrix functions to measure the performance of machine learning models trained with actual data. This matrix consists of four main components, as shown in Table 1.

TABLE I
CONFUSION MATRIX

	Actual Positive (+)	**Actual Negative (-)**
Predicted Positive (+)	True Positive (TP)	False Positive (FP)
Predicted Negative (-)	False Negative (FN)	True Negative (TN)

True Positive (TP) and True Negative (TN) indicate that the predictions made by the model are correct, while False Positive (FP) and False Negative (FN) indicate that the predictions made by the model are incorrect. TP is a positive stunting status correctly predicted, while TN is a negative stunting status also correctly predicted. FP is a negative stunting status incorrectly predicted as positive, and FN is a positive stunting status incorrectly predicted as negative. The values from this matrix can be used to calculate the model's performance, which has been created using accuracy, precision, recall, and F1-score matrices. These matrices are calculated using the following formulas:

1) Accuracy

$$(accuracy) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

2) Precision

$$(Precision) = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

3) Recall

$$(Recall) = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

2) F1-Score

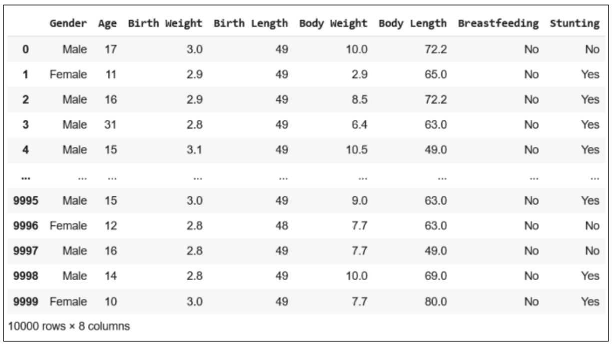
$$(F1 - score) = 2 \times \frac{recall\ \times\ precision}{recall\ -\ precision} \quad (4)$$

Subsequently, the evaluation results of the three models are compared to determine which algorithm can predict stunting better and more accurately than the others.

III. RESULT AND DISCUSSION

A. Data Collection

The dataset used in this study was obtained from the Kaggle platform, and the dataset titled "faktor stunting" was uploaded by Harnelia. The dataset used in this study can be seen in Figure 2.



	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Breastfeeding	Stunting
0	Male	17	3.0	49	10.0	72.2	No	No
1	Female	11	2.9	49	2.9	65.0	No	Yes
2	Male	16	2.9	49	8.5	72.2	No	Yes
3	Male	31	2.8	49	6.4	63.0	No	Yes
4	Male	15	3.1	49	10.5	49.0	No	Yes
...	...	...	...	...	...	...	...	...
9995	Male	15	3.0	49	9.0	63.0	No	Yes
9996	Female	12	2.8	48	7.7	63.0	No	No
9997	Male	16	2.8	49	7.7	49.0	No	No
9998	Male	14	2.8	49	10.0	69.0	No	Yes
9999	Female	10	3.0	49	7.7	80.0	No	Yes

10000 rows × 8 columns

Figure 2. Research Dataset.

Figure 2 shows that this dataset consists of 10,000 rows of data and 8 columns, which include 7 feature columns, namely Gender, Age, Birth Weight, Birth Length, Body Weight, Body Length, and Breastfeeding, as well as 1 target column to indicate stunting status (Stunting). The explanation regarding each column can be seen in Table 2.

TABLE 2
DESCRIPTION OF DATASET FEATURES/ATTRIBUTES

Features	**Description**
Gender	Child's gender (Male/Female)
Age	Child's age at examination (month)
Birth Weight	Child's weight at birth (kg)
Birth Length	Child's length at birth (cm)
Birth Weight	Child's weight at examination (kg)
Birth Length	Child's length at examination (cm)
Breastfeeding	Breastfeeding status (Yes/No)
Stunting	Stunting status (Yes/No)

B. Data Pre-processing

This stage transforms raw data into data ready for modelling to make the model more optimal. There are several steps in this stage, including:

1) Feature Selection

In this step, some features do not contribute significantly to modelling. Figure 3 below is the distribution of breastfeeding features that are considered to have no contribution and will be deleted.

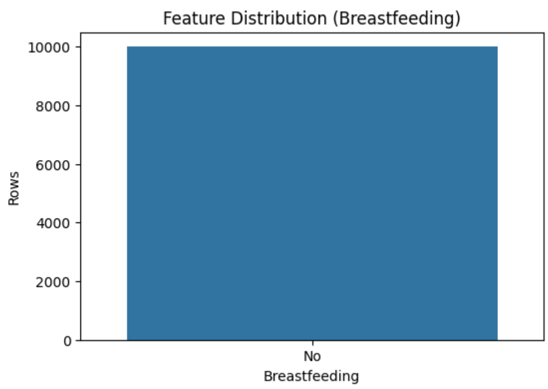


Figure 3. Distribution of Breastfeeding Feature.

As seen in Figure 3, the Breastfeeding feature only has one class, namely “No”, so it contributes less to modelling and will be removed at this stage.

2) Data Cleaning

Data cleaning begins with checking for empty values in the dataset. The checking results can be seen in Figure 4.

```
Missing values in each column:
Gender          0
Age             0
Birth Weight    0
Birth Length    0
Body Weight     0
Body Length     0
Breastfeeding   0
Stunting        0
dtype: int64
```

Figure 4. Results of Checking for Missing Values.

Figure 4 shows no empty values in any column in the dataset, so no handling is needed. The next check is for duplicate data.

```
Number of rows before removing duplicates: 7573
Number of duplicates: 807
Number of rows after removing duplicates: 6766
Number of duplicates: 0
```

Figure 5. Results of Checking for Duplicate Data.

From Figure 5, it can be seen that there are 2427 duplicate data points in the dataset used in this study. Furthermore, after the duplicate data is removed, the dataset has 7573 rows of data remaining.

3) Outlier Handling

Outlier data is data outside the range or very different from other data. This can be seen using a box plot as in Figure 6.

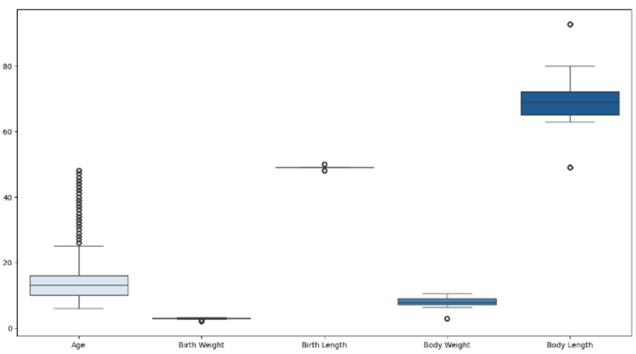


Figure 6. Boxplot of Outlier Data.

Figure 6 shows that all numeric features have outlier data outside the main area and whisker lines. Therefore, handling is needed to ensure that the data distribution becomes more representative and not biased towards the model. Handling is done by changing values that exceed the upper bound to the maximum reasonable value and values that are less than the lower bound to the minimum value. This method is called the capping method, the results of which can be seen in Figure 7.

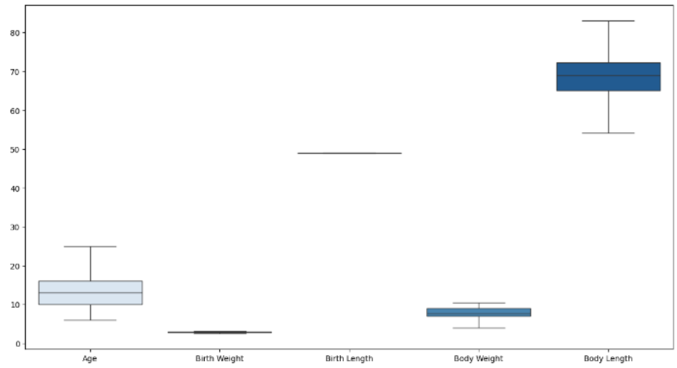


Figure 7. Boxplot After Outlier Handling.

4) Data Balancing

The check at this stage functions to ensure that the data distribution in the target is balanced. The following are the results of checking the distribution of target classes in Figure 8.

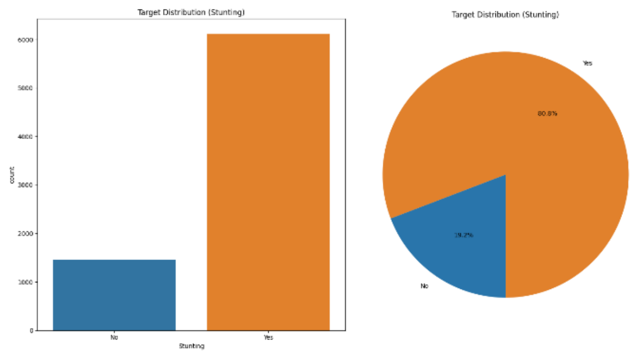
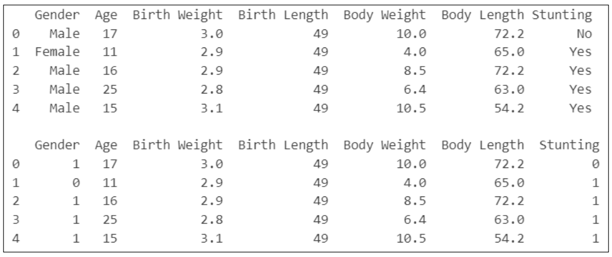


Figure 8. Target Distribution.

From Figure 8, it can be seen that the class distribution in the target is not balanced. In the “Yes” class, there are 6120 rows of data, while in the “No” class, there are 1453 rows of data. With such a large difference, handling this imbalance is needed and one way to overcome this problem is using oversampling techniques. In this study, the SMOTE approach was applied to balance the classes. However, before balancing, encoding must be performed or categorical data must be converted into numerical data so that categorical features can be processed using SMOTE. In the dataset used, there are categorical features, namely the Gender and Stunting columns. The approach used is Label Encoding because the features only have two classes, so it is enough to convert the categories into numbers 0 and 1. The following is a comparison of sample data before and after encoding.



	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
0	Male	17	3.0	49	10.0	72.2	No
1	Female	11	2.9	49	4.0	65.0	Yes
2	Male	16	2.9	49	8.5	72.2	Yes
3	Male	25	2.8	49	6.4	63.0	Yes
4	Male	15	3.1	49	10.5	54.2	Yes

	Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length	Stunting
0	1	17	3.0	49	10.0	72.2	0
1	0	11	2.9	49	4.0	65.0	1
2	1	16	2.9	49	8.5	72.2	1
3	1	25	2.8	49	6.4	63.0	1
4	1	15	3.1	49	10.5	54.2	1

Figure 9. Comparison of Sample Data after Encoding.

As seen in Figure 9, the classes in the Gender column change to 1 (Male) and 0 (Female), and in the Stunting column, they change to 1 (Yes) and 0 (No). After encoding is done, the dataset can be processed using SMOTE.

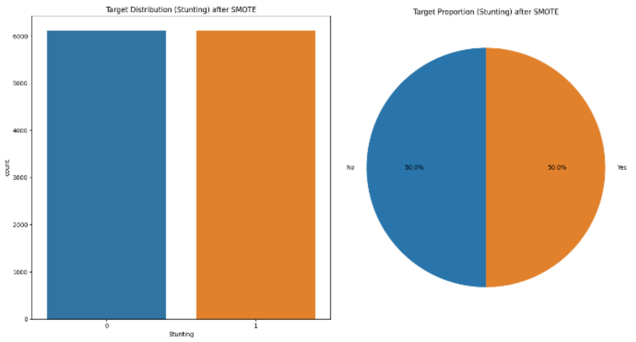


Figure 10. Target Distribution after SMOTE.

After using SMOTE on the dataset, as shown in Figure 10, the class distribution in the target (Stunting) becomes balanced with the same number of data points in each class, namely 6210 rows of data or with a proportion of 50%:50%, and the total rows in the dataset becomes.

After the data is balanced, the next step is to normalize the data. Data needs to be normalized because it can improve model performance by making the data within a uniform range or on the same scale. In this study, the approach used to normalize data is the Standard Scaler, which sets the mean to zero and the standard deviation to one.

TABLE 3
SAMPLE DATA COMPARISON

Gender	Age	Birth Weight	Birth Length	Body Weight	Body Length
Before normalization					
1	17	3.0	49	10.0	72.2
0	11	2.9	49	4.0	65.0
1	16	2.9	49	8.5	72.2
1	25	2.8	49	6.4	63.0
1	15	3.1	49	10.5	54.2
After normalization					
0.801484	0.6945529	1.0173024	0.	1.4568725	0.4639680
-1.24768	-0.531995	0.4511625	0.	-2.351251	-0.552049
0.801484	0.4901281	0.4511625	0.	0.5048416	0.4639680
0.801484	2.3299510	-0.114977	0.	-0.828001	-0.834277
0.801484	0.2857033	1.5834422	0.	1.7742161	-2.076076

Table 3 shows that before normalization, the values between features have quite a large scale, so there are features with an extensive value range that dominate other features. Furthermore, after normalization, the values between features have the same scale.

C. Modelling

Before modelling is performed, the data is divided into two parts, namely 80% training data and 20% testing data. This proportion is chosen because there is enough training data to train the model while leaving enough testing data for evaluation. The results of the data division are presented in Table 4.

TABLE 4
DATA DIVISION

Condition	Training Data (80%)	Testing Data (20%)	Total Data
Before SMOTE	6058	1515	7573
After SMOTE	9792	2448	12240

After the data division is done, the next step is modelling using the three machine learning algorithms, namely Random Forest (RF), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). There are two different data conditions in this modelling process: the condition of imbalanced data and the condition of data that has already been balanced with the SMOTE technique. Therefore, there are six different models whose test results will be compared in the evaluation stage.

D. Evaluation

After the data is trained into a machine learning model, the next stage is the evaluation stage. At this stage, the trained model will be tested for performance using testing data. After being tested, the test results will be evaluated using a confusion matrix. The following is the confusion matrix for each model.

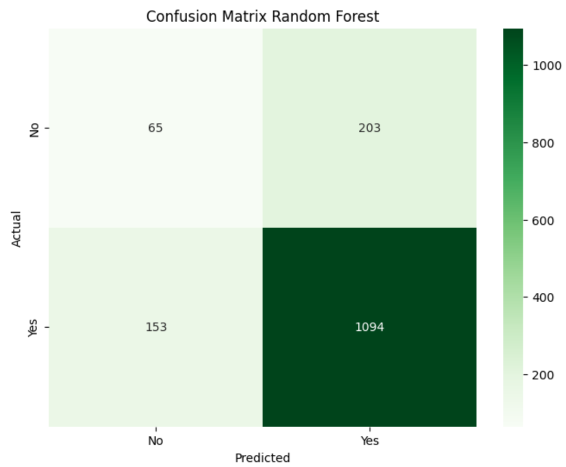


Figure 11. RF Confusion Matrix before SMOTE.

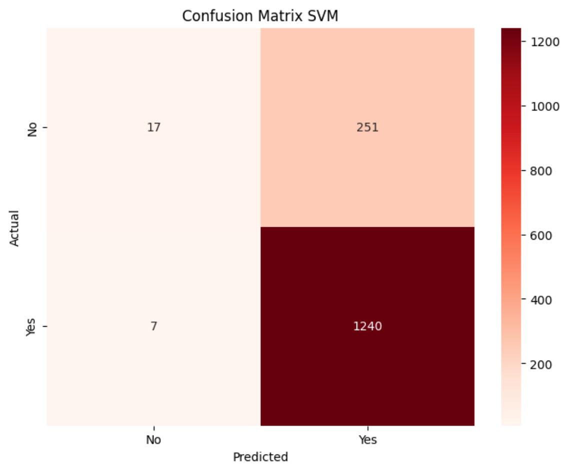


Figure 12. SVM Confusion Matrix before SMOTE.

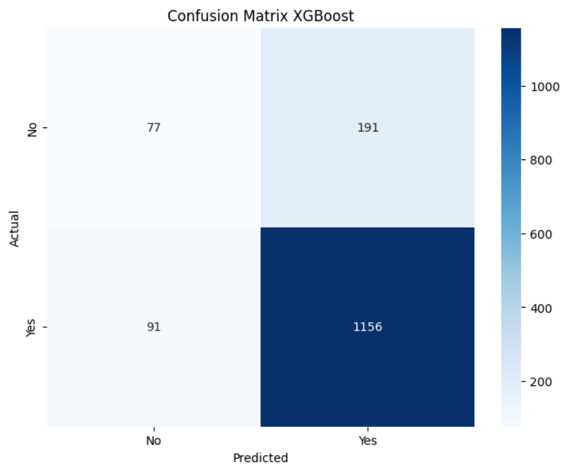


Figure 13. XGBoost Confusion Matrix before SMOTE.

From Figures 11, 12, and 13, it can be seen that the three models show different characteristics. Random Forest produces 65 True Negatives (TN), 203 False Positives (FP), 153 False Negatives (FN), and 1094 True Positives (TP). This model is good at recognizing positive stunting cases. SVM performs better with 17 TN, 251 FP, 7 FN, and 1240 TP. This model is very good at recognizing positive stunting cases with few missed cases, as evidenced by its call value (99.44%). However, the high FN value indicates that this model predicts non-stunting cases as stunting. Meanwhile, the XGBoost model with 77 TN, 191 FP, 91 FN, and 1156 TP makes this model more balanced between precision (85.82%) and recall (92.70%) values. All three models show significant prediction imbalances, reflecting the dataset's imbalance before the data balancing process. The following are the model's confusion matrices after the SMOTE balancing process.

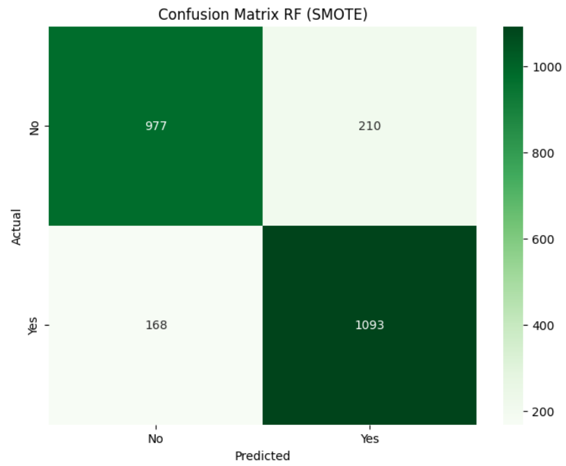


Figure 14. RF Confusion Matrix after SMOTE.

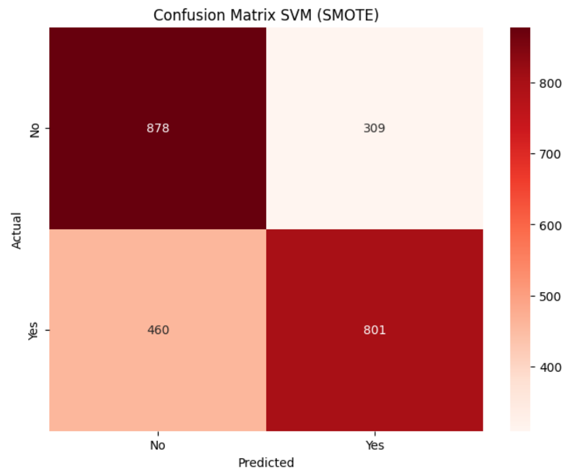


Figure 15. SVM Confusion Matrix after SMOTE.

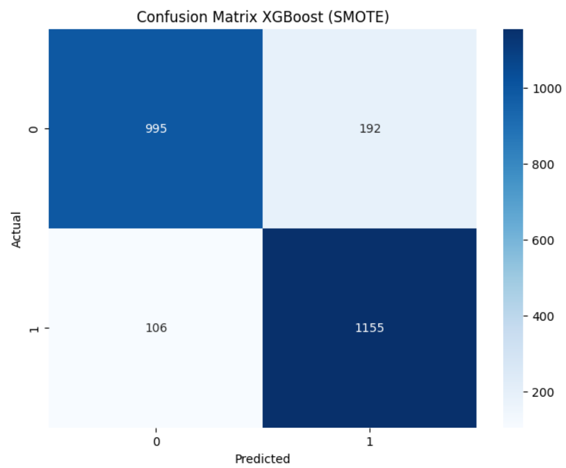


Figure 16. XGBoost Confusion Matrix after SMOTE.

After the data balancing process using SMOTE, model performance evaluation shows significant changes as reflected in the confusion matrices in Figures 14, 15, and 16. After SMOTE, the Random Forest model becomes more balanced in line with the increasing number of True Positives and True Negatives proportionally. Additionally, this model shows a decrease in False Negatives, which means fewer stunting cases are missed.

The SVM model tends to be aggressive in predicting stunting cases after data balancing. However, this increases False Positives, where quite a few non-stunting cases are incorrectly predicted as stunting. Meanwhile, the XGBoost model also shows improved performance after SMOTE. There is an increase in True Positives and True Negatives and a decrease in False Negatives. This model can recognize more stunting cases while accurately identifying non-stunting cases.

The following is a comparison of accuracy, precision, recall, and F1-score matrices to calculate the performance of the models, which can be seen in Table 5.

TABLE 5
PERFORMANCE EVALUATION

Model	Accuracy	Precision	Recall	F1-Score
RF	76.50 %	84.35 %	87.73 %	86.01 %
RF + SMOTE	84.56 %	83.88 %	86.68 %	85.26 %
SVM	82.97 %	83.17 %	99.44 %	90.58 %
SVM + SMOTE	68.59 %	72.16 %	63.53 %	67.57 %
XGB	81.39 %	85.82 %	92.70 %	89.13 %
XGB + SMOTE	87.83 %	85.75 %	91.59 %	88.57 %s

IV. CONCLUSION

Based on the research results, using machine learning algorithms with the SMOTE technique significantly improves model performance in predicting stunting. Of the three tested algorithms, XGBoost with SMOTE assistance recorded the best performance with 87.83% accuracy, 85.75% precision, 91.59% recall, and 88.57% F1-score, outperforming Random Forest (+3.27% accuracy) and SVM (+19.24% accuracy). SMOTE has been proven effective in overcoming data imbalance by increasing the detection of stunting cases (True Positive) while reducing classification errors (False Negative). However, SVM experienced a decrease in performance due to its tendency to overfit synthetic data. These results confirm that the combination of XGBoost and SMOTE is optimal for stunting prediction and feasible to implement in public health systems to detect stunting risk more quickly, accurately, and efficiently.

REFERENCES

[1] A. A. G. Y. Pramana, M. F. Maulana, M. C. Tirtayasa, and D. A. Tyas, "Enhancing Early Stunting Detection: A Novel Approach using Artificial Intelligence with an Integrated SMOTE Algorithm and Ensemble Learning Model," in *Proceedings - 2024 IEEE Conference on Artificial Intelligence, CAI 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 486–493. doi: 10.1109/CAI59869.2024.00098.

[2] N. Oktavin Idris and N. Umasugi, "Applying the Multi-Attribute Utility Theory (MAUT) to Accurately Determine Stunting Susceptibility Levels in Toddlers," 2024. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[3] S. Sumardiyono, "Pengaruh Usia, Tinggi Badan Dan Riwayat Pemberian Asi Eksklusif Terhadap Stunting Pada Balita," *Medika Respati : Jurnal Ilmiah Kesehatan*, vol. 15, no. 1, p. 1, Feb. 2020, doi: 10.35842/mr.v15i1.269.

[4] I. S. I. Putri, R. S. Pradini, and M. Anshori, "Decision Tree Regression Untuk Prediksi Prevalensi Stunting di Provinsi Nusa Tenggara Timur," *Jurnal Teknologi Informatika dan Komputer*, vol. 10, no. 2, pp. 413–427, Sep. 2024, doi: 10.37012/jtik.v10i2.2179.

[5] N. Faoziatun Khusna *et al.*, "Implementasi Random Forest dalam Klasifikasi Kasus Stunting pada Balita dengan Hyperparameter Tuning Grid Search," *Seminar Nasional Sains Data*, vol. 2024, 2024.

[6] "Pengembangan PAUD HI Untuk Mendukung Penurunan Stunting di Indonesia," https://www.kemenkopmk.go.id/index.php/pengembangan-paud-hi-untuk-mendukung-penurunan-stunting-di-indonesia.

[7] Peraturan Presiden Nomor 18 Tahun 2020 Tentang Rencana Pembangunan Jangka Menengah Nasional Tahun 2020-2024.

[8] Y. Wiratama and R. Abdul Aziz, "Perbandingan Prediksi Penyakit Stunting Balita Menggunakan Algoritma Support Vektor Machine dan Random Forest," *Technology and Science (BITS)*, vol. 6, no. 2, pp. 1159–1168, 2024, doi: 10.47065/bits.v6i2.5543.

[9] A. Mizwar, A. Rahim, P. Hartato, A. Ridwan, and F. Asharudin, "Machine Learning-Based Approach for HIV/AIDS Prediction: Feature Selection and Data Balancing Strategy," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[10] H. Hasanah, "Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung," 2023.

[11] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 257–265, Jan. 2024, doi: 10.57152/malcom.v4i1.1078.

[12] T. Sugihartono, B. Wijaya, Marini, A. F. Alkayes, and H. A. Anugrah, "Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 667–682, Jan. 2025, doi: 10.47738/jads.v6i1.494.

[13] N. Rizki Febriyanti and A. Dwi Hartanto, "Edumatic: Jurnal Pendidikan Informatika Analisis Perbandingan Algoritma SVM, Random Forest dan Logistic Regression untuk Prediksi Stunting Balita," *Edumatic: Jurnal Pendidikan Informatika*, vol. 9, no. 1, pp. 149–158, 2025, doi: 10.29408/edumatic.v9i1.29407.

[14] D. Nurmalasari, H. R. Yuliantoro, D. Hidayatul Qudsi, T. Informasi, P. Caltex Riau, and A. Perpajakan, "Improving Panic Disorder Classification Using SMOTE and Random Forest," *Journal of Applied Informatics and Computing (JAIC)*, vol. 8, no. 2, p. 272, 2024, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[15] M. I. Maulana, F. Nugraha, and A. Setiawan, "Model Prediksi Kualitas Air Untuk Budidaya Ikan Lele Dengan Algoritma Extreme Gradient Boosting," *Technology and Science (BITS)*, vol. 6, no. 3, 2024, doi: 10.47065/bits.v6i3.5998.