Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 24, No. 2, March 2025, pp. 235∼246
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v24i2.4092

Determining Toddler's Nutritional Status with Machine Learning Classification Analysis Approach

Taufik Hidayat[1], **Mohammad Ridwan**[1], **Muhamad Fajrul Iqbal**[1], **Sukisno**[1], **Robby Rizky**[2], **William Eric Manongga**[3]
[1]Universitas Islam Syekh-Yusuf, Tangerang, Indonesia
[2]Universitas Mathla'ul Anwar, Banten, Indonesia
[3]Chaoyang University of Technology, Taichung City, Taiwan

Article Info

Article history:

Received May 29, 2024
Revised February 16, 2025
Accepted March 06, 2025

Keywords:

Analysis model;
Classification;
Machine learning;
Nutritional status;
Toddlers.

ABSTRACT

The nutritional status of toddlers is a common issue many countries face worldwide. Various facts indicate that malnutrition is a primary focus for many researchers. Several efforts have been made to address this problem, including developing analytical models for identification, classification, and prediction. This study aims to evaluate the nutritional status of children by utilizing a classification analysis approach using Machine Learning. This research aims to improve the accuracy of the classification system and facilitate better decision-making in stunted toddlers, which is a priority, especially in the health sector. The Machine Learning classification analysis process will later utilize the performance of the Naive Bayes algorithm, the Support Vector Machine algorithm, and the Multilayer Perceptron algorithm. ML performance can be optimized using gridsearchCV to produce optimal classification analysis patterns. The data set of this study uses 6812 toddler data sourced from the Health Center at the Tangerang Regency Health Office. Based on the research presented, Machine Learning performance in analyzing nutritional status classification provides maximum results. The results are reported based on a precision level with an accuracy of 88%. The results of this analysis can also present a classification of nutritional status based on knowledge. This study can contribute to and update the analysis model in determining nutritional status. The results of this study can also provide benefits in handling nutritional status problems that occur in children.

Corresponding Author:

Taufik Hidayat, +6281398475099.,
Faculty of Engineering and Informatics Engineering,
Universitas Islam Syekh-Yusuf, Tangerang, Indonesia,
Email: thidayat@unis.ac.id.

How to Cite:
T. Hidayat, M. Ridwan, M. Iqbal, S. Sukisno, R. Rizky, and W. Manongga, "Determining Toddler's Nutritional Status with Machine Learning Classification Analysis Approach", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 24, No. 2, pp. 235-246, March, 2025.

1. INTRODUCTION

Stunting is a condition characterized by impaired linear growth during childhood, representing the most prevalent form of malnutrition worldwide. The physical and neurocognitive impairments associated with this growth disorder can be potentially permanent, posing a major challenge to human development [1]. Malnutrition affects muscle function, weakens the immune system, impairs brain function, and can lead to issues with neurological development [2]. Between 2013 and 2021, Indonesia saw an annual decrease in the stunting rate, averaging between 1.28% and 2.1% each year. In 2021, the Indonesian government set a goal, outlined in Government Regulation No. 72 of 2021 on Accelerating Stunting Reduction, to lower the stunting rate to 14% by 2024. This requires a total reduction of 10.4% or an average annual decrease of 3.4% [3].

This study [4] applies the Naïve Bayes Classifier algorithm to identify stunting nutritional status in toddlers based on demographic and anthropometric attributes, using a dataset of 1,000 records. The data is divided with a ratio of 6% for training and 35% for testing, achieving an accuracy rate of 94.65%. No data handling techniques were applied for imbalanced data, as the study included an unequal distribution of stunting nutritional status, including categories of very short, short, and normal. In another study [5], the sample consisted of 200 toddlers, 159 of whom were diagnosed with stunting, while 41 were not. The classification results demonstrated significant effectiveness in both methods used, achieving success rates of over 90%. Research by [6] showed that the Naïve Bayes method can be used to classify stunting status in toddlers. The applied Naïve Bayes algorithm achieved an average performance with 58% accuracy, 68% precision, and 58% recall from the confusion matrix test, using 30% testing data and 70% training data. A study conducted by [7] aimed at detecting children's nutritional status using a dataset from Kaggle consisting of 121,000 records with four main features: age, gender, height, and nutritional status. The experimental results showed that SVM with an RBF kernel and CNN achieved the highest accuracy of 98%, while Logistic Regression and MLP reached 76% and 97%, respectively. SVM with an RBF kernel was selected as the best model due to its high accuracy and computational efficiency. In research by [8], the dataset on nutritional status (Weight-for-Age) from the Health Department of Medan City consisted of 1,528 records. Classification performance with optimization of the Support Vector Machine algorithm based on Particle Swarm Optimization, using Radial Basis Function and Polynomial kernels, demonstrated improved performance. The best performance achieved was an increase in Accuracy to 78%, Precision to 89%, Recall to 66%, and F1-Score to 72% using the Radial Basis Function kernel. Finally, the study by [9] used a toddler dataset consisting of 8 attributes: gender, age, birth weight, birth length, weight, height, breastfeeding, and stunting, with a total of 10,000 records. Encoding was performed to convert categorical data into numerical attributes for gender, breastfeeding, and stunting. The study employed MLP integrated with GridSearchCV hyperparameter optimization, achieving an accuracy of 82.37%.

This study has 6,812 toddler records, which are the results of direct measurements of toddlers. We balance the data on toddler measurement data, which some researchers did not do. Handling data imbalance is an important step in determining whether a majority class or a minority class is handling missing data. In the proposed study, we will use Stratified k-fold and hyperparameter tuning with GridSearchCV [10].

The difference in this study lies in the handling of data imbalance, as the data obtained is imbalanced and undergoes balancing. The purpose of data balancing is to address situations where class differences are too wide so that balanced data will yield less disparity between classes, thus reducing uncertainty in the toddler nutritional status dataset, which typically contains a majority class. This study employs three algorithms: Support Vector Machine (SVM), NaNaïveve Bayes (NB), and Multi-Layer Perceptron (MLP). The SMOTE technique is proposed to address data imbalance to achieve class balance within the dataset. The results show that Multi-Layer Perceptron (MLP) achieves an accuracy of 65% and a precision of 88% through hyperparameter optimization with GridSearchCV. These findings can positively contribute to relevant stakeholders, particularly in the healthcare sector, predicting toddlers' stunting.

This paper is structured as follows: Section 1 presents an introduction that includes a problem background, a literature review, research gaps, and systematic information about the article. Section 2 explains the research methodology used, while section 3 contains research results and comments. Section 4 is the conclusion, which contains the findings of the investigation.

2. RESEARCH METHOD

This study includes several stages in classifying data sets using machine learning. The steps involved are initiating, developing a classification model, and generating classification results that can produce predictions processed in the machine learning model. The first stage, initiation, involves identifying the population prevalence [10] of stunting from the collected data, data initiation on stunting aims to measure the prevalence among children under five years of age. The second stage focuses on model development and selection of machine learning algorithms, such as Multilayer Perceptrons (MLP), which can detect implicit patterns in data [11], Naïve Bayes algorithm addresses data uncertainty in health-related datasets [12], Support Vector Machines (SVM) can predict stunting by analyzing nutritional factors [13]. Using dataset samples representing all dataset variables' minimum, mean, and maximum values ultimately classifies the dataset to extract information and insights from the available data.

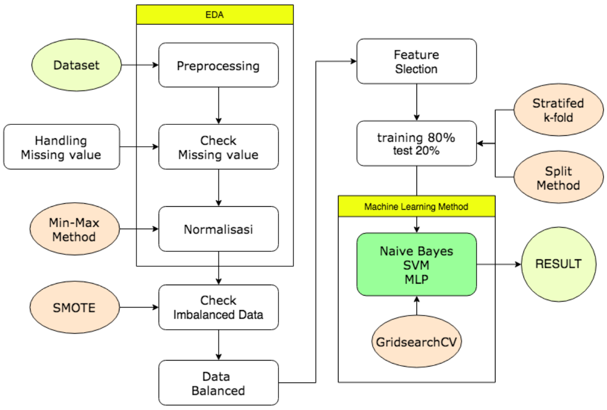

Figure 1. Proposed machine learning method design

We propose a machine learning design as in Figure 1, where there is a process that begins with checking the toddler dataset; starting from cleaning and missing values, we normalize the dataset, and then we check the balance of the dataset. If there is unbalanced data or one class is monotonically higher than the other class, we use the Synthetic Minority Oversampling Technique (SMOTE) to handle compositional data inequality. The SMOTE algorithm generates synthetic data points in a particular class by combining its features [14]. After the classes (data) are balanced, feature selection is performed using the split method to divide the dataset, which is then processed with machine learning. We initialize the dataset with 7 columns, 6 attributes, and 1 target. Table 1 provides the distribution of attribute data. In Table 1, BB/TB (birth weight/birth height) is the target for toddler nutritional status (stunting).

Table 1. Table Dataset Column/Variable Stunting

Variable/column	**Count**
Age at measurement	6813
Birth weight	6806
Birth height	6806
Weigh	6813
Height	6813
upper arm circumference (LiLA)	3585
Birth weight/ Birth height	6813

Each number's distribution is known thanks to the dataset in Table 1. Next, we examined the dataset to see if any missing values existed. The figure below illustrates how processing missing values resulted in over half of the upper arm circumference (LiLA) measurements being missing. In Table 2, we show the process of handling missed values, where previously there were 13 columns, and handling missing values or missing data was carried out. Then, a dataset of 6812 was obtained with 7 columns. In Table 2, the proposed variables will be processed to determine the nutritional classification consisting of age at measurement, birth weight, birth height, weight, height, and upper arm circumference. This data is taken based on measurements of the toddler's body condition, with the distribution of data sets with weights or values so that the normalization process can be carried out, which we propose in Table 3.

Table 2. Weighted Values Derived Following the Accounting for Missing Data

	Age at measurement	**Birth Weight (BB)**	**Birth Height (TB)**	**Weight**	**Height**	**Upper arm circumference (LiLA)**
0	36.0	2.6	52.0	15.0	95.0	15.731.726
1	37.0	2.8	48.0	14.5	95.0	15.731.726
2	23.0	2.9	50.0	11.0	83.0	15.731.726
3	34.0	2.7	47.0	12.5	92.0	15.731.726
4	50.0	2.7	48.0	15.5	99.0	15.731.726
...	...	...	...	...	...	...
6808	25.0	3.4	50.0	13.0	87.0	15.731.726
6809	20.0	2.8	48.0	11.3	83.0	15.731.726
6810	50.0	2.9	49.0	15.0	98.0	15.731.726
6811	46.0	3.0	49.0	17.0	101.0	15.731.726
6812	27.0	3.5	49.0	13.2	85.3	15.500.129

Table 3 displays the data in relation to the pre-normalization process. When displaying the data's descriptive statistics, note its distribution, as well as its minimum, maximum, average, and standard deviation, and if it only collects within specific quartile ranges. For us to perform further preprocessing [15], this is crucial.

Table 3. Minimum and Maximum Value Results

	Age at measurement	Birth Weight	Birth Height	Weight	Height	Upper arm circumference	Gender	BB/TB
count	6.813.000.000	6.813.000.000	6.813.000.000	6.813.000.000	6.813.000.000	6.813.000.000	6.813.000.000	6.813.000.000
mean	39.039.337	10.851.270	49.633.072	15.624.616	94.932.071	15.693.493	0.459269	0.548657
std	13.842.539	148.898.215	57.607.358	106.767.134	10.451.184	1.713.650	0.498375	1.522.734
min	0.000000	0.000000	0.000000	2.000.000	1.000.000	0.160000	0.000000	0.000000
25%	29.000.000	2.900.000	48.000.000	12.800.000	90.000.000	15.500.129	0.000000	0.000000
50%	41.000.000	3.000.000	49.000.000	14.600.000	98.000.000	15.731.726	0.000000	0.000000
75%	50.000.000	3.300.000	50.000.000	16.400.000	102.000.000	16.000.000	1.000.000	0.000000
max	60.000.000	3.300.000.000	4.749.000.000	8.824.000.000	115.000.000	98.000.000	1.000.000	6.000.000

We propose the Min-Max normalization [16] method to perform scaling on data so that the value is within a certain range, 0 to 1. This method is also known as Min-Max scaling or Min-Max normalization. Min-Max Normalization aims to change the values of various features in the toddler nutritional status dataset so that they all have a similar scale without changing the relative distribution between the feature values of a toddler nutritional status dataset that we are working on, but this must be ensured in sensitivity to outliers if not handled properly, the formula 12.

$$\text{Min-Max normalization} X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

$$\text{Min-Max scaling} X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

A variation called Min-Max scaling is used to normalize the training data for machine learning models or to scale numerical values uniformly. Because the gradient varies consistently rather than abruptly owing to disparate scales, the model learns and converges more quickly as a result. The dataset is displayed in Table 4 following Min-Max scaling normalization. As you can see, there are 6812 datasets and 6 columns in this data once it has been normalized.

Table 4. Dataset Normalization Results with Min-Max Scaling

	Age at measurement	Birth Weight (BB)	Birth Height (TB)	Weight	Height	Upper arm circumference (LiLA)
0	0.600000	0.000788	0.010950	0.001474	0.824561	0.159155
1	0.616667	0.000848	0.010107	0.001417	0.824561	0.159155
2	0.383333	0.000879	0.010529	0.001020	0.719296	0.159155
3	0.566667	0.000818	0.009697	0.001190	0.796246	0.159155
4	0.833333	0.000818	0.010107	0.001530	0.859649	0.159155
...	...	...	...	...	...	...
6808	0.416667	0.001030	0.010529	0.001247	0.754386	0.159155
6809	0.333333	0.000848	0.010107	0.001054	0.719296	0.159155
6810	0.833333	0.000879	0.010318	0.001474	0.850877	0.159155
6811	0.766667	0.000909	0.010318	0.001700	0.877193	0.159155
6812	0.450000	0.001061	0.010318	0.001270	0.156788	0.156788

3. RESULT AND ANALYSIS

We propose this research and take several steps starting from collecting toddler nutritional status dataset, conducting exploratory data analysis with the aim of preprocessing, overcoming missing values to know the completeness of variables from a data, and handling incomplete data or missing values, then normalizing. We propose checking the data to see if there is unbalanced data. SMOTE generates additional instances using linear interpolation between existing sample data from the minority class and its nearest neighbors in a collection of toddler nutritional status datasets. This technique aims to reduce doubts in the nutritional status dataset of toddlers, which usually contains a majority class. Figure 2 shows class imbalance in the toddler nutritional status dataset. We propose SMOTE to be able to overcome the class imbalance in the toddler nutritional status dataset by generating synthetic samples for the minority class, thereby balancing the class distribution.

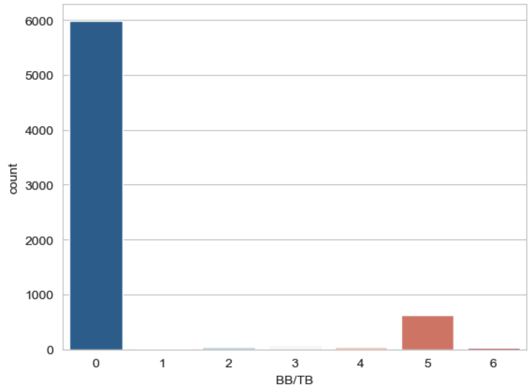

Figure 2. Imbalanced dataset

We propose an over-sampling and under-sampling technique with SMOTE to balance the data, as seen in Figure 3. Previously, the toddler nutrition dataset showed an imbalance in Figure 2. The synthetic minority over-sampling technique duplicates instances of the minority class to balance the class distribution so that the toddler nutrition dataset shows balance, which can then be processed in the next step.

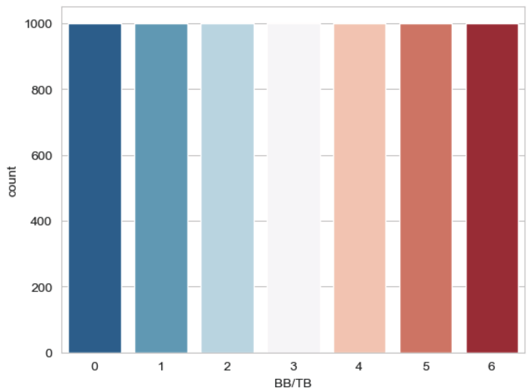

Figure 3. Balanced dataset

Table 5 shows a dataset with class imbalance; class imbalance affects the performance of classification models [17], indicating substantial inequality between classes. This imbalance affects the performance of the model [18], leading to inaccuracies in evaluation and biased learning in machine learning towards the majority class, which results in better classification performance [19], especially regarding recall and F1 score, and underlines the importance of accurately identifying the minority class, which is often critical in many scenarios.

Table 5. The Number of Samples for Each Class Imbalance

Sampler class	Count
0	5992
1	1000
2	1000
3	1000
4	1000
5	1000
6	1000

Table 6 shows the adjusted information; this proposes an arbitrary beneath sampler to adjust the number of tests in an imbalanced dataset. With a programmed inspecting technique, which means it'll adjust the minority lesson with the same number as the larger part course with settled arbitrary esteem of 42, this settled arbitrary esteem of 42 is utilized to guarantee that the comes about gotten from the irregular prepare (such as information part or test determination) will be steady each time the code is run, with the esteem 42 utilized as the esteem to initialize the irregular number generator. Utilized to guarantee steady comes about. The undersampled information is 'x_resampled' and 'y_resampled'. After the total undersampling handle, the code prints the number of tests for each undersampled course utilizing 'Counter.'

```
from imblearn.under_sampling import RandomUnderSampler
from collections import Counter

RandomUnderSampler = RandomUnderSampler(sampling_strategy='auto', random_state=42)
X_resampled, y_resampled = RandomUnderSampler.fit_resample(X_resampled
, y_resampled)
print("Jumlah sampel per kelas setelah SMOTE:", Counter(y_resampled))
```

Figure 4. Python source code of random under sampler module from imblearn.under_sampling and Counter

Table 6. The Number of Samplers for Each Class is Balanced

Sampler class	Count
0	1000
1	1000
2	1000
3	1000
4	1000
5	1000
6	1000

We propose a box plot visualization that represents the data distribution by visually displaying the spread and central tendency of the data to compare the distributions in terms of extreme values, medians, and quartiles, thus getting an idea of the presented attributes/variables that we generally suspect could be the main differentiators. between 6 (six) classes. Figure 5 visualizes body weight data based on the original class or before normalization, where the data does not yet show a distribution of weight data values from each class, so this improves the visualization of reading the data.

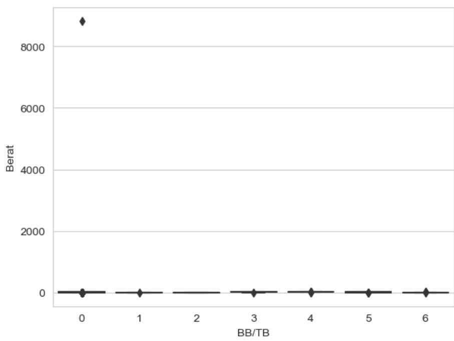

Figure 5. Boxplot of original BB data

Meanwhile, Figure 6 illustrates how body weight data across different classes have been normalized for easy visualization. The figure shows classes where outlier data, representing values significantly different from most toddler nutrition datasets, are absent. These toddler nutrition datasets typically reflect average weights from normally distributed data.

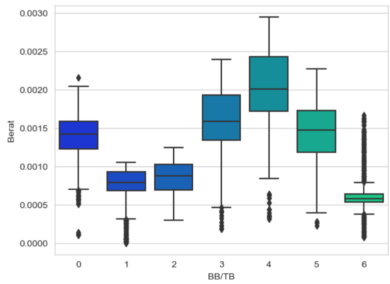

Figure 6. Boxplot of normalized BB data

Meanwhile, Figure 7 shows normalized toddler nutritional data, with a distribution showing that one class has no outliers and several classes have outliers; however, on average, the data is normally distributed.

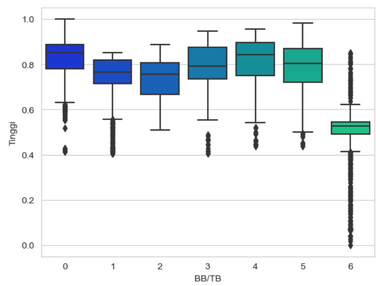

Figure 7. Boxplot of normalized TB data

Figure 8 shows the distribution of nutritional data for toddlers based on the relationship between variables. For example, the variable relationship between upper arm circumference (LiLA) and age, when measured, shows that almost all toddlers have a measurement of 0 – 25 cm from the age of 0 – 60 months. Thus, Figure 8 will provide an information visualization of each relationship between variables.

Figure 8. Shows the relationship between variables

Figure 9 displays the distribution of nutritional data for toddlers categorized by gender, encompassing both male and female groups. The visualization provides insight into how height correlates with age at measurement. Notably, it highlights a prevalence of higher values among males compared to females.

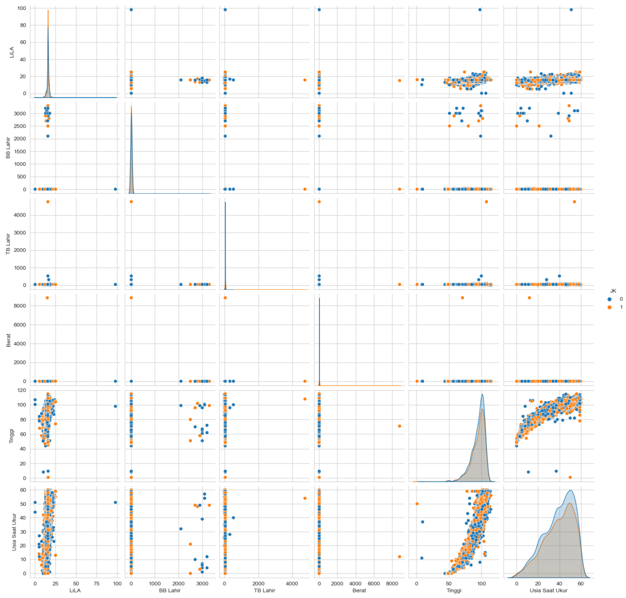

Figure 9. Variable relationships based on gender

The classes in the sklearn.feature_selection module can be used to improve the estimator's performance on very high-dimensional datasets or to raise its accuracy score for feature selection and dimensionality reduction on sample sets. Figures 10 and 11 illustrate the variations in feature selection; Figure 10 details feature selection before normalization and class balance, while Figure 11 indicates that the data has undergone these processes.

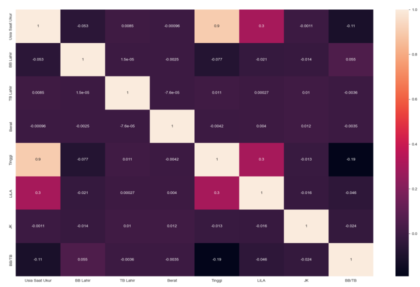

Figure 10. Original data feature selection

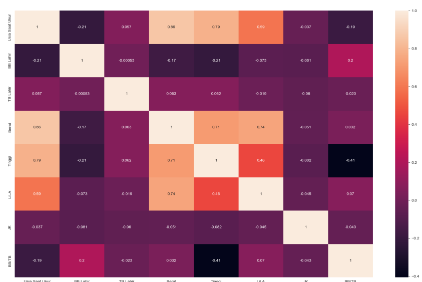

Figure 11. Selection of normalized and balanced data features

Splitting data into training data (training set) and test data (testing set) with train_test_split where; X_train, X_test, y_train, y_test = train_test_split(df_feat, df_target, test_size=0.30, random_state=0), with information namely X_train: training data attributes, X_test: test data attributes, y_train: training data label/class, y_test: test data label/class.

Finally, two good working processes were obtained on Naïve Bayes, Support Vector Machine, and Multilayer Perceptron, namely the first with Statifed k-fold then with GridsearchCV, for more details we show the results below;

3.1. Startifed k-fold

In Table 7, we present the outcomes of our testing using stratified k-fold cross-validation [20], where stratified K-fold ensures that each fold maintains the same proportion of class labels as the entire dataset, which is especially important for imbalanced datasets such as those often found in medical diagnostics. Another research [5], the sample used was 200 toddlers, of which 159 toddlers were diagnosed with stunting, and 41 toddlers were not diagnosed; the accuracy results of both methods showed a success rate of more than 90%; this study did not use data balance treatment. While we obtained an SVM value reaching an accuracy of 48%, Naïve Bayes reached 37%, and MLP reached 65%; from the values we obtained, we have used stratified K-fold in data handling to ensure that model evaluation is not biased toward the majority class.

Table 7. Test Results with Stratified k-fold

Stratifed k-fold	**SVM %**	**NB %**	**MLP %**
Accuracy	48%	37%	61%
Precision	75%	76%	83%
Recall	95%	27%	94%
f1-score	84%	40%	88%

3.2. GridsearchCV

We have improved or selected the optimal algorithm by gridsearchCV to find out which of these three algorithms has a value that is optimal enough to be selected. Table 8 shows the increase in test results obtained by the multilayer perceptron (MLP) with an accuracy of 0.65 from the previous 0.61 and a precision of 0.88 from the previous 0.83. Compared to other research [9] using stratified k-fold and hyperparameters achieving an accuracy value of 0.82 and a precision of 0.55, we think the results we obtained are quite good.

Table 8. Optimal Algorithm Test Results

GridsearchCV	**SVM %**	**NB %**	**MLP %**
Accuracy	61%	37%	65%
Precision	82%	76%	88%
Recall	99%	73%	96%
f1-score	90%	61%	92%

4. CONCLUSION

This study recommends a data separation technique to distinguish training and testing datasets, followed by feature selection to eliminate irrelevant attributes, enhancing the precision and accuracy of learning and result interpretation. The training process employs a Support Vector Machine (SVM), Naïve Bayes (NB), and Multilayer Perceptron (MLP) with Stratified K-Fold cross-validation. Hyperparameter tuning using GridSearchCV results in SVM achieving a precision of 0.88% and an accuracy of 0.61%, NB reaching 0.37%, and MLP attaining 0.65%. This research provides insights into methods such as data balancing, Stratified K-Fold cross-validation, and hyperparameter optimization (GridSearchCV), comparing the findings with previous studies. Future research directions include exploring additional features, utilizing ensemble algorithms, and implementing cross-validation techniques to improve model generalization to new data. This study employs three algorithms: Support Vector Machine (SVM), Naïve Bayes (NB), and Multilayer Perceptron (MLP).

5. ACKNOWLEDGEMENTS

We are grateful to our seniors and colleagues, as well as the university and Tangerang District Health Service, for providing the facilities needed for us to conduct this research. It is our desire that the information shared will lead to acts of worship.

6. DECLARATIONS

AUTHOR CONTIBUTION

Taufik Hidayat: Conceptualization, Methodology, Validation. **Mohammad Ridwan**: Formal analysis resources, Writing the original draft. **Muhamad Fajrul Iqbal**: Visualization and Software. **Sukisno**: Wrtiting review and editing. Robby Rizky: Formal analysis, Investigation, and Data curation. **William Eric Manongga**: Formal Analysis.

FUNDING STATEMENT

Universities provided financial assistance for this study, and the author also covered the costs of writing it, as well as his own costs for study design, data collection, analysis, and interpretation.

COMPETING INTEREST

For this article, the writers have declared no conflicts of interest.

REFERENCES

[1] M. Hadi, "the Relationship Between Socio - Economic Factors and Risk of Stunting : a Systematic Review," *Journal of Advance Research in Medical & Health Science (ISSN: 2208-2425)*, vol. 9, no. 7, pp. 25–30, 2023, https://doi.org/10.53555/nnmhs.v9i7.1751.

[2] H. Shi, D. Yang, K. Tang, C. Hu, L. Li, L. Zhang, T. Gong, and Y. Cui, "Explainable machine learning model for predicting the occurrence of postoperative malnutrition in children with congenital heart disease," *Clinical Nutrition*, vol. 41, no. 1, pp. 202–210, 2022, https://doi.org/10.1016/j.clnu.2021.11.006.

[3] I. Soesanti, "Komitmen Pimpinan pada Penurunan Stunting di Kabupaten Tuban Jawa Timur," *Media Gizi Indonesia (National Nutrition Journal) Special Issue: International Conference on Stunting 2022*, vol. 17, no. 1, pp. 211–214, 2024, https://doi.org/10.20473/mgi.v17i1SP.211-214.

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

[5] S. Melyani, S. Z. Harahap, and I. Irmayanti, "Prediction of Stunting in Toddlers Combining the Naive Bayes Method and the C4.5 Algorithm," *Sinkron*, vol. 8, no. 2, pp. 1160–1168, 2024, https://doi.org/10.33395/sinkron.v8i2.13651.

[6] A. Rozaq and A. J. Purnomo, "Classification of Stunting Status in Toddlers using Naive Bayes Method in The City of Madiun Based on Website," *Jurnal Techno Nusa Mandiri*, vol. 19, no. 2, pp. 69–76, 2022, https://doi.org/10.33480/techno.v19i2.3337.

[7] F. T. Sabilillah, C. A. Sari, R. B. Abiyyi, and Y. P. Astuti, "Comparison of Machine Learning Algorithms on Stunting Detection for ' Centing ' Mobile Application to Prevent Stunting," vol. 8, no. 4, pp. 2361–2368, 2024, https://doi.org/10.33395/sinkron.v8i4.13967.

[8] S. Y. Andriyani, M. S. Lydia, and S. Efendi, "Optimization of Support Vector Machine Algorithm Using Stunting Data Classification," *Prisma Sains : Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, vol. 11, no. 1, p. 164, 2023, https://doi.org/=10.33394/j-ps.v11i1.6619.

[9] I. A. Cahyani, P. I. Ashuri, C. Sri, and K. Aditya, "Stunting Disease Classification Using Multi- Layer Perceptron Algorithm with GridSearchCV," *SinkrOn*, vol. 8, no. 1, pp. 392–401, 2024, https://doi.org/10.33395/sinkron.v9i1.13245.

[10] S. P. Manongga, H. Hendry, and D. H. F. Manongga, "Predicting a Stunting Prevalence Using Semi-Supervised Learning Models in East Nusa Tenggara," *ICIC Express Letters*, vol. 19, no. 4, pp. 1073–1086, 2023, https://doi.org/10.24507/ijicic.19.04.1073.

[11] M. G. Rojas, A. C. Olivera, and P. J. Vidal, "Optimising Multilayer Perceptron weights and biases through a Cellular Genetic Algorithm for medical data classification," *Array*, vol. 14, no. January, pp. 1–15, 2022, https://doi.org/10.1016/j.array.2022.100173.

[12] K. Gohari, A. Kazemnejad, M. Mohammadi, F. Eskandari, S. Saberi, M. Esmaieli, and A. Sheidaei, "A Bayesian latent class extension of naive Bayesian classifier and its application to the classification of gastric cancer patients," *BMC Medical Research Methodology*, vol. 23, no. 1, pp. 1–15, 2023, https://doi.org/10.1186/s12874-023-02013-4.

[13] S. M. Rahman, N. A. Ahmed, M. M. Abedin, B. Ahammed, M. Ali, M. J. Rahman, and M. Maniruzzaman, "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," *PLoS ONE*, vol. 16, no. 6 June 2021, pp. 1–11, 2021, https://doi.org/10.1371/journal.pone.0253172.

[14] T. Nguyen, K. Mengersen, D. Sous, and B. Liquet, "SMOTE-CD: SMOTE for compositional data," *PLoS ONE*, vol. 18, no. 6 June, pp. 1–19, 2023, https://doi.org/10.1371/journal.pone.0287705.

[15] G. Wang, B. Zhao, B. Wu, C. Zhang, and W. Liu, "Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases," *International Journal of Mining Science and Technology*, vol. 33, no. 1, pp. 47–59, 2023, https://doi.org/10.1016/j.ijmst.2022.07.002.

[16] P. P. Allorerung, A. Erna, and M. Bagussahrir, "Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, https://doi.org/10.14421/jiska.2024.9.3.178-191.

[17] L. Yuningsih, G. A. Pradipta, D. Hermawan, P. D. W. Ayu, D. P. Hostiadi, and R. R. Huizen, "IRS-BAG-Integrated Radius-SMOTE Algorithm with Bagging Ensemble Learning Model for Imbalanced Data Set Classification," *Emerging Science Journal*, vol. 7, no. 5, pp. 1501–1516, 2023, https://doi.org/10.28991/ESJ-2023-07-05-04.

[18] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 13, no. 6, pp. 2–34, 2023, https://doi.org/10.3390/app13064006.

[19] L. D. Cahya, A. Luthfiarta, J. I. T. Krisna, S. Winarno, and A. Nugraha, "Improving Multi-label Classification Performance on Imbalanced Datasets Through SMOTE Technique and Data Augmentation Using IndoBERT Model," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 3, pp. 290–298, 2024, https://doi.org/10.25077/TEKNOSI.v9i3.2023.290-298.

[20] S. Rana and A. K. Gautam, "K-Fold Cross-Validation through Identification of the Opinion Classification Algorithm for the Satisfaction of University Students," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 9, pp. 122–130, 2023, https://doi.org/10.3991/ijoe.v19i11.39887.

[This page intentionally left blank.]