Vol. 12 No.1 Tahun 2026, pp 10-20
Jurnal Teknologi dan Manajemen Informatika
http:// http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604 E-ISSN: 2580-8044

Comparative Analysis of SMOTE, WMOTE, and ADASYN Oversampling Methods on Multinomial Naive Bayes Performance in Classifying Toddlers Nutritional Status

Dewi Sartika [1], Naretha Kawadha Pasemah Gumay [2], Rendra Gustriansyah [3], Yesinta Florensia [4], Miftahul Falah [5]

[1,2,4,5] Universitas Sriwijaya, Indonesia
[3] Universitas Indo Global Mandiri, Indonesia

Article Info

Article History
Received : 24-11-2025
Revised : 12-12-2025
Accepted : 15-04-2026

Keywords
ADASYN;
MNB;
SMOTE;
WMOTE

✉ Corresponding Author
Naretha Kawadha Pasemah Gumay,
Universitas Sriwijaya,
narethakawadha@unsri.ac.id

ABSTRACT

Class imbalance in toddler nutritional status data often reduces the ability of classification models, especially in predicting minority classes. This study aims to analyze the impact of three oversampling techniques, namely SMOTE, WMOTE, and ADASYN, on improving the performance of the Multinomial Naive Bayes (MNB) algorithm. A dataset of 243 data was processed through a preprocessing stage by converting categorical variables using numeric labels. To meet the MNB algorithm's requirement for non-negative data, continuous numeric features (such as birth weight, birth height, weight, height, and age) were normalized using the Min-Max Scaler to the range [0, 1]. This process discretizes continuous values onto a probability scale to ensure feature compatibility with the Multinomial distribution. Data balancing was performed only on the training dataset, where the SMOTE method produced 374 data, ADASYN produced 375 data, and WMOTE produced 373 data. The evaluation results show that although all three oversampling methods experienced a slight decrease in global accuracy, the model's ability to detect minority classes improved, as evidenced by increases in G-Means and Balanced Accuracy. The test results concluded that MNB-ADASYN was the best model for prioritizing high sensitivity to all class labels, while MNB-WMOTE provided the most consistent global accuracy stability while maintaining performance on minority classes.

INTRODUCTION

Nutritional status among toddlers is an essential indicator that reflects the overall quality of public health and serves as a foundation for future human resource development. One of the most critical nutritional issues in Indonesia is stunting, a condition of impaired growth resulting from prolonged nutritional deficiencies [1]. The impacts of stunting are not only visible in physical growth but also affect cognitive development, intelligence, and productivity as children grow into adulthood [2]. Although the stunting rate in Indonesia has shown a declining trend over the past decade, it still has not reached the national target set in the 2020–2024 RPJMN [3]. Therefore, the use of technology to support early detection of nutritional status in toddlers has become increasingly relevant.

In the fields of data mining and machine learning, nutritional status classification serves as an approach that can assist health workers in identifying the condition of toddlers quickly and accurately. The Multinomial Naive Bayes method is one of the widely used algorithms because it offers simple computational processes, efficiency, and stable performance even when the amount of data is limited [4], [5]. However, one of the main challenges in child health data is class imbalance. The number of children categorized as "malnourished" or "obese" is often far smaller than those classified as having "normal nutrition," causing the model to be biased toward the majority class [6]. This situation creates a high risk of misclassification in the form of false negatives, where toddlers who actually require medical intervention are instead detected as normal. Clinically, this failure of early detection is very dangerous because critically ill toddlers miss the opportunity to receive immediate treatment, which ultimately can increase morbidity and even mortality. Therefore, this study not only focuses on global accuracy but also provides a thorough justification for improving sensitivity metrics to minimize the risk of such clinical failures.

To address the issue of class imbalance, several synthetic data–based oversampling techniques have been developed. One of the most widely used techniques is the Synthetic Minority Oversampling Technique (SMOTE), which has been shown to effectively improve the overall performance of the Multinomial Naive Bayes model [7] by generating new samples through feature interpolation between minority data points and their nearest neighbors [8]. However, SMOTE still has limitations, such as maintaining the data distribution without considering the varying levels of difficulty in classifying each sample, and it does not account for the complexity of individual instances. As a result, the synthetic samples generated may be suboptimal or insufficiently representative. In addition, the SMOTE method concentrates on generating data in borderline areas—regions close to both majority and minority classes—which are often sources of misclassification [9]

To address these weaknesses, the Weighted Minority Oversampling Technique (WMOTE) was developed, assigning higher weights to minority instances deemed important, especially those that are difficult to classify, so that the resulting synthetic samples are more representative [10]. This method works by identifying difficult-to-learn minority samples near the decision boundary, then assigning them greater weight based on their proximity to the majority class [11]. Through this selective weighting mechanism, MWMOTE produces more informative and accurate synthetic data, thereby significantly improving the model's generalization ability in detecting critical minority classes without exacerbating the overfitting that often occurs in conventional oversampling methods [12].

Furthermore, the Adaptive Synthetic Sampling (ADASYN) method is an adaptive oversampling technique. ADASYN automatically generates more synthetic samples in regions that are difficult to classify, based on the distribution of classification error probabilities [13]. This approach encourages better learning of minority classes, particularly those located near the decision boundaries. Considering the various characteristics of each oversampling technique, this study focuses on analyzing how the application of SMOTE, WMOTE, and ADASYN can improve the quality of imbalanced data and enhance the performance of the Multinomial Naive Bayes algorithm in classifying the nutritional status of toddlers. Each oversampling method employs a different approach in generating synthetic samples, which may lead to varying effects on evaluation metrics such as accuracy, precision, recall, f-measure, and g-means.

Therefore, this study is expected to provide a comprehensive overview of which oversampling technique is the most effective in producing a more accurate and fair classification model, particularly one that can better recognize minority classes in early identification of nutritional problems among toddlers.

METHODS

The system workflow shown in Figure 1 is applied in this study, starting with data loading, preprocessing, data validation, feature scaling, handling imbalanced datasets (training dataset balancing using the oversampling method), training, and model performance evaluation.

Figure 1. Research Model

The first step in this study is entering the dataset, which consists of nutritional status data for toddlers. The dataset includes several variables, namely gender, birth weight, birth height, age, body weight, body height, and weight gain status obtained during the nutritional assessment process. Next, preprocessing is performed, which involves handling missing data, noise (data cleaning), transforming categorical data into numeric data (label encoding), normalizing/standardizing if necessary, and identifying class distributions to detect imbalance [14]. This stage is carried out to ensure all variables are numeric and positive, that there are no blanks, and that the data align with the characteristics of the Multinomial Naïve Bayes Method.

Then, the data validation stage is carried out. There are two types of data validation methods: split validation and cross-validation [15], [16]. The data validation method used in this study is split validation with a test size of 0.3. This division uses a stratified sampling technique (via the stratify=y parameter) to ensure the class distribution in the test data remains representative of the original dataset, given the limited number of samples. Split validation is a model validation method that randomly divides the training and test data according to a specified proportion [17]. This validation method was chosen because the number of datasets used is limited. Then, data transformation is performed using the Min-Max Scaler to normalize all variables into the range of 0 to 1. This aims to ensure that all features are assessed as equally important; features with large scales do not dominate the model's probability calculation. The scaling process is carried out by fitting only to the training data, then applying (transforming) it to the test data to avoid information leakage (data leakage). Next, handling of imbalanced datasets is carried out; training datasets with unbalanced class distributions are processed using three methods: SMOTE, WMOTE, and ADASYN. Given the limited dataset, the nearest neighbor parameter k is dynamically adjusted to prevent system errors when the number of minority samples is very small, using Equation (1).

$$k = max(1, n\_minoritas\ - 1) \qquad (1)$$

SMOTE is a resampling technique in which the number of minority samples is increased to balance the majority class. This technique is suitable for small datasets because it does not reduce essential information. SMOTE works by identifying k nearest neighbors for each minority sample and generating synthetic data based on the required percentage through random interpolation of feature vectors between the minority sample and its k nearest neighbors [8]. The value of k is determined as a minimum of $n-1$, where n is the number of minority samples. The SMOTE calculation is expressed in Equation (2).

$$x_{new} = x_i + \lambda \cdot (x_j - x_i) \qquad (2)$$

Where x_i is the original minority data point, x_j is the nearest neighbor of x_i, and λ is a random value between 0 and 1 that determines how far the synthetic point x_{new} is positioned from x_i toward x_j [18].

In this study, the WMOTE approach is adapted from the *Majority Weighted Minority Oversampling Technique (MWMOTE)* proposed by Barua et al. (2014) [10]. The implementation used in this research does not fully replicate the original MWMOTE algorithm, but adopts its fundamental concept of assigning weights to minority instances based on their level of classification difficulty. Therefore, WMOTE in this study should be understood as a simplified adaptation of MWMOTE, where the weighting mechanism is utilized to prioritize minority samples that are more difficult to classify. This adaptation is tailored to the characteristics of the dataset and the experimental design, and does not represent a distinct or newly established oversampling method.

The adapted WMOTE approach assigns higher weights to minority instances that are considered important and difficult to classify in order to generate more representative synthetic samples [10]. In general, the WMOTE formula consists of three main stages. The first stage calculates the weight of each minority instance based on its level of classification difficulty. If x_i is a minority sample and m_i is the number of its majority-class neighbors. One formulation of the weight is expressed as (3):

$$wi=\left(\frac{mi}{\sum_{j-1}^{n} m_i}\right) \tag{3}$$

where a higher value of m_i indicates greater importance of the sample. The second stage selects minority samples based on their weights (4):

$$p(x_i)=\left(\frac{wi}{\sum_{j-1}^{n} wj}\right) \tag{4}$$

Next, in Stage 3, synthetic sample generation is performed for each selected sample x_i and its nearest neighbor x_{nn} (5):

$$x_{new} = x_i + \delta\ X\ (x_{nn} - x_i) \tag{5}$$

with:

$$\delta \sim U(0,1)$$

Meanwhile, ADASYN generates synthetic samples adaptively based on the level of classification difficulty and produces more data in borderline areas [13]. The ADASYN calculation is expressed in Equation (6):

$$x_{new} = x_i + \delta\ .\ (x_{zi} - x_i) \tag{6}$$

The next stage is model training, in which the Multinomial Naive Bayes algorithm is used as the classification model because it performs well on categorical data and is simple to implement [19]. Naive Bayes is a supervised learning method used to predict the probability of each class label based on the training data [20]. The probability calculation in Naive Bayes uses the Gaussian distribution for continuous data, while the Multinomial distribution is used for discrete data. Previous studies have shown that the Multinomial Naive Bayes method is capable of producing classification models with relatively high average performance [21]. The Multinomial distribution is used to determine probabilities for outcomes categorized into more than two groups. In general, if an experiment produces one of *k* possible outcomes $E_1, E_2, \ldots, E_k$ with corresponding probabilities $p_1, p_2, \ldots, p_k$, then the Multinomial distribution provides the probability that E_1 occurs x_1 times, E_2 occurs x_2 times, …, and E_k occurs x_k times in *n* independent trials, where:

$$x_1 + x_2 + \cdots + x_k = n \tag{7}$$

$$p_1 + p_2 + \cdots + p_k = 1 \quad (8)$$

With the probability distribution (9),

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k; n) \quad (9)$$

Each sequence that results in x_1 occurrences of E_1, x_2 occurrences of E_2, …, and x_k occurrences of E_k occurs with a probability of ${p_1}^{x_1}{p_2}^{x_2} \ldots {p_k}^{x_k}$. The total number of sequences that produce the same outcome distribution in *n* trials is equal to the number of ways to partition *n* items into *k* groups, with x_1 items in the first group, x_2 in the second group, …, and x_k in the *k* group. This can be done in (10):

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! x_2! \ldots x_k!} \quad (10)$$

If each trial yields *k* possible outcomes $E_1, E_1, \ldots E_k$ with corresponding probabilities $p_1, p_2, \ldots, p_k$, then the probability distribution of the random variables $x_1, x_2, \ldots, x_k$, which represent the number of occurrences of $E_1, E_1, \ldots E_k$ in *n* trials, is given by (11):

$$f(x_1, x_2, \ldots, x_k; p_1, p_2, \ldots, p_k; n) = \binom{n}{x_1, x_2, \ldots, x_k} {p_1}^{x_1}{p_2}^{x_2} \ldots {p_k}^{x_k} \quad (11)$$

with $\sum_{i=1}^{k} x_i = n$ dan $\sum_{i=1}^{k} p_i = 1$

The final stage is prediction and evaluation, which uses the Confusion Matrix to measure performance based on accuracy, precision, recall, f-measure, specificity, g-means, and balanced accuracy (the latter specifically for evaluating models on imbalanced data). The Confusion Matrix is a classification performance metric obtained by calculating the number of correctly and incorrectly assigned class labels. In this study, the data are classified into more than two class labels, making it a multi-class classification problem. The multi-class Confusion Matrix can be expressed using Equation (12). Other classification performance criteria derived from the Confusion Matrix [22] include recall (13), precision (14), accuracy (15), f-measure (16), specificity (17), g-means (18) [20], and balanced accuracy (19) [23].

$$M(r, c) = \sum_{i=1}^{m} \big(I(y_i = r) I(h(x_i) = c)\big), \forall_r, c \in \{0, \ldots, q-1\} \quad (12)$$

Where *M* is the value of the multi-class Confusion Matrix, *r* and *c* represent the rows and columns of *M*, and *m* is the total number of data points. *I(.)* is the indicator function, x_i is the *i*-th input classified by *h(.)*, y_i is the true label of input x_i, and *q* is the number of classes [24].

$$recall = \frac{TP}{TP+FN} \quad (13)$$

$$precision = \frac{TP}{TP+FP} \quad (14)$$

$$accuracy = \frac{TP+TN}{n} \quad (15)$$

$$fmeasure = \frac{2TP}{2TP+FP+FN} \quad (16)$$

$$Specificity = \frac{TN}{TN+FP} \quad (17)$$

$$gmeans = \sqrt{recall - specificity} \quad (18)$$

$$Balanced\ Accuracy = \frac{recall + specificity}{2} \quad (19)$$

Where n is the number of data instances. True Positive (TP) and True Negative (TN) represent the number of class labels that are correctly classified, while False Positive (FP) and False Negative (FN) represent the number of class labels that are incorrectly classified. Previous studies have noted that accuracy is not an appropriate metric for evaluating model performance on imbalanced data because minority classes have little influence on the overall accuracy value [9].

In addition to the evaluation metrics derived from the confusion matrix, this study employs macro-averaged and weighted-averaged metrics to provide a more comprehensive evaluation of multi-class imbalanced data. Macro-average computes the arithmetic mean of the evaluation metrics across all classes, treating them equally regardless of their distributions. In contrast, weighted-average calculates the metrics by considering the number of instances (support) in each class, thereby reflecting the actual data distribution. The use of macro-averaged and weighted-averaged metrics is recommended for evaluating multi-class imbalanced classification problems to avoid bias toward majority classes [25], [26].

$$\text{Macro Average Precision} = \frac{\sum_{k=1}^{K} Precision_k}{K} \quad (18)$$

$$\text{Macro Average Recall} = \frac{\sum_{k=1}^{K} Recall_k}{K} \quad (19)$$

$$\text{Macro Average F1} - \text{score} = \frac{\sum_{k=1}^{K} F1_k}{K} \quad (20)$$

$$\text{Weighted Average Precision} = \frac{\sum_{k=1}^{K} nk \,.\, Precision_k}{\sum_{k=1}^{K} n_k} \quad (21)$$

$$\text{Weighted Average Recall} = \frac{\sum_{k=1}^{K} nk \,.\, Recall_k}{\sum_{k=1}^{K} n_k} \quad (22)$$

$$\text{Weighted Average F1} - \text{score} = \frac{\sum_{k=1}^{K} nk \,.\, F1_k}{\sum_{k=1}^{K} n_k} \quad (23)$$

Where K is the number of classes, n_k represents the number of instances (support) in class K, and $Precision_k$, $Recall_k$, and $F1_k$ denote the evaluation metrics for each class. The output of this system includes a classification model for the nutritional status of children under five that has been balanced using the three oversampling methods (SMOTE, WMOTE, and ADASYN), a comparison of the model's performance on the original dataset versus the oversampled datasets, and the identification of the most effective oversampling method for the toddlers' nutritional status classification case.

RESULTS AND DISCUSSION

This study used a dataset of 243 toddler nutritional status data with nine variables. The variables used consisted of four variables representing the toddler's identity, two variables obtained from measurement results, and three variables representing comparisons of the measured variables. The obtained data was then processed. Details of the variables used to determine toddler nutritional status can be seen in Table 1.

Table 1. Variable Description

Name	Variable	Type
jk	Gender	Categorical
BBLahir	Birth Weight (in Gram)	Numeric
TBLahir	Birth Height (in Cm)	Numeric
usia	Age (in Months)	Numeric
BB	Measured Weight (in Gram)	Numeric
TB	Measured Height (in Gram)	Numeric
BBUsia	Age weight	Categorical

Name	Variable	Type
TBUsia	Age Height	Categorical
naikBB	Weight gain status	Categorical

In the preprocessing stage, the dataset is adjusted again by changing variables with categorical types, such as gender, weight for age, height for age, and weight gain status, into numeric labels, which can be seen in Tables 2, 3, 4, and 5.

Table 2. Label of Gender Variable (jk)

Label	Description
1	Male
2	Female

Table 3. Label of Age Weight Variable (BBUsia)

Label	Description
1	Less
2	Normal
3	High Risk
4	Very Less

Table 4. Label of Age Height Variable (TBUsia)

Label	Description
1	Very Short
2	Short
3	Normal
4	Tall

Table 5. Label of Weight Gain Status Variable (naikBB)

Label	Description
1	No Increase
2	No Previous Data
3	Increase

The nutritional status of toddlers is determined based on the variables described in Table 1-5, while the label details of their nutritional status, as well as the amount of data and the distribution of training and testing data, are shown in Table 6. Furthermore, the dataset is scaled to the range 0 to 1.

Table 6. Class Label Description and Distribution of Training and Testing Data

Label	Details	Count	Training	Testing
1	Malnutrition	5	4	1
2	Good Nutrition	192	134	58
3	Risk of Overnutrition	30	21	9
4	Overnutrition	11	8	3
5	Obesity	5	3	2

Based on Table 6, the dataset used in this study is unbalanced. In the next stage, synthetic data is generated from the training data to balance the number of labels. The methods used in creating synthetic data are SMOTE, WMOTE, and ADASYN. In all three methods, the oversampling ratio is set to balance the minority class to an equal number of samples in the majority class. In SMOTE and MWMOTE, the nearest neighbor parameter (k) and in ADASYN (parameter n) are determined dynamically based on the number of minority samples available in the training data using equation (1). This equation ensures that the algorithm stops generating synthetic data exactly when the class distribution is balanced, while preventing system failure when the neighbor parameter exceeds the number of available samples. Details on the number of datasets for each class label after adding synthetic data are shown in Table 7.

Table 7. Description of Synthetic Data Amount

Label	Details	Count of Training Data			
		Original	SMOTE	ADASYN	WMOTE
1	Malnutrition	4	4	4	134
2	Good Nutrition	134	134	134	134
3	Risk of Overnutrition	21	21	21	21
4	Overnutrition	8	8	8	8
5	Obesity	3	134	135	3

After oversampling, the initial dataset of 243 data items increased in volume to achieve equilibrium: 374 with SMOTE, 375 with ADASYN, and 373 with MWMOTE. Using MWMOTE has more complex computational implications than SMOTE, because it involves a difficult minority sample selection process and assigns weights based on proximity to the majority class. Therefore, although this method can improve accuracy and stability, there is a more specific risk of overfitting. This is due to the extremely large amount of training data in the malnutrition and obesity classes, which can cause the resulting synthetic data to be too similar to the original samples. Next, training was carried out using the Multinomial Naive Bayes (MNB) classification model. Based on the prediction and evaluation results, the MNB model's performance was compared, as shown in Table 8.

Table 8. Comparison of Model Performance

Variable	MNB (%)	MNB-SMOTE (%)	MNB-WMOTE (%)	MNB-ADASYN (%)
Accuracy	79,45	67,12	65,75	69,86
Balance Accuracy	20,00	26,55	36,21	27,24
g-means	40,41	49,14	45,08	50,19
Weighted Precision	63,00	64,00	62,00	64,00
Weighted Recall	79,00	67,00	66,00	70,00
Weighted f-measure	70,00	65,00	63,00	67,00
Macro Precision	16,00	18,00	17,00	18,00
Macro Recall	20,00	27,00	36,00	27,00
Macro f-measure	18,00	19,00	19,00	20,00

The comparison results in Table 8 show a decrease in the MNB model's accuracy when oversampling is applied. The decrease in accuracy level was 12.33% in the MNB-SMOTE classification model, 13.70% in the MNB-WMOTE classification model, and 9.59% in the MNB-ADASYN classification model. There was a decrease in the value of Weighted Recall (from 79% to 67-78%) and Weighted F-measure (from 70% to 65-70%) indicating a trade-off (Interest Exchange), but tended to be stable in Weighted Precision indicating that although the model made more predictions on the minority class, the quality of the model's predictions on each class was maintained. This indicates a high level of model stability (robustness). The model does not just guess or produce biased or random predictions; it does so based on synthetic data that has been generated to maintain robust classification quality. The oversampling technique significantly reduced the false-negative rate in the minority class, as evidenced by a 16% increase in Macro Recall from 20.00% to 36.00% in the MNB-WMOTE model. This 16% quantitative increase is clinically crucial in minimizing the risk of undernourished toddlers going undetected, supported by a 16.21% increase in Balanced Accuracy, indicating the model is now much more objective and no longer biased towards the majority class.

MNB-ADASYN produces the highest G-Means value of 50.19% due to its adaptive nature, which focuses the generation of synthetic samples on low-density areas that are difficult for the model to learn. MNB-WMOTE achieves the best global accuracy and stability among balancing methods, thanks to its specific weighting mechanism for critical samples near the decision boundary. This analysis confirms that each technique has functional advantages, ADASYN is very effective in maximizing sensitivity balance across categories, while WMOTE is a more robust solution in maintaining overall accuracy while still paying special attention to extreme data limitations, such as in malnutrition and obesity samples whose number is very limited in this study, namely there are only 4 training data samples in the malnutrition label and 3 samples in the obesity label.

CONCLUSIONS AND SUGGESTIONS

This study concludes that using global accuracy is often irrelevant in highly imbalanced datasets on nutritional status because it tends to mask model failures in detecting minority classes. Failure to detect (false negatives) in the malnutrition category is clinically significant because it risks overlooking infants requiring immediate medical attention. Test results show that applying oversampling techniques such as MWMOTE and ADASYN significantly improves macro-metrics and G-Means, thereby quantitatively reducing the number of false negatives in minority classes.

However, it should be explicitly emphasized that the performance of this study is highly dependent on the training-testing split used for validation and the characteristics of the dataset. Because there were only four training data samples labeled as malnutrition and three samples labeled as obesity, the model has limited generalizability to a broader population. The results of this study demonstrate that the effectiveness of oversampling methods is context-dependent, as changes in the distribution of test samples can drastically impact evaluation scores. Therefore, no single method is best for all conditions; ADASYN excels at adaptive sensitivity, while WMOTE is more stable at maintaining global accuracy in skewed data. Therefore, the results of this study are preliminary and limited, as the very limited data set is unable to represent the diversity of malnutrition conditions in the general population and lacks generalizability for broader clinical use. Therefore, it is recommended that future researchers use a more robust validation scheme. Furthermore, increasing the number of original samples in the minority class is essential to ensure that the generated synthetic data has higher variance and does not simply replicate the original data pattern (overfitting).

REFERENCES

[1] K. Rahmadhita, “Permasalahan Stunting dan Pencegahannya Stunting Problems and Prevention,” *Jurnal ilmiah Kesehatan Sandi Husada*, vol. 11, no. 1, pp. 225–229, 2020, doi: 10.35816/jiskh.v10i2.253.

[2] C. R. Titaley, I. Ariawan, D. Hapsari, A. Muasyaroh, and M. J. Dibley, “Determinants of the stunting of children under two years old in Indonesia: A multilevel analysis of the 2013 Indonesia basic health survey,” *Nutrients*, vol. 11, no. 5, pp. 1–13, May 2019, doi: 10.3390/nu11051106.

[3] Kementerian Kesehatan Republik Indonesia. (2023). Stunting di Indonesia dan Determinannya. [Online]. Available: repository.badankebijakan.kemkes.go.id factsheet PDF. [Accessed: Aug. 7, 2026].

[4] E. R. Arumi, S. A. Subrata, and A. Rahmawati, “Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City,” *Jurnal RESTI*, vol. 7, no. 2, pp. 201–207, Apr. 2023, doi: 10.29207/resti.v7i2.4438.

[5] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, “Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter,” in *2018 6th International Conference on Cyber and IT Service Management, CITSM 2018*, Institute of Electrical and Electronics Engineers Inc., Mar. 2018. doi: 10.1109/CITSM.2018.8674352.

[6] E. Apriliyani and Y. Salim, “Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset,” *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 2, pp. 47–54, 2022.

[7] D. Sartika, Y. Florensia, and M. Utari, "The Effect of the SMOTE Method on the Classification of Toddler Nutritional Status Using the Naive Bayes Method," *SISFOKOM Jurnal (Computer and Information System)*, vol. 14, no. 3, pp. 324–329, 2025.

[8] A. Surya Firmansyah, A. Aziz, and M. Ahsan, "Optimasi K-Nearest Neighbor Menggunakan Algoritma SMOTE Untuk Mengatasi Imbalance Class Pada Klasifikasi Analisis Sentimen," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, pp. 3341–3347, 2023.

[9] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE Dan K-Nearest Neighbor," *Jurnal ISD*, vol. 3, no. 1, pp. 2528–5114, 2018.

[10] S. Barua, M. M. Islam, X. Yao, and K. Murase, "MWMOTE-Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 2, pp. 405–425, Feb. 2014, doi: 10.1109/TKDE.2012.232.

[11] M. A. N. M. Yusuf and M. C. Untoro, "Evaluate of Random Undersampling Method and Majority Weighted Minority Oversampling Technique in Resolve Imabalanced Dataset," *IT Journal Research and Development (ITJRD)*, vol. 8, no. 1, pp. 1–13, Aug. 2023, doi: 10.25299/itjrd.2023.12412.

[12] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "NI-MWMOTE: An improving noise-immunity majority weighted minority oversampling technique for imbalanced classification problems," *Expert Syst. Appl.*, vol. 158, Nov. 2020, doi: 10.1016/j.eswa.2020.113504.

[13] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the International Joint Conference on Neural Networks*, 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

[14] I. Witten, E. Frank, and F. Azuaje, "Data Mining: Practical Machine Learning Tools and Techniques 2nd edition," *Biomed. Eng. Online*, vol. 5, no. 1, Dec. 2006, doi: 10.1186/1475-925x-5-51.

[15] I. Tougui, A. Jilbab, and J. El Mhamdi, "Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning-Based Diagnostic Applications," *Healthc. Inform. Res.*, vol. 27, no. 3, pp. 189–199, Jul. 2021, doi: 10.4258/HIR.2021.27.3.189.

[16] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation," *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61–71, Dec. 2021, doi: 10.5815/ijitcs.2021.06.05.

[17] F. Ramadhani, A. Satria, and I. P. Sari, "Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah," *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 2, pp. 58–62, May 2023, doi: 10.56211/helloworld.v2i2.253.

[18] A. Prawita Ningrum, S. Winarno, and V. Praskatama, "Klasifikasi Kualitas Biji Kedelai Menggunakan Transfer Learning Convolutional Neural Network dan SMOTE," *Journal of Applied Computer Science and Technology*, vol. 5, no. 2, pp. 155–164, Dec. 2024, doi: 10.52158/jacost.v5i2.1002.

[19] T. T. Wong and H. C. Tsai, "Multinomial Naive Bayesian classifier with generalized Dirichlet priors for high-dimensional imbalanced data," *Knowl. Based. Syst.*, vol. 228, pp. 1–8, Sep. 2021, doi: 10.1016/j.knosys.2021.107288.

[20] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS One*, vol. 14, no. 11, Nov. 2019, doi: 10.1371/journal.pone.0224365.

[21] C. Muhamad Sidik Ramdani, A. Rachman Nur, and R. Setiawan, "Comparison of the Multinomial Naive Bayes Algorithm and Decision Tree with the Application of AdaBoost in Sentiment Analysis Reviews PeduliLindungi Application," *International Journal of Information System & Technology*, vol. 6, no. 4, pp. 419–430, 2022.

[22] W. Irmayani, “Visualisasi Data pada Data Mining Menggunakan Metode Klasifikasi Naive Bayes,” *Jurnal Khatulistiwa Informatika*, vol. 9, no. 1, pp. 68–72, 2021.

[23] Y. Pristyanto and A. A. Zein, “Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 9–15, Mar. 2023, doi: 10.32736/sisfokom.v12i1.1399.

[24] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[25] M. Grandini, E. Bagli, and G. Visani, “Metrics for Multi-Class Classification: an Overview,” *A White Paper*, pp. 1–17, Aug. 2020, [Online]. Available: http://arxiv.org/abs/2008.05756

[26] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.