Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online at: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 4 (2024) 496 - 505 e-ISSN: 2580-0760

The Effect of Resampling Techniques on Model Performance Classification of Maternal Health Risks

Nia Mauliza[1*], Aisha Shakila Iedwan[2], Yoga Pristyanto[3], Anggit Dwi Hartanto[4], Arif Nur Rohman[5]
[1,2,3,4,5] Department of Information Systems, Faculty of Computer Science, Amikom Yogyakarta University, Yogyakarta, Indonesia
[1]niamauliza08@students.amikom.ac.id, [2]aisha.shakila@students.amikom.ac.id, [3]yoga.pristyanto@amikom.ac.id, [4]anggit@amikom.ac.id, [5]arifrahman@amikom.ac.id

Abstract

Indonesia's maternal mortality rate was the second highest in ASEAN, reflecting the problem of class imbalance in maternal health data. This research aimed to improve prediction accuracy in the classification of pregnant women's diseases through the application of various resampling methods. The methods used in this research included Synthetic Minority Over-sampling Technique (SMOTE), SMOTE-Edited Nearest Neighbor (SMOTE-ENN), Adaptive Synthetic Sampling (ADASYN), and ADASYN-ENN, using five classification algorithms: Decision Tree, K-Nearest Neighbor (KNN), Naïve Bayes, Random Forest, and Support Vector Machine (SVM). Performance evaluation was carried out using accuracy, precision, recall, and F1-score metrics to determine the best method and algorithm. The results showed that the SMOTE-ENN and ADASYN-ENN methods significantly improved the model's performance in predicting maternal disease. Random Forest and Decision Tree algorithms showed the best results in terms of accuracy and consistency. These findings provided practical guidance for the application of resampling techniques in the classification of pregnant women's health data, which could contribute to improving the quality of maternal health services in Indonesia.

Keywords: class imbalance; resampling methods; classification algorithms; maternal health; prediction accuracy; machine learning

How to Cite: Nia Mauliza, Aisha Shakila Iedwan, Yoga Pristyanto, Anggit Dwi Hartanto, and Arif Nur Rohman, "The Effect of Resampling Techniques on Model Performance Classification of Maternal Health Risks", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 8, no. 4, pp. 496 - 505, Aug. 2024.
DOI: https://doi.org/10.29207/resti.v8i4.5934

1. Introduction

The high maternal mortality rate was a global concern, despite significant reductions due to medical advances in various countries, including developing nations. Data from the Ministry of Health's Maternal Perinatal Death Notification (MPDN) indicated that Indonesia had the second highest maternal and infant mortality rate in ASEAN. The number of maternal deaths increased from 4,005 in 2022 to 4,129 in 2023. Additionally, the number of infant deaths rose from 20,882 to 29,945 within the same period. This condition underscored the necessity of enhancing maternal health prevention and intervention efforts to concurrently reduce maternal and infant mortality rates [1]. The application of technology in the world of health is increasingly becoming the main focus to improve efficiency, accuracy and quality of service, especially to optimize various operational aspects in hospitals and the health sector as a whole [2]. One step that can be taken is to maximize the role of technology, especially with advances in the fields of artificial intelligence and machine learning. Machine learning is considered important because of its capabilities in the classification and analysis of infections and diseases. This allows difficult-to-diagnose diseases to be more easily managed. Machine learning, as an advanced method, can make decisions independently and without requiring reprogramming by humans [3].

In cases related to maternal health classification, research using machine learning has been carried out. Research [4] using the decision tree method produces 90% accuracy, the KNN method produces 86% accuracy and the Naïve Bayes method produces 65% accuracy. Research [5] using the decision tree method produces 89% accuracy, light GBM produces 84% accuracy, catBoost produces 83% accuracy, random forest produces 81% accuracy, gradient boosting machines produce 73% accuracy and the KNN method produces 68% accuracy. In this research, cases of class

Received: 19-07-2024 | Accepted: 13-08-2024 | Published Online: 19-08-2024

imbalance were found in the data used, where the majority class was larger than the minority class. This condition of imbalance can cause the resulting prediction results to be biased so it is feared that it could affect the accuracy of the research results [6]. The minority class is more difficult to predict than the majority class even though sometimes the minority class has more important information [7]. Therefore, special strategies are needed to overcome this class imbalance so that machine learning models can provide more accurate and useful predictions.

The solution that can be taken is to handle class imbalance in the dataset because overcoming class imbalance problems can improve the performance of classification methods in terms of accuracy, sensitivity and specificity [8]. There are two approaches to dealing with class imbalance problems, namely approaches at the algorithmic level and approaches at the data level [9], [10]. The algorithmic level approach is more focused on improving the classifier algorithm, while the data level approach is focused on using various data resampling techniques to make the class distribution balanced. Between the two approaches, the data level approach is proven to have better performance than the algorithmic level approach. This approach involves three methods that can be used, namely over-sampling, under-sampling and hybrid-sampling [11]. Various studies related to handling class imbalance using a level approach have been carried out. For example, research [12] uses three sampling methods, namely ADASYN, SMOTE, and the SMOTE-ENN combination applied to the KNN, Adaboost, and Random Forest classification models. This research provides results that the use of the resampling method provides an increase in accuracy values. The research results show that the combination of ADASYN and Random Forest produces the best performance, with higher accuracy, precision, true positive rate, true negative rate and g-mean values compared to other models

Apart from that, research [13] also highlights the problem of class imbalance. This research compares 6 resampling methods on 15 datasets that will be balanced. Based on the mean of ranks, this research showed that the best resampling techniques is SMOTEENN(1,700), ADASYN(2,767), RUS(3,333), SMOTETomek(3,867), SMOTE(4,000), ROS(5,333). study [14] compared classification methods on heart disease datasets. The results of this study show that the use of the SMOTE method for oversampling heart disease datasets produces effective improvements in increasing accuracy and precision in classification using SVM and Logistic Regression (LR). Handling class imbalance with a similar approach was also carried out in research [15] addressing the problem of unbalanced classes in classifying cases of diarrhoea in toddlers in Indonesia. This research addresses the problem of unbalanced classes using the Synthetic Minority Oversampling Technique (SMOTE) method and several ensemble techniques, such as Random Forest, AdaBoost, and XGBoost. The results show that all SMOTE-based methods provide competitive performance, with SMOTE-XGBoost achieving the highest level of accuracy (0.88), precision (0.96), and F1-score (0.86).

Various studies have shown that treating class imbalance is a crucial step to improving the effectiveness of classification model performance. This research aims to compare the effect of four resampling methods, namely SMOTE, SMOTE-ENN, ADDASYN, and ADDASYN-ENN on the performance of five classification algorithm models. The five classification algorithms, namely decision tree, naïve babies, KNN, random forest, and SVM are used in maternal health classification. The performance of these models will be evaluated using a confusion matrix to analyze the accuracy, precision, recall and F1-score of each model, then the results will be compared with each other. The main aim of this study is to provide an in-depth understanding of which resampling methods are most effective in dealing with class imbalance in maternal health datasets. In addition, this research will also identify classification algorithm models that provide the best performance after applying the resampling method, with the ultimate aim of providing practical recommendations for the use of these techniques in improving the quality of disease classification in the population of pregnant women.

2. Research Methods

In this research, the research stages include data acquisition, resampling, classification model and model evaluation as shown in Figure 1.

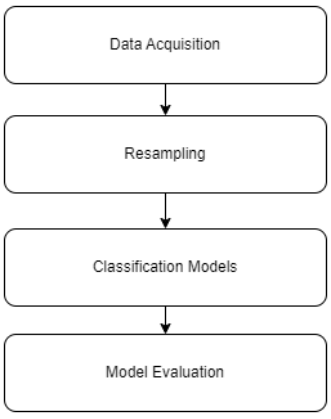

Figure 1. Research Method

2.1 Data Acquisition

The dataset used in this research is the Maternal Health Risk dataset taken from the UCI Machine Learning Repository, which provides a collection of ready-to-use datasets for machine learning implementation. Researchers collected data from various places using an IoT-based risk monitoring system. This dataset has no missing values. Table 1 shows the sample dataset used.

Table 1. Software and supporting hardware

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
35	120	60	6.1	98	76	Low Risk
32	120	90	6.9	98	70	Mid Risk
42	130	80	18	98	70	High Risk

In this study, samples were taken randomly to ensure optimal representativeness and minimize selection bias. The dataset was divided using a consistent scheme: 80% of the data was used for training and 20% for testing. This scheme was uniformly applied to evaluate the random sampling method, allowing for accurate performance assessment of the classification models and fair comparison between them.

2.2 Resampling

Addressing class imbalance in datasets was critical for improving the performance of classification models. In this research, four resampling methods were used and compared: SMOTE (Synthetic Minority Oversampling Technique), SMOTE-ENN (SMOTE with Edited Nearest Neighbors), ADASYN (Adaptive Synthetic Sampling), and ADASYN-ENN.

The Synthetic Minority Oversampling Technique (SMOTE) approach was employed to balance unbalanced classes through oversampling. SMOTE created synthetic samples from minority data to equalize the number with the majority data. This method was effective in reducing overfitting, a common issue with oversampling techniques. However, a drawback of SMOTE was that the synthetic data generated did not account for the majority class, potentially causing data overlap [16], [17].

SMOTE-ENN was a resampling method that combined oversampling with SMOTE and undersampling with ENN. This approach initially oversampled the minority class through interpolation and subsequently removed redundant samples using the ENN method. Ultimately, this technique produced data with a balanced class distribution, suitable for use with machine learning algorithms to achieve the desired performance [18] [19].

ADASYN was a resampling technique designed to generate synthetic data for minority classes, with a particular focus on observations that were difficult to learn. This approach utilized the distribution of their nearest neighbors to address the problem of data imbalance [20].

Combining ADASYN with ENN (ADASYN-ENN) involves applying ADASYN first to create synthetic samples for the minority class, and then using the ENN to remove irrelevant samples from the majority class. This hybrid approach ensures that the dataset remains balanced and clean, thereby improving classifier performance by reducing overfitting and improving the quality of training data [21].

2.3 Classification Method

After going through the resampling process, the next step is to implement the classification model. This research adopts five types of classification models that have been proven effective in various data analysis contexts: Random Forest, Decision Tree, SVM (Support Vector Machine), Naïve Bayes, and KNN (K-Nearest Neighbors).

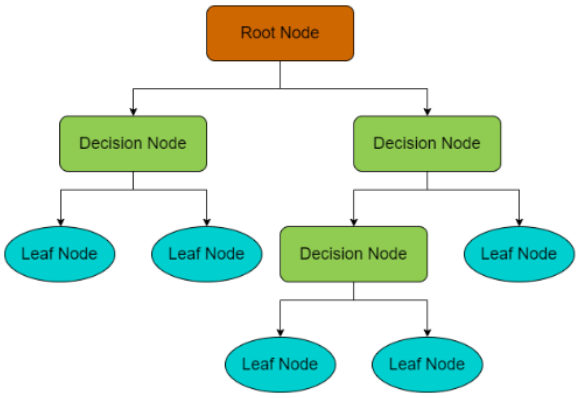

Figure 2. Decision Tree

Decision trees are a very popular algorithm and one of the most widely used models in classification [22]. This method uses a tree structure where each node represents the attribute being tested, each branch shows the test results, and each leaf represents a specific class group. The top node is the root node, which is usually the most influential attribute. Decision Trees generally look for solutions from top to bottom as shown in Figure 2. To classify unknown data, attribute values are tested by following the path from root to leaf, and then the new data class is predicted. To build a decision tree, one of the metrics used is entropy as seen in Equation 1.

$$Entropy\ (S) = \sum_{i=1}^{n} -P_i log_2 p_i \quad (1)$$

S denotes the set of cases, partitioned into N subsets, where p-i represents the probability derived from the number of classes divided by the total cases [23-25].

K-Nearest Neighbor (KNN) is a classification method that is based on the distance between new data and several of its nearest neighbors. This method determines a new data class based on the majority class of the nearest neighbors [26]. In implementing the KNN algorithm, Equation 2 is used [27].

$$d_i = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2)$$

d represents the distance between the training data (X) and the test data (Y), where N denotes the data dimension and I signifies variable data.

Naïve Bayes is a statistical algorithm used to calculate the probability of a hypothesis. In the context of classification, Naïve Bayes estimates the probability of a label based on the attributes it has and selects the label that has the highest probability as a classification result [28].

Random Forest is an algorithm that consists of several decision trees. This algorithm is used to make predictions by dividing data into several categories based on certain attributes and making decisions by comparing these values. By combining the results of several decision trees, Random Forest can increase prediction accuracy and reduce the risk of overfitting which often occurs in single decision trees [29-31]. The formulas and equations for obtaining random forests can be seen in Equation 3 [32].

$$Entrophy\ (S) = \sum_{i=1}^{n} -P_i * Log_2 P_i \quad (3)$$

Support Vector Machine (SVM) is an algorithm that is often used for data classification analysis [33]. The way the SVM (Support Vector Machine) model works for classification is by trying to separate each class or label by making as wide a margin as possible between the classes. SVM searches for an optimal hyperplane that maximizes the distance between data points from different classes, resulting in a clear and robust separation between categories [34-35]. The hyperlane value can also be formulated as Equation 4.

$$f(x) = In^T x + b \quad (4)$$

2.4 Model Evaluation

Performance evaluation is carried out using a confusion matrix, which describes how well the model or algorithm can predict data classes. This matrix provides an overview of the number of correct predictions (True Positive and True Negative) as well as prediction errors (False Positive and False Negative) [36]. which is used to calculate evaluation metrics such as accuracy, precision, recall, and F1-Score [37]. The matrix is described in Table 2.

Table 2. Description of evaluation matrix

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

Accuracy measures how often the model classifies correctly overall. The accuracy formula can be written with Equation 5.

$$Accuracy = \frac{TP+TN}{Total} \quad (5)$$

Precision measures how often positive predictions from the model are correct. The precision formula can be written with Equation 6:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

Recall measures how often the model predicts positive when the actual class is positive. The recall formula can be written with Equation 7.

$$Recall = \frac{Tp}{TP+FN} \quad (7)$$

F1-Score is the harmonic average of precision and recall. The F1-Score formula can be written with Equation 8.

$$F1 - Score = 2 \times \frac{precision \times recall}{precision+recall} \quad (8)$$

3. Results and Discussions

3.1. Results

In the results section, the discussion focused on the influence of various resampling methods on the performance of five classification algorithms: Decision Tree, K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, and Support Vector Machine (SVM). The evaluation concentrated on metrics such as accuracy, precision, recall, and F1-score to determine the extent to which the four resampling methods—SMOTE, SMOTE-ENN, ADASYN, and ADASYN-ENN—improved the performance of each algorithm in addressing class imbalance in maternal health datasets.

The study utilized the Maternal Health dataset, which comprised 1,013 data points across seven variables: Age, Systolic Blood Pressure (SystolicBP), Diastolic Blood Pressure (DiastolicBP), Respiratory Rate (BS), Body Temperature (BodyTemp), Heart Rate (HeartRate), and the target variable Risk Level. The initial analysis revealed a significant class imbalance within this dataset. The class distribution prior to resampling was as follows: Low Risk with 406 cases, Mid Risk with 336 cases, and High Risk with 272 cases. This imbalance was illustrated in the class distribution visualization in Figure 3.

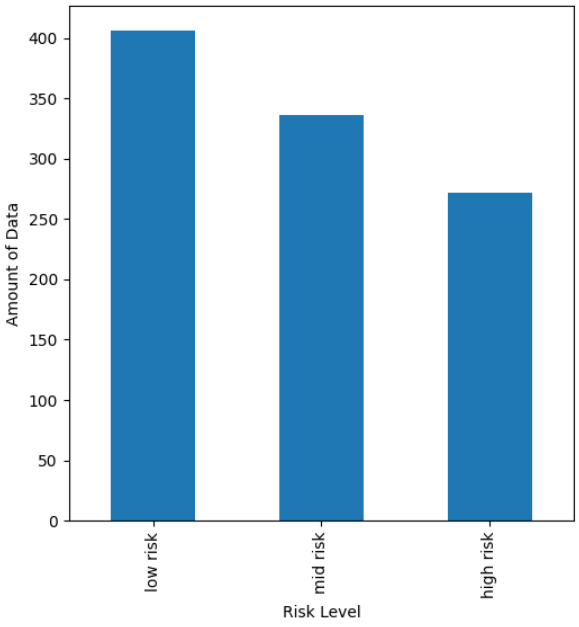

Figure 3. Class distribution before resampling

After applying SMOTE (Synthetic Minority Oversampling Technique), the class distribution became balanced with each class having 406 cases. SMOTE is effective in adding synthetic samples to the minority class so that the number is balanced with the majority class. A visualization of class distribution can be seen in Figure 4.

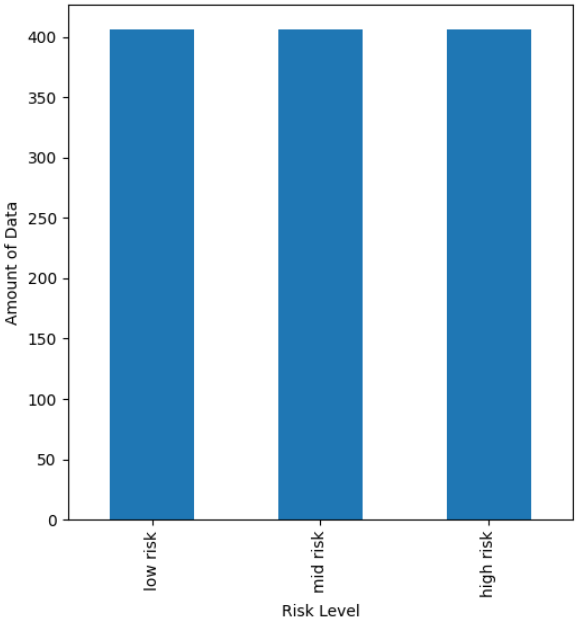

Figure 4. Class distribution after SMOTE resampling

The application of SMOTE-ENN resulted in a more balanced class distribution with 300 cases for High Risk, 235 cases for Mid Risk, and 226 cases for Low Risk. SMOTE-ENN combines oversampling with SMOTE and undersampling with Edited Nearest Neighbors (ENN) to clean datasets from oversampling and noise. A visualization of class distribution is attached in Figure 5.

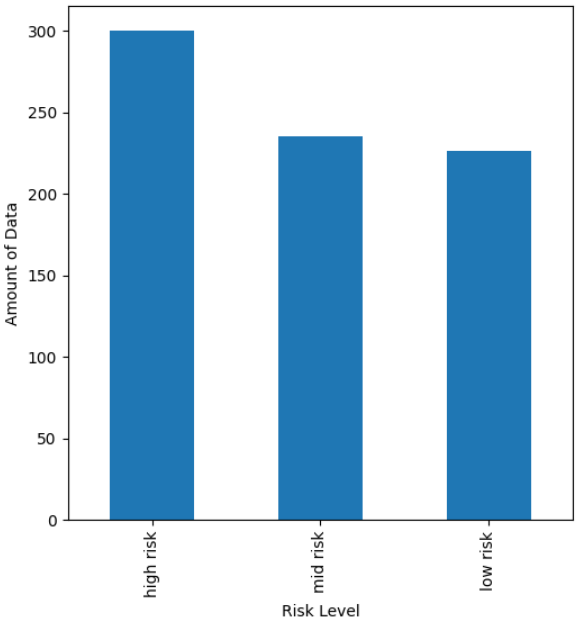

Figure 5. Class distribution after SMOTE-ENN resampling

ADASYN (Adaptive Synthetic Sampling) is used to address class imbalance by focusing on difficult-to-study observations from minority classes. After resampling with ADASYN, the class distribution became 420 cases for High Risk, 378 cases for Mid Risk, and 406 cases for Low Risk. ADASYN improves the representation of minority data to improve the performance of classification models. A visualization of class distribution can be seen in Figure 6.

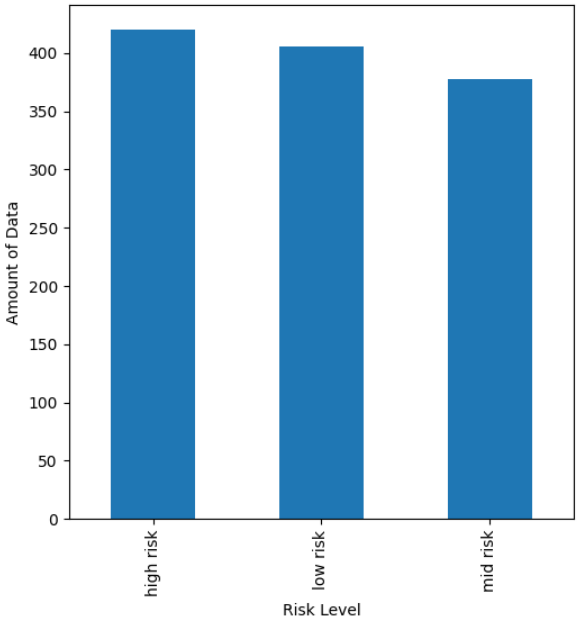

Figure 6. Class distribution after ADASYN resampling

Finally, ADASYN-ENN produces a class distribution with 272 cases for High Risk, 192 cases for Mid Risk, and 232 cases for Low Risk after the combination of ADASYN and ENN resampling. This approach helps in maintaining data diversity and reduces overfitting in the model. A visualization of class distribution can be seen in Figure 7.

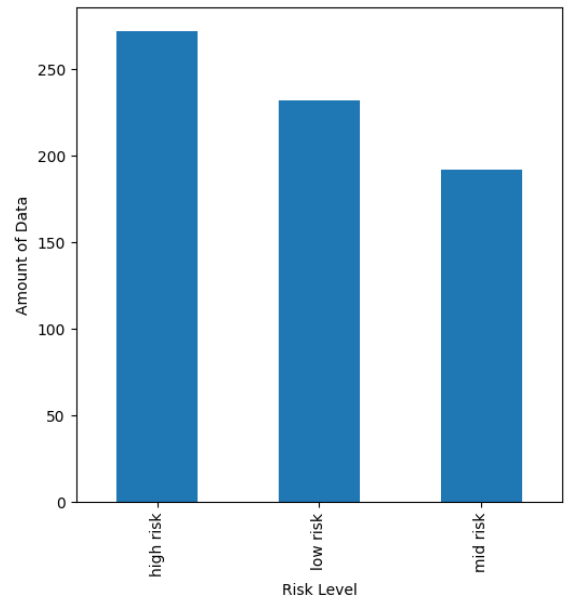

Figure 7. Class distribution after ADASYN-ENN resampling

After resampling to address the class imbalance in the maternal health dataset, five primary classification algorithms were evaluated: Decision Tree, K-Nearest Neighbors (KNN), Naïve Bayes, Random Forest, and Support Vector Machine (SVM). The evaluation focused on accuracy, precision, recall, and F1-score metrics to determine the relative performance of each method following resampling.

Table 3 presents the initial performance of the five primary methods in classification before employing the resampling methods to manage class imbalance in the maternal health dataset.

Table 3. Classification Model Performance Without Resampling

	Accuracy	Precision	Recall	F1-Score
Decesion Tree	81%	81%	81%	81%
KNN	66%	68%	67%	66%
Naive Bayes	57%	59%	58%	53%
Random Forest	81%	81%	81%	81%
SVM	66%	68%	66%	63%

Table 3 provided a preliminary evaluation of the performance of five classification methods before applying resampling techniques to the maternal health dataset. Decision Tree and Random Forest exhibited similar results, with accuracy, precision, recall, and F1-score reaching 81%. KNN showed slightly lower performance, with an accuracy of 66%, precision of 68%, recall of 67%, and F1-score of 66%. Naïve Bayes demonstrated lower performance, with 57% accuracy, 59% precision, 58% recall, and 53% F1-score. SVM had comparable results to KNN, with 66% accuracy, 68% precision, 66% recall, and 63% F1-score.

A comparison of classification accuracy values for the Maternal Health dataset before resampling is visualized in Figure 8.

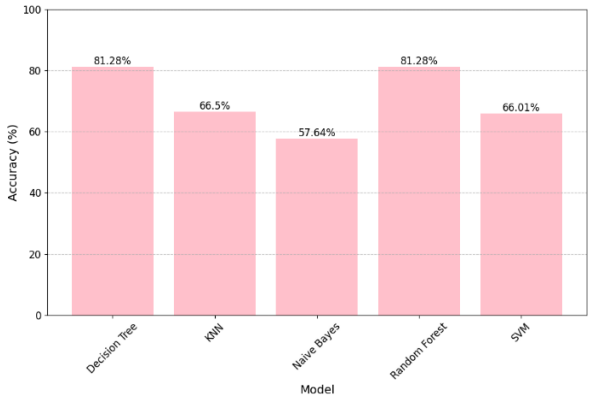

Figure 8. Comparison of classification accuracy values before resampling

This initial evaluation provides a basis for comparing performance improvements after applying the resampling technique.

Table 4 shows the performance evaluation results of five classification methods after applying the SMOTE resampling technique to the maternal health dataset.

Table 4 Classification Model Performance after SMOTE Resampling

	Accuracy	Precision	Recall	F1-Score
Decesion Tree	83%	85%	83%	83%
KNN	73%	75%	74%	74%
Naive Bayes	60%	62%	61%	57%
Random Forest	82%	84%	82%	83%
SVM	67%	69%	67%	68%

After the SMOTE resampling technique was applied to the maternal health dataset, the evaluation results demonstrated improved performance of several primary classification methods. Decision Tree recorded an accuracy of 83%, with a precision of 85%, a recall of 83%, and an F1-score of 83%. KNN exhibited significant improvements, with an accuracy of 73%, a precision of 75%, a recall of 75%, and an F1-score of 74%. Naïve Bayes experienced a slight increase, achieving an accuracy of 60%, a precision of 62%, a recall of 61%, and an F1-score of 57%. Random Forest yielded excellent results, with an accuracy of 82%, a precision of 84%, a recall of 82%, and an F1-score of 83%. Meanwhile, SVM achieved an accuracy of 67%, a precision of 69%, a recall of 67%, and an F1-score of 68%. These results indicated that the application of SMOTE in resampling improved the model's ability to handle class imbalance in maternal health datasets, particularly with the Decision Tree and Random Forest methods, which showed significant performance enhancements. The comparison of classification accuracy values for the Maternal Health dataset after resampling using the SMOTE method was visualized in Figure 9.

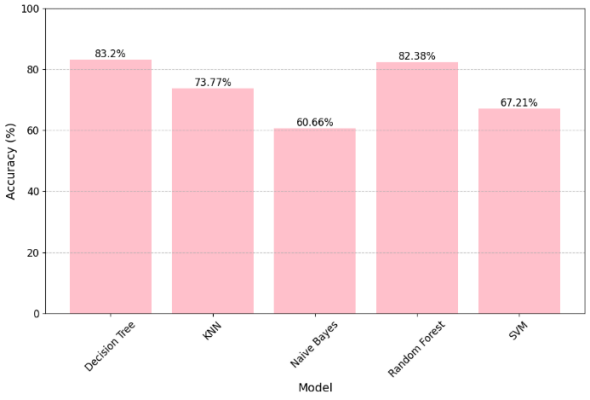

Figure 9. Comparison of resampling classification accuracy values using the SMOTE method

Table 5 shows the performance evaluation results of five classification methods after applying the SMOTE-ENN resampling technique to the maternal health dataset.

Table 5 Classification Model Performance after SMOTE-ENN Resampling

	Accuracy	Precision	Recall	F1-Score
Decesion Tree	96%	97%	97%	97%
KNN	79%	81%	79%	80%
Naive Bayes	63%	64%	63%	63%
Random Forest	96%	96%	96%	96%
SVM	69%	72%	69%	69%

Evaluation results after applying resampling using the SMOTE-ENN method show significant improvements in the performance of the five classification methods used on the maternal health dataset. Decision Tree and Random Forest show excellent results with all performance metrics (accuracy, precision, recall, and F1-score) reaching 96%, showing consistent improvement from before. KNN also showed a marked improvement with 79% accuracy, 81% precision, 79% recall, and 80% F1 score. Naïve Bayes showed stable performance with slight improvements, achieving 63% accuracy, 64% precision, 63% recall, and 63% F1-score. SVM showed visible improvements in precision (72%) and F1-score (69%), although recall (69%) is slightly lower than before. Overall, the use of SMOTE-ENN was effective in improving the model's ability to resolve class imbalance in this dataset.

A comparison of classification accuracy values for the Maternal Health dataset after resampling using the SMOTE-ENN method is visualized in Figure 10.

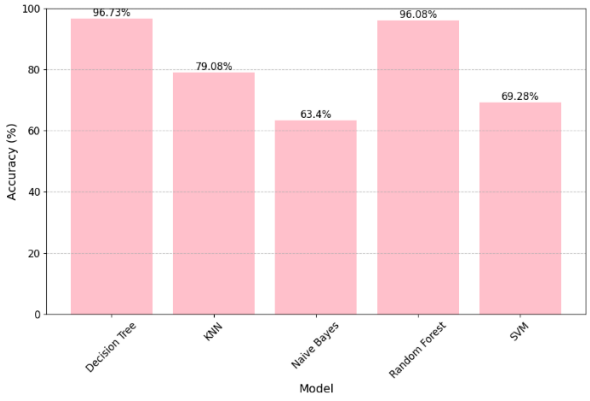

Figure 10 Comparison of resampling classification accuracy values using the SMOTE-ENN method

Table 6 shows the performance evaluation results of five classification methods after applying the ADASYN resampling technique to the maternal health dataset.

Table 6 Classification Model Performance after ADASYN Resampling

	Accuracy	Precision	Recall	F1-Score
Decesion Tree	80%	81%	80%	80%
KNN	65%	66%	66%	65%
Naive Bayes	55%	56%	56%	53%
Random Forest	83%	84%	83%	83%
SVM	66%	68%	66%	65%

The results of this evaluation show the performance of five classification methods after applying resampling using the ADASYN method. Decision Tree showed an accuracy of 80%, with precision, recall, and F1-score reaching 81%, 80%, and 80% respectively. KNN showed slightly lower performance with 65% accuracy, 66% precision, 66% recall, and 65% F1 score. Naïve Bayes showed lower results than other methods, with 55% accuracy, 56% precision, 56% recall, and 53% F1-score. Random Forest showed good performance with 83% accuracy, 84% precision, 83% recall, and 83% F1 score. SVM showed comparable results to KNN, with 66% accuracy, 68% precision, 66% recall, and 65% F1-score. Overall, resampling using ADASYN succeeded in improving the performance of several classification methods, especially Decision Tree and Random Forest, although Naïve Bayes showed a smaller improvement.

A comparison of classification accuracy values for the Maternal Health dataset after resampling using the ADASYN method is visualized in Figure 11.

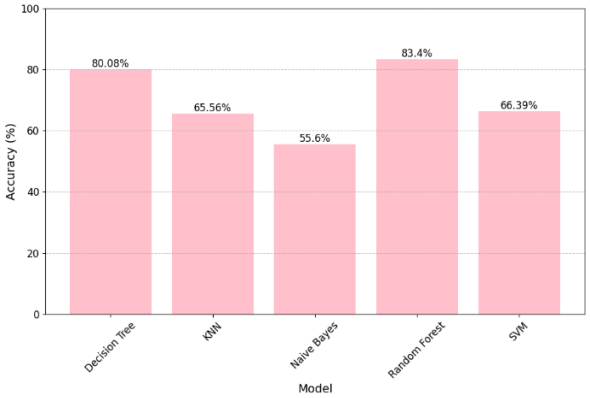

Figure 11 Comparison of resampling classification accuracy values using the ADASYN method

Table 7 shows the performance evaluation results of five classification methods after applying the ADASYN-ENN resampling technique to the maternal health dataset.

Table 7 Classification Model Performance after ADASYN-ENN Resampling

	Accuracy	Precision	Recall	F1-Score
Decesion Tree	92%	93%	92%	92%
KNN	81%	82%	81%	82%
Naive Bayes	76%	79%	76%	77%
Random Forest	90%	91%	91%	91%
SVM	75%	79%	75%	75%

The results of this evaluation show the performance of five classification methods after applying resampling using the ADASYN-ENN method. Decision Tree showed an accuracy of 92%, with precision, recall, and F1-score reaching 93%, 92%, and 92% respectively. KNN showed excellent performance with 81% accuracy, 82% precision, 81% recall, and 82% F1 score. Naïve Bayes showed significant improvements compared to before, with 76% accuracy, 79% precision, 76% recall, and 77% F1-score. Random Forest showed strong performance with 90% accuracy, 91% precision, 91% recall, and 91% F1 score. SVM also showed improvement with 75% accuracy, 79% precision, 75% recall, and 75% F1 score. Overall, resampling using ADASYN-ENN successfully improved the performance of almost all classification methods, demonstrating its effectiveness in dealing with class imbalance in maternal health datasets.

A comparison of classification accuracy values for the Maternal Health dataset after resampling using the ADASYN-ENN method is visualized in Figure 12.

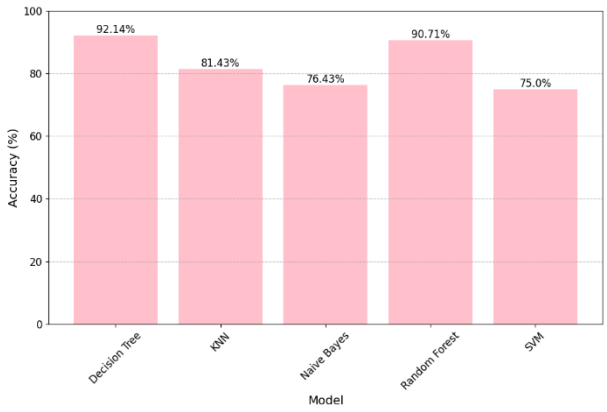

Figure 12 Comparison of resampling classification accuracy values using the ADASYN-ENN method

3.2 Discussions

This research demonstrates that resampling techniques such as SMOTE, SMOTE-ENN, ADASYN, and ADASYN-ENN can significantly enhance the performance of classification models when addressing class imbalance in maternal health datasets. Through a comprehensive evaluation, it was found that SMOTE-ENN is the most effective resampling method, with substantial improvements in accuracy, precision, recall, and F1-score for classification models such as Decision Tree and Random Forest. Specifically, Decision Tree and Random Forest achieved an accuracy of up to 96% with the application of SMOTE-ENN, highlighting its capability to effectively address class imbalance issues and enhance data quality for better modelling.

Although SMOTE is less effective compared to SMOTE-ENN, it still provides significant improvements, particularly for Decision Tree and Random Forest models, achieving accuracies of 83% and 82% respectively. SMOTE improves class distribution by adding synthetic samples, thereby enhancing model performance. However, its results are not as strong as SMOTE-ENN because SMOTE does not handle outliers and noise as effectively as SMOTE-ENN, which combines oversampling with undersampling to clean the data from noise. Conversely, the ADASYN technique, which focuses on improving challenging minority data, showed more variable improvements, with accuracies of 80% for the Decision Tree and 83% for the Random Forest. ADASYN tends to introduce additional noise that can affect data quality and overall model performance. ADASYN-ENN, while attempting to address these issues by incorporating ENN, does not fully resolve the problems related to noise and outliers, resulting in less consistent and stable outcomes compared to SMOTE-ENN and SMOTE.

Overall, this research affirms that SMOTE-ENN is the best resampling method for tackling class imbalance in maternal health datasets. SMOTE-ENN effectively combines oversampling and undersampling techniques to enhance data balance, reduce noise, and achieve superior and more stable model performance, with the highest accuracy reaching 96%. SMOTE is effective but not as superior as SMOTE-ENN, while ADASYN and ADASYN-ENN provide more variable and less significant results, with accuracies of 92% for ADASYN-ENN and 90% for Random Forest. This approach underscores the importance of selecting the appropriate resampling technique to improve the accuracy and reliability of classification models in the context of health datasets.

The application of resampling techniques to handle imbalanced data proves effective because these techniques address class imbalance issues that often cause classification models to be biased towards the majority class. When a dataset has an imbalanced class distribution, models tend to predict the majority class more accurately while neglecting the minority class. By balancing the class distribution through resampling, models can learn from a more equitable sample size, leading to improved prediction accuracy and reduced bias. Techniques like SMOTE and ADASYN add minority samples, providing models the opportunity to learn from more representative patterns in the minority class, while SMOTE-ENN also reduces noise that could disrupt the training process, resulting in more robust and accurate models.

After achieving good model performance, this research offers two types of benefits. First, the technical benefit of producing a more accurate maternal health risk classification model, aids medical practitioners in more precisely identifying maternal health risk levels. This research also develops resampling techniques (SMOTE, SMOTE-ENN, ADASYN, and ADASYN-ENN) and evaluates various classification algorithms (Decision Tree, Naive Bayes, KNN, Random Forest, and SVM), providing insights into the effectiveness and performance of each technique and algorithm. Second, the non-technical benefits include improving the quality of maternal health services by assisting hospitals and healthcare providers in making more informed and data-driven medical decisions, potentially reducing maternal and infant mortality rates in Indonesia. Additionally, the results of this research can serve as a reference for future researchers in developing or applying resampling techniques and classification algorithms across various domains. This research also supports increasing the operational efficiency of healthcare institutions and helps the government in formulating more effective health policies.

4. Conclusions

This research demonstrates that resampling techniques such as SMOTE, SMOTE-ENN, ADASYN, and ADASYN-ENN can significantly enhance the performance of classification models in handling class imbalance within maternal health datasets. The comprehensive evaluation revealed that SMOTE-ENN is the most effective technique for improving accuracy, precision, recall, and F1-score across various

classification models, including Decision Trees and Random Forests. Specifically, Decision Tree and Random Forest achieved an impressive accuracy of 96% using SMOTE-ENN. While SMOTE also provided notable performance improvements, particularly for Random Forest and Decision Tree models, it did not perform as well as SMOTE-ENN. In contrast, ADASYN and ADASYN-ENN offered some improvements but yielded more variable results compared to SMOTE and SMOTE-ENN. The SMOTE-ENN method effectively combines SMOTE's advantage of increasing minority samples with Edited Nearest Neighbors (ENN) to reduce outliers, resulting in a more balanced and higher-quality dataset. ADASYN, which targets challenging data points, tends to introduce additional noise, and ADASYN-ENN, although incorporating ENN, does not fully resolve issues related to noise and outliers. SMOTE-ENN's simplicity and consistency make it a more reliable and stable method for enhancing classification performance compared to ADASYN and ADASYN-ENN. For future research, it is recommended to test SMOTE-ENN and other resampling techniques on different datasets with varying class imbalances. Additionally, exploring the combination of these resampling techniques with advanced machine learning methods, such as neural networks and ensemble methods, could yield more optimal results. Investigating the integration of resampling techniques with more sophisticated models may further enhance the accuracy and reliability of classification systems across diverse domains.

Acknowledgements

Thanks, Information System Study Program, Faculty of Computer Science, Department of Research and Community Service, and Amikom Yogyakarta University, for funding, supporting this research and striving to finish this research.

References

[1] Rokom, "For Mother and Baby to be Safe," Healthy My Country, Jan. 25, 2024. Accessed: Jul. 17, 2024. [Online]. Available: https://sehatnegeriku.kemkes.go.id/baca/blog/20240125/3944849/agar-ibu-dan-bayi-selamat/

[2] "View of Applications and Benefits of Machine Learning in Hospitals." https://www.jurnal.medanresourcecenter.org/index.php/MULTIVERSE/article/view/1207/1089

[3] Agnes, "The Benefits of Machine Learning in the Healthcare Industry," DQLab | Indonesian R Python Online Data Science Course, Jul. 21, 2022. Accessed: Jul. 17, 2024. [Online]. Available: https://dqlab.id/manfaat-machine-learning-di-industri-healthcare

[4] "View of Comparative Analysis of Machine Learning Algorithms for Classifying Risk Levels of Pregnant Women." https://journal-stiayappimakassar.ac.id/index.php/srj/article/view/846/867

[5] H. B. Mutlu, F. Durmaz, N. Yücel, E. CengiL, and M. Yildirim, "Prediction of Maternal Health Risk with Traditional Machine Learning Methods," NATURENGS MTU Journal of Engineering and Natural Sciences Malatya Turgut Ozal University, Jun. 2023, doi: 10.46572/naturengs.1293185.

[6] "An Optimized View of Unbalanced Data on Drug Target Interactions with Sampling and Ensemble Support Vector Machine." https://jtiik.ub.ac.id/index.php/jtiik/article/view/2857/pdf

[7] "SMOTE–IPF: Addressing the noisy and borderline examples problem in imbalanced classification by a re-sampling method with filtering," Information Sciences, vol. 291, pp. 184–203, doi: 10.1016/j.ins.2014.08.051

[8] A. M. W. Saputra and A. W. Wijayanto, "IMPLEMENTATION OF ENSEMBLE TECHNIQUES FOR DIARRHEA CASES CLASSIFICATION OF UNDER-FIVE CHILDREN IN INDONESIA," JITK (Journal of Computer Science and Technology), vol. 6, no. 2, pp. 175–180, Feb. 2021, doi: 10.33480/jitk.v6i2.1935.

[9] C. Kaope and Y. Pristyanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 22, no. 2, pp. 227–238, Mar. 2023, doi: 10.30812/matrik.v22i2.2515.

[10] M. Fatourechi, R. K. Ward, S. G. Mason, J. Huggins, A. Schlögl, and G. E. Birch, "Comparison of Evaluation Metrics in Classification Applications with Imbalanced Datasets," in 2008 Seventh International Conference on Machine Learning and Applications, 2008. Accessed: Jul. 17, 2024. [Online]. Available: http://dx.doi.org/10.1109/icmla.2008.34

[11] I. Lin, O. Loyola-González, R. Monroy, and M. A. Medina-Pérez, "A Review of Fuzzy and Pattern-Based Approaches for Class Imbalance Problems," Applied Sciences, vol. 11, no. 14, p. 6310, Jul. 2021, doi: 10.3390/app11146310.

[12] C. Kaope and Y. Pristyanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," MATRIK: Journal of Management, Information Engineering and Computer Engineering, vol. 22, no. 2, pp. 227–238, Mar. 2023, doi: 10.30812/matrik.v22i2.2515.

[13] Y. A. Sir and A. H. H. Soepranoto, "Data Resampling Approach to Handling Class Imbalance Problems," Journal of Computers and Informatics, vol. 10, no. 1, pp. 31–38, Mar. 2022, doi: 10.35508/jicon.v10i1.6554.

[14] C. Sonjaya, A. F. Nur Masruriyah, D. Sulistya Kusumaningrum, and A. Rizky Pratama, "The Performance Comparison of Classification Algorithms in Order to Detecting Heart Disease," INTERNAL (Information System Journal), vol. 5, no. 2, pp. 166–175, Dec. 2022, doi: 10.32627/internal.v5i2.595

[15] A. M. W. Saputra and A. W. Wijayanto, "IMPLEMENTATION OF ENSEMBLE TECHNIQUES FOR DIARRHEA CASES CLASSIFICATION OF UNDER-FIVE CHILDREN IN INDONESIA," JITK (Journal of Computer Science and Technology), vol. 6, no. 2, pp. 175–180, Feb. 2021, doi: 10.33480/jitk.v6i2.1935.

[16] S. Maula Chamzah, M. Lestandy, N. Kasan, and A. Nugraha, "Application of Synthetic Minority Oversampling Technique (SMOTE) to Imbalance Class on Text Data Using kNN," Syntax: Journal of Informatics, vol. 11, no. 02, pp. 56–67, Nov. 2022, doi: 10.35706/syji.v11i02.6940.

[17] W. I. Sabilla and C. Bella Vista, "Implementation of SMOTE and Under Sampling on Imbalanced Dataset for Corporate Bankruptcy Prediction," Journal of Applied Computing, vol. 7, no. 2, pp. 329–339, Dec. 2021, doi: 10.35143/jkt.v7i2.5027.

[18] A. Indrawati, H. Subagyo, A. Sihombing, W. Wagiyah, and S. Afandi, "ANALYZING THE IMPACT OF RESAMPLING METHOD FOR IMBALANCED DATA TEXT IN INDONESIAN SCIENTIFIC ARTICLES CATEGORIZATION," READ: DOCUMENTATION AND INFORMATION JOURNAL, vol. 41, no. 2, p. 133, Dec. 2020, doi: 10.14203/j.baca.v41i2.702

[19] M. Muntasir Nishat et al., "A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset," Scientific Programming, vol. 2022, pp. 1–17, Mar. 2022, doi: 10.1155/2022/3649406

[20] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in 2008 IEEE International Joint Conference on

Neural Networks (IEEE World Congress on Computational Intelligence), Jun. 2008. Accessed: Jul. 17, 2024. [Online]. Available: http://dx.doi.org/10.1109/ijcnn.2008.4633969

[21] H. Kamal and M. Mashaly, "RETRACTED: Enhancing Multi-Class Intrusion Detection through Hybrid Auto Encoder-Deep Neural Network Classifiers: A Comprehensive Analysis of Class Imbalance Mitigation Strategies Using Data Resampling Techniques," Research Square Platform LLC, May 2024. Accessed: Jul. 17, 2024. [Online]. Available: http://dx.doi.org/10.21203/rs.3.rs-4438556/v1

[22] A. P. Wibawa, M. G. A. Purnama, M. F. Akbar, and F. A. Dwiyanto, "Classification Methods," in Proceedings of the Computer Science and Information Technology Seminar, Mar. 2018, Vol. 13.

[23] R. Rusito and M. Firmansyah, "IMPLEMENTATION OF THE DECISION TREE METHOD AND C4.5 ALGORITHM FOR CLASSIFICATION OF BANK CUSTOMER DATA," Infokam Scientific Journal, vol. 12, no. 2, Oct. 2016, doi: 10.53845/infokam.v12i2.103.

[24] Baiq Nurul Azmi, Arief Hermawan, and Donny Avianto, "Analysis of the Influence of the Composition of Training Data and Testing Data on the Use of PCA and Decision Tree Algorithms for Classifying Liver Disease Patients," JTIM: Journal of Information Technology and Multimedia, vol. 4, no. 4, pp. 281–290, Feb. 2023, doi: 10.35746/jtim.v4i4.298.

[25] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[26] D. Sebastian, "Implementation of the K-Nearest Neighbor Algorithm to Classify Products from several E-marketplaces," Journal of Informatics Engineering and Information Systems, vol. 5, no. 1, May 2019, doi: 10.28932/jutisi.v5i1.1581.

[27] S. K. P. Loka and A. Marsal, "Comparison of the K-Nearest Neighbor Algorithm and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 3, no. 1, pp. 8–14, May 2023, doi: 10.57152/malcom.v3i1.474.

[28] J. Huang, J. Lu, and C. X. Ling, "Comparing naive Bayes, decision trees, and SVM with AUC and accuracy," in Third IEEE International Conference on Data Mining. Accessed: Jul. 17, 2024. [Online]. Available: http://dx.doi.org/10.1109/icdm.2003.1250975

[29] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifebri Purbolaksono, "Improving Classification Results in the Random Forest Algorithm for Detecting Diabetic Patients Using the Normalization Method," RESTI Journal (Information Systems and Technology Engineering), vol. 5, no. 1, pp. 114–122, Feb. 2021, doi: 10.29207/resti.v5i1.2880.

[30] A. Arisusanto, N. Suarna, and G. Dwilestari, "Classification Analysis of Mobile Phone Price Data Using the Random Forest Algorithm with Optimize Grid Parameters," Journal of Computer Science Technology, vol. 1, no. 2, pp. 43–47, Jul. 2023, doi: 10.56854/jtik.v1i2.51.

[31] Jan Melvin Ayu Soraya Dachi and Pardomuan Sitompul, "Comparative Analysis of the XGBoost Algorithm and the Random Forest Ensemble Learning Algorithm in Credit Decision Classification," RESEARCH JOURNAL OF MATHEMATICS AND NATURAL SCIENCES, vol. 2, no. 2, pp. 87–103, Jul. 2023, doi: 10.55606/jurrimipa.v2i2.1470.

[32] P. Handayani, Abd. C. Fauzan, and H. Harliana, "Machine Learning Classification of Toddler Nutritional Status Using the Random Forest Algorithm," CLICK: Scientific Study of Informatics and Computers, vol. 4, no. 6, pp. 3064–3072, Jun. 2024, doi: 10.30865/klik.v4i6.1909.

[33] S. Sudianto, A. D. Sripamuji, I. R. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Application of Support Vector Machine Algorithms and Multi-Layer Perceptron in News Topic Classification," National Journal of Informatics Engineering Education: JANAPATI, vol. 11, no. 2, pp. 84–91, Aug. 2022.

[34] G. A. Lustiansyah, H. Prasetyo, B. K. Widodo, B. A. Wibisono, and D. S. Prasvita, "Comparative Analysis of SVM and CNN Algorithms for Fruit Classification," Proceedings of the National Student Seminar on Computer Science and its Applications, vol. 2, no. 2, pp. 1–11, Jan. 2021.

[35] E. Ramon, A. Nazir, N. Novriyanto, Y. Yusra, and L. Oktavia, "CLASSIFICATION OF NUTRITIONAL STATUS OF POSYANDU BABIES IN BANGUN PURBA DISTRICT USING THE SUPPORT VECTOR MACHINE (SVM) ALGORITHM," Journal of Information Systems and Informatics (Simika), vol. 5, no. 2, pp. 143–150, Aug. 2022, doi: 10.47080/simika.v5i2.2185.

[36] A. Tharwat, "Classification assessment methods," Applied Computing and Informatics, vol. 17, no. 1, pp. 168–192, Jul. 2020, doi: 10.1016/j.aci.2018.08.003.

[37] M. D. N. Alif and N. F. Fahrudin, "Performance Analysis of Oversampling and Undersampling on Telco Churn Data Using Naive Bayes, SVM And Random Forest Methods," E3S Web of Conferences, vol. 484, p. 02004, 2024, doi: 10.1051/e3sconf/202448402004.

