JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 1, June 2025, pp. 390~397
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i1.4360

Optimization of Random Forest Model with SMOTE for Fetal Health Classification Based on Cardiotocography

Abdul Latif[1], Siti Khotimatul Wildah[2]
[1]Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia
[2]Teknologi Komputer, Universitas Bina Sarana Informatika, Indonesia

Article Info

Article history:

Received 05 15, 2025
Revised 06 09, 2025
Accepted 06 20, 2025

Keywords:

Cardiotocography
Fetal Health
Imbalanced Data
Random Forest
SMOTE

ABSTRACT

Fetal health during pregnancy is a crucial aspect in ensuring the optimal growth and development of a child, particularly during the golden period of life from the womb to the age of two. In the medical field, monitoring fetal conditions is vital to detect potential risks as early as possible. One of the tools commonly used in this process is cardiotocography (CTG), which provides essential data on fetal heart activity and movement. With technological advancements, machine learning-based approaches are increasingly being utilized to process CTG data more effectively. However, a major challenge in classifying medical data such as CTG lies in class imbalance, where the distribution between majority and minority classes is uneven. This study evaluates the effectiveness of the Synthetic Minority Over-sampling Technique (SMOTE) in addressing this imbalance and assesses the performance of the Random Forest algorithm in classifying fetal health conditions. The results show that the combination of SMOTE and Random Forest achieves the best performance compared to other methods, with an accuracy of 94.40%, precision of 94.45%, recall of 94.40%, and an F1-score of 94.38%. These findings indicate that SMOTE is effective in improving the representation of minority classes, while Random Forest demonstrates superior and consistent classification performance on CTG data.

Corresponding Author:

Abdul Latif
Sistem Informasi
Universitas Bina Sarana Informatika
Jakarta, Indonesia
Email: abdul.bll@bsi.ac.id

1. Introduction

The early period of life, starting from the fetal stage in the womb up to the age of two, is a crucial phase for a child's growth and development. This phase is known as the golden period, but it is also a very vulnerable stage to various negative influences [1]. Therefore, routine health examinations during pregnancy are essential to ensure the fetus remains in optimal condition.

Medical diagnosis, particularly in the field of obstetrics, plays a crucial role in determining fetal health status and detecting potential risks at an early stage. Obstetrics itself is a branch of medicine that specifically deals with everything related to pregnancy, childbirth, and the health of both the mother and the fetus [2].

Monitoring fetal health conditions, both before and during childbirth, is highly important. The risks that may be experienced by the fetus and the mother can be identified through the observation of the fetal heart rate [3]. Fetal cardiotocography is an important tool used to monitor fetal health conditions throughout pregnancy. This tool provides crucial data about the fetus, such as heart rate and movement activity [4].

In recent years, rapid advancements in signal processing technology and artificial intelligence have encouraged many researchers to explore the use of machine learning algorithms for intelligent assessment of fetal conditions [5]. The development of new strategies to support early diagnosis, screening, and risk assessment has the potential to reduce the severity of emerging disorders and minimize their negative impact on both maternal and infant health. In line with these developments, machine learning-based approaches have been utilized as a potential solution to these problems [6].

One of the techniques in machine learning that can be applied is data mining, which is a systematic process aimed at uncovering, extracting, and identifying important patterns or hidden information from large datasets using various specific methods and analytical techniques [7]. Thus, data mining enables the classification of complex medical data into categories that can be further processed and provide deeper insights into fetal health status.

Several previous studies have developed various approaches to improve accuracy in classifying fetal health conditions, utilizing machine learning technology and Fetal Cardiotocography data. One of them is a study conducted by Muhamad Rian Santoso and Purnawarman Musa, focusing on classifying fetal conditions using the C5.0 algorithm, applied to a Cardiotocography dataset. This study includes the calculation of entropy, information gain, split information, and gain ratio as part of the classification process, which is evaluated using a confusion matrix. The most optimal results were achieved in the scenario of using 90% of the data as training data and 10% as testing data, with an accuracy of 93.40% and generating 257 rules to classify fetal conditions [8].

Another study by Mehbodniya et al. expanded the understanding of machine learning algorithms that can be used in fetal health classification by testing several algorithms, such as Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron (MLP), and K-Nearest Neighbors (K-NN). The testing was performed on cardiotocography data to classify fetal conditions into three categories: normal, suspect, and pathological. The experimental results showed that the Random Forest algorithm performed the best with an accuracy of 94.5%, recall of 94%, and F1-score of 93%, while the SVM algorithm also showed high performance with an accuracy and F1-score of 93% [9].

Next, the research by Mochammad Ilham Aziz developed an ensemble method based on the Decision Tree algorithm to detect fetal health anomalies in pregnant women. This study tested the effectiveness of this method using a public dataset consisting of 2,126 patient data. The results showed that using the Decision Tree algorithm alone achieved an accuracy of 89.80%, but applying the ensemble method based on the Decision Tree improved the accuracy to 92.66%. The application of this ensemble method proved to improve the model's performance by 2.86%, which could contribute to more accurate early detection of fetal conditions [10].

Then, the study conducted by Indah Sulihati proposed a method for detecting fetal health in pregnant women by utilizing a Decision Tree classification algorithm combined with a feature forward selection technique. This study used a public dataset consisting of 2,126 patient data to test the effectiveness of this method. The experimental results showed that using the Decision Tree without feature selection achieved an accuracy of 89.84%. Meanwhile, applying the feature forward selection technique was able to increase the accuracy to 91.06%. These results show that applying feature forward selection can improve accuracy by 1.22%, which could strengthen the system's ability to detect fetal health conditions more accurately [11].

From several related studies, it is evident that choosing the right method impacts the results of the research and data processing. Therefore, selecting the appropriate algorithm or method plays a crucial role in achieving the desired outcomes. Therefore, a deep understanding of the analysis objectives and the data processing workflow is required so that the chosen algorithm can produce optimal results [12].

Before applying classification algorithms, it is important to ensure that the data has undergone adequate preprocessing. One common challenge often encountered in medical data, including cardiotocography (CTG) data, is class imbalance. Class imbalance refers to a condition where the distribution of data in each class of a dataset is uneven, or there is a significant difference between the number of samples in one class compared to the others [13]. This condition can hinder machine learning modeling because algorithms tend to have higher accuracy in predicting the majority class, while their performance declines for the minority class with fewer data samples [14].

To address this challenge, one technique that can be used is the Synthetic Minority Over-sampling Technique (SMOTE), which is specifically designed to tackle data imbalance issues. SMOTE is an

interpolation-based oversampling technique aimed at balancing the dataset by adding synthetic samples to the minority class. This method works by selecting a random sample from the minority class, then finding its nearest neighbor, and generating a new data point between them using linear interpolation—an estimation process that creates a point between two data points by drawing a straight line between them [15]. The application of SMOTE to CTG data is expected to improve classification model performance by providing fairer weighting to both classes during the learning process.

In addition to SMOTE, the Adaptive Synthetic Sampling (ADASYN) method is also a widely used alternative for addressing class imbalance problems. Unlike SMOTE, which generates synthetic data evenly along the line between minority samples and their nearest neighbors, ADASYN balances the data by adaptively generating synthetic samples for the minority class, so the model becomes more responsive to uneven data distributions [16]. Therefore, ADASYN not only balances class distribution but also enhances the model's ability to recognize the characteristics of minority classes that were previously difficult to learn.

In the context of cardiotocography data classification, the choice of classification algorithm is also an important aspect to consider. Two popular algorithms frequently used are Decision Tree (DT) and Random Forest (RF). Decision Tree is a tree-based classification method that is simple and easy to interpret [17], while Random Forest builds a number of decision trees randomly, and the classification result is determined based on the most frequent class (mode) among the predictions provided by all the trees formed in the process [18]. Each of these methods has its own advantages, and their performance may vary depending on the characteristics of the dataset used.

Based on this background, this study aims to compare the performance of two resampling techniques, namely SMOTE and ADASYN, in handling class imbalance in cardiotocography data. In addition, this study will also compare the effectiveness of two classification algorithms, namely Random Forest and Decision Tree, in classifying fetal health status. Model performance will be evaluated using relevant metrics such as accuracy, precision, recall, and F1-score to obtain a comprehensive understanding of the influence of resampling techniques and classification algorithms on prediction results. With this approach, it is expected that an optimal classification model can be obtained to detect fetal conditions based on CTG data more accurately and in a balanced manner..

2. Research Method

This study aims to evaluate the performance of classification algorithms in predicting fetal health conditions based on Cardiotocography (CTG) data. To achieve this objective, the research was conducted through several stages as illustrated in Figure 1.

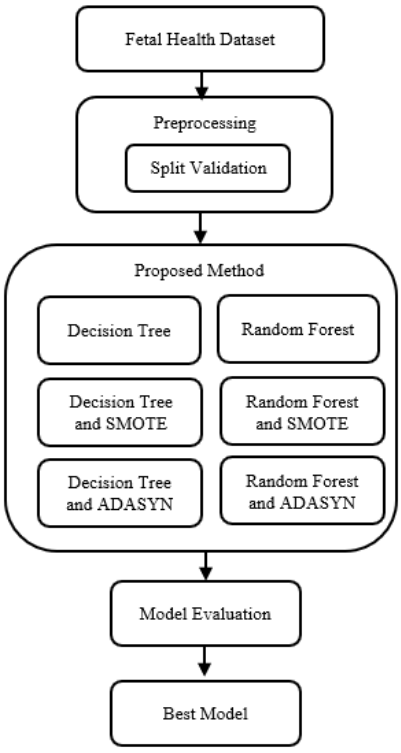

Figure 1. Research Method

2.1 Dataset

The dataset used in this study is sourced from the Kaggle platform, containing 2,126 records of cardiotocogram examination results. This dataset was developed by Ayres de Campos [19] to support efforts in reducing infant and maternal mortality rates. The features in this dataset are derived from extracted CTG signal data, which were then classified by three obstetricians into three fetal health categories: normal, suspect, and pathological. The class distribution in this dataset is imbalanced, with the "Normal" category having a significantly higher number of instances compared to "Suspect" and "Pathological." Therefore, data balancing techniques such as SMOTE and ADASYN are required to enable machine learning models to perform optimally. The attributes in the fetal health dataset along with their descriptions are as follows:

a. baseline_value: Baseline FHR (fetal heart rate) in beats per minute.
b. accelerations: Number of fetal heart rate accelerations per second.
c. fetal_movement: Number of fetal movements per second.
d. uterine_contractions: Number of uterine contractions per second.
e. light_decelerations: Number of mild decelerations in heart rate per second.
f. severe_decelerations: Number of severe heart rate decelerations per second.
g. prolongued_decelerations: Number of prolonged decelerations per second.
h. abnormal_short_term_variability: Percentage of time showing abnormal short-term variability.
i. mean_value_of_short_term_variability: Mean value of short-term variability.
j. percentage_of_time_with_abnormal_long_term_variability: Percentage of time showing abnormal long-term variability.
k. mean_value_of_long_term_variability: Mean value of long-term variability.
l. histogram_width: Width of the fetal heart rate histogram.
m. histogram_min: Minimum value in the histogram (low frequency).
n. histogram_max: Maximum value in the histogram (high frequency).
o. histogram_number_of_peaks: Number of peaks in the histogram.
p. histogram_number_of_zeroes: Number of zero values in the histogram.
q. histogram_mode: Mode value in the histogram.
r. histogram_mean: Mean value of the histogram.
s. histogram_median: Median value of the histogram.
t. histogram_variance: Variance of the histogram.
u. histogram_tendency: Tendency (positive or negative) of the histogram.
v. fetal_health: Target class: 1 (Normal), 2 (Suspect), and 3 (Pathological).

2.2 Split Validation

This stage aims to prepare the data before model training. The data is divided into two subsets: training data and test data, with a ratio of 67:33. This split is intended to allocate the majority of the data for training the model, while the remaining portion is used to test the model's performance on unseen data, thereby evaluating the model's generalization ability.

2.3 Proposed Method

This study examines six classification scenarios by utilizing two machine learning algorithms, namely Decision Tree and Random Forest, as well as two data balancing methods, namely SMOTE and ADASYN. The selection of the Decision Tree algorithm is based on its simplicity and ease of interpretation. This algorithm works by constructing a decision tree structure based on data attributes to map patterns that distinguish one class from another [17]. In general, Decision Tree is used to identify classification patterns from training data so that the model can predict the class of new data whose category is not yet known [20].

Meanwhile, the Random Forest algorithm is selected because it is a popular ensemble method widely used in both classification and regression tasks. This algorithm combines the principles of ensemble methods with decision trees [21]. Random Forest is designed to generate a number of decision trees randomly, and the classification result is determined based on the most frequently occurring class (mode) from the predictions provided by all the trees generated in the process [18].

To address the issue of class imbalance in the dataset, this study also applies two data balancing methods, namely SMOTE (Synthetic Minority Over-sampling Technique) and ADASYN (Adaptive Synthetic Sampling). SMOTE is an interpolation-based oversampling technique used to balance data by adding synthetic samples to the minority class. This method works by selecting random samples from the minority class, finding their nearest neighbors, and generating new data between them using linear

interpolation, which is the process of estimating a point between two data points by drawing a straight line between them [15].

On the other hand, ADASYN (Adaptive Synthetic Sampling) is another oversampling method designed to address the problem of data imbalance in classification processes. Its main principle is to assign varying weights to class data based on their level of difficulty. With this approach, the minority class is strengthened through the addition of synthetic samples, thereby making the distribution between classes more balanced [22]. In simple terms, ADASYN balances data by adaptively creating synthetic samples for the minority class so that the model becomes more responsive to uneven data distributions [16].

The tested scenarios include the separate use of Decision Tree and Random Forest, as well as the combination of these algorithms with the SMOTE and ADASYN data balancing methods. Each model is trained using data that has been split into training and testing sets, with balanced data used for the scenarios involving the SMOTE or ADASYN methods.

2.4 Model Evaluation

The evaluation was conducted with the aim of assessing the accuracy level achieved by the classification results, by calculating several predefined metrics. This process relied on confusion matrix analysis as the primary tool for evaluating the overall performance of the model [23]. In the confusion matrix, errors in the classification process can be seen, indicated by False Positive (FP) and False Negative (FN), which is when the model incorrectly detects the positive or negative class. Meanwhile, True Negative (TN) indicates that the model has successfully classified both positive and negative classes correctly [24]. Additional evaluations were carried out using various metrics, such as accuracy, precision, recall (or sensitivity), and F1-score [25].

2.5 Best Model

The best model is selected based on its overall performance in classifying fetal health conditions. The chosen model demonstrates the best capability in providing a balance between high accuracy and the ability to effectively distinguish among all categories. It also exhibits good stability when tested on previously unseen data, indicating optimal generalization performance. This ensures that the model is not only effective on the training data but also robust when applied to diverse.

3. Result and Discussion

The dataset was evaluated by comparing the performance of the Random Forest and Decision Tree algorithms in classifying fetal health conditions based on Cardiotocography (CTG) data, both before and after the application of data balancing techniques, namely SMOTE and ADASYN. The data was split into two subsets: training data and testing data, with a ratio of 67:33. To avoid overfitting and bias in data splitting, a 10-fold cross-validation technique was also used. The performance evaluation was carried out using accuracy, precision, recall, and F1-score metrics. Table 1 below presents the comparison of experimental results across various scenarios:

Table 1. Model Evaluation Results

Model	Accuracy	*Precision*	*Recall*	*F1-Score*
Decision Tree (without balancing)	93.59%	93.78%	93.59%	93.64%
Decision Tree and SMOTE	92.06%	92.11%	92.06%	92.01%
Decision Tree and ADASYN	87.82%	87.81%	87.82%	87.77%
Random Forest (without balancing)	93.73%	93.55%	93.73%	93.56%
Random Forest and SMOTE	94.40%	94.44%	94.40%	94.38%
Random Forest and ADASYN	93.97%	93.98%	93.97%	93.92%

From the table, it can be seen that the Random Forest algorithm performs better than the Decision Tree in all scenarios. The application of SMOTE on Random Forest results in the best performance, with an accuracy of 94.40%, followed by Random Forest with ADASYN at 93.97%. Meanwhile, the performance of the Decision Tree is consistently lower, especially when combined with ADASYN. This suggests that Random Forest is more robust in handling imbalanced datasets and benefits more significantly from resampling techniques compared to Decision Tree. Further evaluation was carried out using the confusion

matrix for the Random Forest and SMOTE model. The results of the confusion matrix testing are displayed in Figure 2.

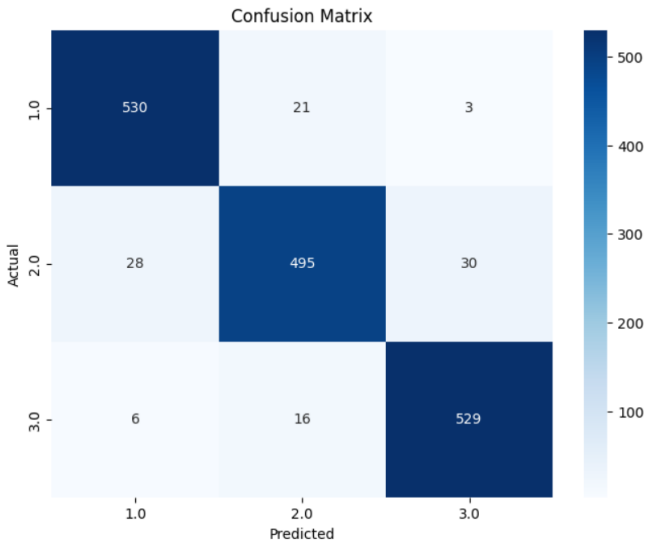

Figure 2. Confusion Matrix

Figure 2 shows the confusion matrix for the Random Forest model trained using data that has been balanced with the SMOTE method. This result illustrates the model's ability to classify data into three fetal health classes: Normal (1.0), Suspect (2.0), and Pathological (3.0).

a. For the Normal (1.0) class, 530 data points were correctly classified, while 21 data points were misclassified as Suspect and 3 data points as Pathological.
b. For the Suspect (2.0) class, the model correctly classified 495 data points. However, 28 data points were misclassified as Normal and 30 as Pathological.
c. For the Pathological (3.0) class, 529 data points were correctly classified, with 6 misclassified as Normal and 16 as Suspect.

Overall, the model shows a very high classification rate for all classes, with relatively small classification errors, reflecting the accuracy and precision of the model in predicting fetal health conditions. This shows that the application of the SMOTE (Synthetic Minority Over-sampling Technique) method has succeeded in increasing the sensitivity of the model to minority classes, namely Suspect and Pathological, which in the original dataset have a much smaller proportion of data compared to the Normal class. Without the application of SMOTE, the model often has difficulty in recognizing patterns in these minority classes due to significant data imbalance. The balanced predictions across all three classes show that the model is not biased towards the majority class and maintains stable performance in multi-class classification. Furthermore, the model's performance was also evaluated through the following classification report:

Table 2. Random Forest and SMOTE Classification Report

Class	*Precision*	*Recall*	*F1-Score*	*Support*
1.0 (Normal)	0.94	0.96	0.95	554
2.0 (*Suspect*)	0.93	0.90	0.91	553
3.0 (*Pathological*)	0.94	0.96	0.95	551
Average	**0.94**	**0.94**	**0.94**	**1658**

The classification report shows average precision and recall values of 94% across the three classes, indicating that the model performs consistently and accurately in identifying all class categories. These high scores reflect the model's ability to correctly classify positive instances (recall) while minimizing false positives (precision), even in the presence of class imbalance. Such balanced performance highlights the effectiveness of the model in handling multi-class classification tasks without bias toward the majority class,

reinforcing its robustness and reliability in practical applications. Another evaluation is presented in the form of a ROC curve, which can be seen in Figure 3.

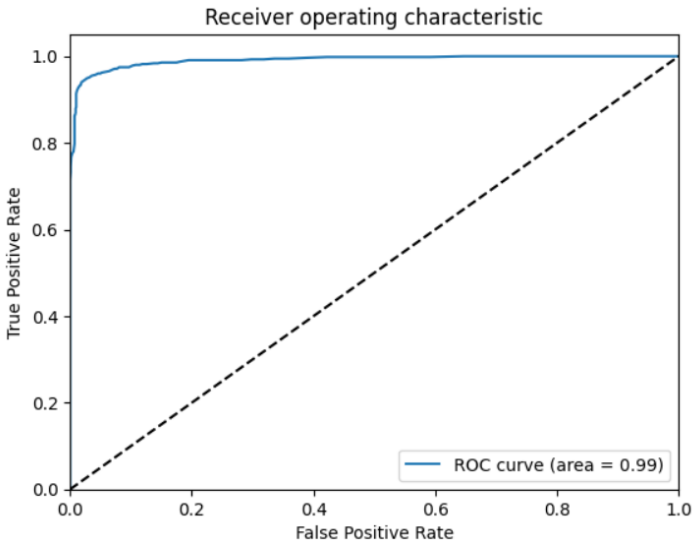

Figure 3. ROC Curve

In this chart, the ROC curve lies very close to the top-left corner, indicating excellent classification performance. The Area Under the Curve (AUC) value of 0.99 shows that the model has an almost perfect ability to distinguish between the three classes (Normal, Suspect, Pathological) in the fetal health classification dataset. An AUC value approaching 1.0 indicates that the model consistently makes correct predictions for the positive class compared to the negative class

4. Conclusion

Based on the results of this study, it can be concluded that in classifying fetal health conditions using cardiotocography (CTG) data, the combination of the Synthetic Minority Over-sampling Technique (SMOTE) and the Random Forest algorithm delivers the best performance compared to other methods. This model achieved an accuracy of 94.40%, precision of 94.45%, recall of 94.40%, and an F1-score of 94.38%. These results indicate that SMOTE is effective in addressing class imbalance by improving the representation of minority data, while Random Forest provides stable and accurate classification for medical data. Compared to the Decision Tree method and the ADASYN resampling technique, the SMOTE and Random Forest model demonstrated superior performance in accurately and evenly detecting fetal health status.

Acknowledgement (12 Pt)

Xx xxx

References (12 Pt)

[1] N. A. Tama and H. Handayani, “Determinan Status Perkembangan Bayi Usia 0 – 12 Bulan,” *J. Mhs. BK An-Nur Berbeda, Bermakna, Mulia*, vol. 7, no. 3, p. 73, 2021, doi: 10.31602/jmbkan.v7i3.5762.

[2] U. Yeni Tri, W. Linda, and S. Santi, “Analisis Ketepatan Kode Diagnosis Dan Tindakan Kasus Obstetri Pasien Rawat Inap Di Rsud Waras Wiris Boyolali,” *Infokes J. Ilm. Rekam Medis dan Inform. Kesehat.*, vol. 14, no. 1, pp. 14–21, 2024, doi: 10.47701/infokes.v14i1.3773.

[3] M. Ramla, S. Sangeetha, and S. Nickolas, “Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements,” *Proc. 2nd Int. Conf. Intell. Comput. Control Syst. ICICCS 2018*, no. June, pp. 1799–1803, 2018, doi: 10.1109/ICCONS.2018.8663047.

[4] G. A. Pradipta and Putu Desiana Wulaning Ayu, “Klasifikasi Fetal Cardiotocography Menggunakan Pendekatan Boosting Classifier,” *J. Sist. dan Inform.*, vol. 18, no. 1, pp. 103–110, 2023, doi: 10.30864/jsi.v18i1.594.

[5] J. Li and X. Liu, “Fetal Health Classification Based on Machine Learning,” *2021 IEEE 2nd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. ICBAIE 2021*, no. Icbaie, pp. 899–902, 2021, doi: 10.1109/ICBAIE52039.2021.9389902.

[6] D. Mennickent *et al.*, “Machine learning applied in maternal and fetal health: a narrative review focused on pregnancy diseases and complications,” *Front. Endocrinol. (Lausanne).*, vol. 14, no. May, pp. 1–22, 2023, doi: 10.3389/fendo.2023.1130139.

[7] S. K. Wildah, S. Agustiani, M. R. R. S, W. Gata, and H. M. Nawawi, "Deteksi Penyakit Alzheimer Menggunakan Algoritma Naïve Bayes Dan Correlation Based Feature Selection," *J. Inform.*, vol. 7, no. 2, pp. 166–173, 2020, doi: 10.31294/ji.v7i2.8226.

[8] M. R. Santoso and P. Musa, "Rekomendasi Kesehatan Janin Dengan Penerapan Algoritma C5.0 Menggunakan Classifying Cardiotocography Dataset," *J. Simantec*, vol. 9, no. 2, pp. 65–76, 2021, doi: 10.21107/simantec.v9i2.10730.

[9] A. Mehbodniya *et al.*, "Fetal health classification from cardiotocographic data using machine learning," *Expert Syst.*, vol. 39, no. 6, 2022, doi: 10.1111/exsy.12899.

[10] M. I. Aziz, "Implementasi Machine Learning Untuk Deteksi Anomali Kesehatan Janin Menggunakan Metode Ensemble Berbasis Decision Tree masa kehamilan yang selanjutnya atau kedepannya (Salwa Darin L , 2018). daripada bentuk dan suatu tren yang cenderung untuk menganalis," 2025.

[11] I. Sulihati, A. Syukur, and A. Marjuni, "Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1658–1664, 2022, doi: 10.47065/bits.v4i3.2672.

[12] A. Latif *et al.*, "Analisis Kinerja Algoritma Ensemble dalam Prediksi Perilaku Pembelian Pelanggan," vol. 9, no. 1, pp. 557–563, 2025.

[13] F. Dwi Astuti and F. Nova Lenti, "Implementasi SMOTE untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN," *J. JUPITER*, vol. 13, no. 1, pp. 89–98, 2021.

[14] H. Akbar and W. K. Sanjaya, "Kajian Performa Metode Class Weight Random Forest pada Klasifikasi Imbalance Data Kelas Curah Hujan," *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 3, no. 1, 2023, doi: 10.20885/snati.v3i1.30.

[15] A. Nurhopipah, Y. Ceasar, and A. Priadana, "Improving Machine Leaming Accuracy using Data Augmentation in Recruitment Recommendation Process," *3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021*, pp. 203–208, 2021, doi: 10.1109/EIConCIT50028.2021.9431908.

[16] M. Tiara *et al.*, "Pemanfaatan Algoritma Adasyn dan Support Vector Machine dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru," vol. 8, no. 5, pp. 8773–8778, 2024.

[17] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021, doi: 10.35329/jiik.v7i2.203.

[18] W. Aprilita, Junadhi, Agustin, and H. Asnal, "Analisis Sentimen Layanan Hotel Menggunakan Algoritma Extra Trees: Studi Kasus pada Ulasan Pelanggan," *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[19] D. Ayres-de-Campos, J. Bernardes, A. Garrido, Joaquim Marques-de-Sa, and Luis Pereira-Leite, "SisPorto 2.0: A Program for Automated Analysis of Cardiotocograms," vol. 318, no. November 1999, pp. 311–318, 2000.

[20] E. S. Pribadi, P. Poningsih, and H. S. Tambunan, "Analisa Tingkat Kepuasan Masyarakat Terhadap Pelayanan Pengadilan Agama Pematangsiantar Menggunakan Algoritma C4.5," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 1, pp. 33–40, 2020, doi: 10.30645/brahmana.v2i1.46.

[21] J. Schlenger, "Random Forest," *Comput. Sci. Sport*, pp. 201–207, 2024, doi: 10.1007/978-3-662-68313-2_24.

[22] Y. Li, W. Xu, W. Li, A. Li, and Z. Liu, "Research on hybrid intrusion detection method based on the ADASYN and ID3 algorithms," *Math. Biosci. Eng.*, vol. 19, no. 2, pp. 2030–2042, 2021, doi: 10.3934/MBE.2022095.

[23] A. Karimah, G. Dwilestari, and M. Mulyawan, "Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 767–737, 2024, doi: 10.36040/jati.v8i1.8373.

[24] W. Priatna, "Dampak Pengambilan Sampel Data untuk Optimalisasi Data tidak seimbang pada Klasifikasi Penipuan Transaksi E-Commerce," *Indones. J. Comput. Sci.*, vol. 13, no. 2, pp. 3070–3079, 2024, doi: 10.33022/ijcs.v13i2.3698.

[25] F. Aziz, P. Ishak, and S. Abasa, "Klasifikasi Depresi Menggunakan Support Vector Machine: Pendekatan Berbasis Data Text Mining," *J. Pharm. Appl. Comput. Sci.*, vol. 2, no. 2, pp. 33–38, 2024, doi: 10.59823/jopacs.v2i2.53.