Sistemasi: Jurnal Sistem Informasi
Volume 13, Nomor 3, 2024: 980-990

ISSN:2302-8149
e-ISSN:2540-9719

Comparison of XGboost, Extra Trees, and LightGBM with SMOTE for Fetal Health Classification

[1]Kartika Handayani*, [2]Badaritul Lailiah
[1,2]Sistem Informasi Kampus Kota Pontianak, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Abdul Rahman Saleh No.18, Bangka Belitung Laut, Kec. Pontianak Tenggara, Kota Pontianak, Kalimantan Barat
*e-mail: kartika.kth@bsi.ac.id

(***received***: 7 November 2023, ***revised***: 8 Maret 2024, ***accepted***: 4 April 2024)

Abstract

Cardiotocography (CTG) is widely used by obstetricians to physically access the condition of the fetus during pregnancy. This can provide data to the obstetrician about fetal heart measurements and uterine duration which helps determine whether the fetus is pathological or not. Determining the pathological classification or not can be done using machine learning methods. In this research, there is a problem of unbalanced data or data imbalance. To overcome data instability, testing using SMOTE is used. Then a comparison was made with the classifications, namely XGboost, Extra Trees and LightGBM. XGboost, Extra Trees and LightGBM testing results using SMOTE obtained the best results at 91.52% accuracy, 90.49% recall and 89.12% f1-score produced by LightGBM. Meanwhile, the best results were 89.07% precision and AUC 0.9800 produced by Extra Trees.

Keywords: *Fetal Health, SMOTE, XGboost, Extra Trees and LightGBM*

1 Introduction

The problem of medical diagnosis consists of determining the patient's status from many and various factors, such as diagnosing obstetric complications during childbirth and poor deliveries that affect the health of the fetus and the mother [1]. Cardiotocography (CTG) is the most commonly used in the clinical routine evaluation as the primary approach to detect fetal conditions [2].

There are four fundamental and crucial factors in CTG data. These four factors are the Baseline Fetal Heart Rate (BL), Accelerations (ACC), Decelerations (DCL), and Variability. By using the aforementioned criteria, doctors classify the fetal condition as normal, suspicious, or pathological [1].

Data mining has been effectively assisting medical diagnosis in recent years to prevent diagnostic errors [3]. This means that data mining can also be applied in determining a predictive model of fetal health status using CTG data by classifying fetal conditions as normal, suspicious, or pathological [4].

Data mining is expected to contribute to the efficiency of fetal health operations and the prevention of fetal deaths and diseases by accurately classifying and specifically managing fetal deaths and diseases that may occur [5]. The application of data mining in determining the classification of fetal health status encounters problems of unbalanced or imbalanced data.

Therefore, before implementing the proposed data mining classification model, a preprocessing algorithm is applied. Synthetic Minority Oversampling Technique (SMOTE) is a technique used to address unbalanced data [6]. Next, models generated from the data classification process are built using Extreme Gradient Boosting (XGBoost), Extra Trees Classifier, and Light Gradient Boosting (LightGBM).

XGBoost is an extension of gradient boosted decision trees. XGBoost is applied to solve many classification problems in various fields [7] such as store sales prediction, customer behavior prediction, ad click prediction, hazard prediction, web text prediction, and malware classification [8]. Extra Trees Classifier creates a group of decision trees that are not pruned according to the traditional top-down method. Essentially, this algorithm involves attribute randomization and simultaneous cut-point selection that splits nodes from a tree. Predictions from all trees are combined to assign the final prediction [9][10].

LightGBM is another gradient boosting algorithm that uses the leaf-wise algorithm to grow trees vertically [11]. LightGBM is a meta-algorithm in machine learning used for supervised learning.*learning* [12]. The objective of this research is to achieve the highest accuracy results from the classification algorithm employed. In this study, the data to be processed is the Fetal Health Classification dataset obtained from the Kaggle website..

2 Literature Review

Related research on the classification of fetal health status using cardiotocography data utilized seven classification models (LR, SVM, RF, DT, KNN, GNB, XGBoost) to test their efficiency in fetal health classification within the most relevant feature dataset. To assess fetal health status, the MOGACD approach is proposed as a feature selection mechanism [13].

Furthermore, the application of classifying fetal health status is explored using random forest, which is employed for various purposes. Due to the imbalanced class ratio in fetal health status, data resampling is used to replicate a small number of classes to match the count of the major classes. After data balancing, the results of feature importance analysis indicate that abnormal_short_term_variability exhibits the highest importance [5].

The application of random forest along with another method, neural network, has also been conducted using the CTG dataset from the UCI Repository. Due to high imbalance, weighting methods have been applied to optimize the model [14].

The application of classifying fetal health status also utilizes the decision tree algorithm without feature selection, resulting in an accuracy of 89.84%. Subsequently, employing feature forward selection on this decision tree algorithm yields an accuracy of 91.06% [15].

3 Research Method

The stages in conducting this research begin with dataset analysis and conclude with result evaluation. The research stages are described in the Figure.1 below.:

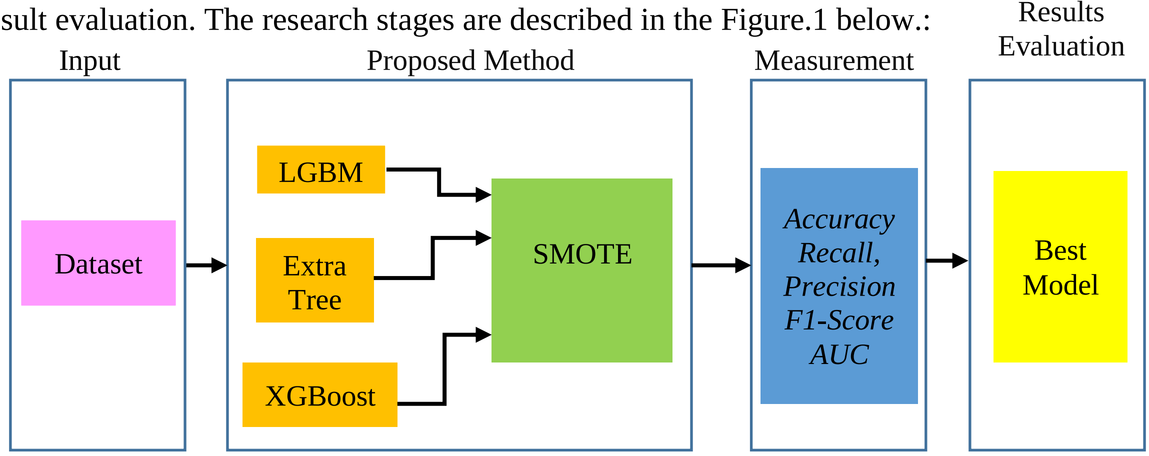


Figure 1. Research Stages

Figure 1 explains the research method from the initial stage of selecting the dataset, testing the LGBM, Extra Tree, XGBoost methods without smote and using. then the results are analyzed using several evaluation matrices such as accuracy, precision, recall, F-Score and AUC.

3.1 Dataset

This research utilizes the Fetal Health Classification dataset, which is publicly available data obtained from Kaggle. This dataset pertains to the classification of fetal health status, comprising 2126 instances, 21 attributes, and 1 class attribute.

3.2 *Pre-Processing* Data

This research employs the SMOTE algorithm to address the issue of imbalanced data. SMOTE generates synthetic samples from the minority class by interpolating existing instances that are very close to each other. For the minority category in the dataset, SMOTE randomly selects instances from the minority class data. The SMOTE algorithm is formulated as follows: Given a sample X (x1, x2, x3, ..., xn, x1, x2, x3, ... :[16]. Xn represents the n-dimensional value of sample X.

The distance from a set of samples to various classes is computed using Euclidean distance D, and the K-nearest neighbors are obtained. The formula is as follows:

$$D = \sqrt{\sum_{k-1}^{n} (X_k - Y_k)^2} \qquad (1)$$

According to the proportion of the imbalanced dataset, the sampling rate N is determined. Six samples closest to D are selected to form one group. Each group of samples is connected to each other to generate several new samples randomly, which are added to the dataset and recycled [17].

3.3 Model Testing

K-Fold = 10 is used in this research. For each iteration, the data is divided into 80% for training and the remainder for testing. Considering the imbalanced data, SMOTE is applied and then tested with XGBoost, Extra Trees Classifier, and LightGBM.

3.4 Evaluation

To compare the overall performance of the proposed research scheme, evaluation is conducted using: accuracy, recall, precision, F1-Score, and AUC. Confusion matrices are also presented to describe the performance of each classification model, and feature importance analysis is conducted to identify the attributes that most influence the classification model results.

4 Results and Analysis

4.1 Dataset

The dataset used is a public dataset taken from Kaggle. This data is a classification dataset of fetal health status. It consists of 2126 data points, 21 attributes, and 1 class attribute. Below is the dataset table of Fetal Health Classification explaining the attributes in the dataset:

Table 1. Dataset Attributes

No	Name	Description
1	baseline value	Fetal Heart Rate (FHR)
2	Accelerations	Number of accelerations per second
3	fetal_movement	Number of fetal movements per second
4	uterine_contractions	Number of uterine contractions per second
5	light_decelerations	Number of LD per second
6	severe_decelerations	Number of SD per second
7	prolongued_decelerations	Number of PD per second
8	abnormal_short_term_variability	Percentage of time with abnormal short-term variability Average value of short-term variability
9	mean_value_of_short_term_variability	Average value of short-term variability
10	percentage_of_time_with_abnormal_long_term_variability	Percentage of time with abnormal long-term variability
11	mean_value_of_long_term_variability	Average value of long-term variability
12	histogram_width	Width of the histogram constructed using all values from the record a
13	histogram_min	Minimum histogram value

14	histogram_max	Maximum histogram value
15	histogram_number_of_peaks	Number of peaks in the test histogram
16	histogram_number_of_zeroes	Number of zeros in the test histogram
17	histogram_mode	Hist mode
18	histogram_mean	Hist mean
19	histogram_median	Hist median
20	histogram_variance	Hist variance
21	histogram_tendency	Histogram trend
22	fetal_health	Fetal health: 1 - Normal 2 - suspect 3 – Pathological

Table 1 explains the 22 attributes in the data set used by the research. The use of these attributes is based on the results of analysis from previous research.

4.2 *Pre-Processing* Data

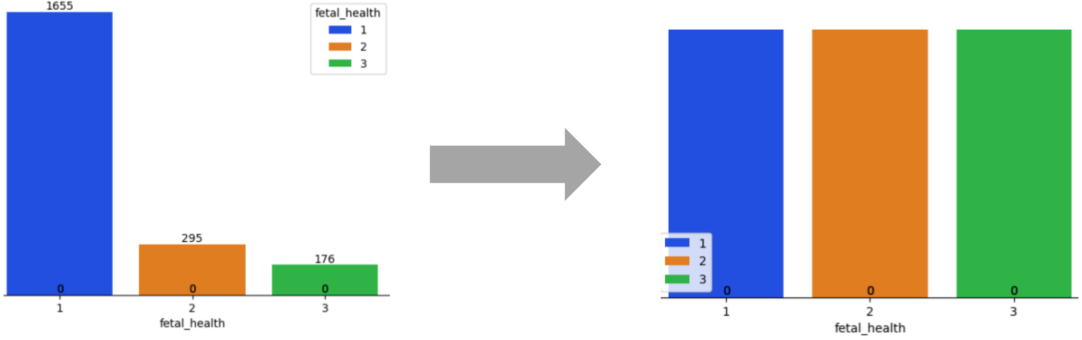


Figure 2. SMOTE Process

Figure 2 illustrates the initial data with 3 imbalanced classes, with each class having 1 (Normal) totaling 1655, class 2 (Suspect) totaling 295, and class 3 (Pathological) totaling 176. Subsequently, data preprocessing is performed using SMOTE to balance the training data of these three classes, making them equal in quantity.

4.3 Model Testing

In the testing, evaluations were conducted both without SMOTE and with SMOTE. Here are the results of the testing without using SMOTE :

Table 2. Results of Testing Without SMOTE

Method	Accuracy	Recall	Precision	F1-Score	AUC
XGBoost	91,07%	81,80%	86,76%	82,85%	0.9764
Extra Trees	89,44%	70,81%	83,90%	75,30%	0.9753
LightGBM	90,84%	79,03%	84,60%	79,70%	0.9761

Table 2 describes the results after testing without SMOTE, XGBoost obtained an accuracy of 91.07%, recall of 81.80%, precision of 86.76%, F1-Score of 82.85%, and an AUC of 0.9764. Extra Trees achieved an accuracy of 89.44%, recall of 70.81%, precision of 83.90%, F1-Score of 75.30%, and an AUC of 0.9753. Meanwhile, LightGBM obtained an accuracy of 90.84%, recall of 79.03%, precision of 84.60%, F1-Score of 79.70%, and an AUC of 0.9761. After testing without SMOTE, testing with SMOTE was conducted as shown in the following table :

Tab1e 3. Results of Testing With SMOTE

Method	Accuracy	Recall	Precision	F1-Score	AUC
XGBoost+SMOTE	91,30%	89,15%	88.32%	88,32%	0.9775
Extra Trees+SMOTE	91,30%	89,40%	89,07%	88,84%	0.9800
LightGBM+SMOTE	91,52%	90,49%	88,34%	89,12%	0.9770

Table 3 describes the results after testing with SMOTE. For clarity, a comparison table is presented based on the evaluation matrices used. Below is the table comparing the results of the three algorithms before and after applying SMOTE in tables 4, 5, 5, 7, and 8:

Table 4. Accuracy Comparison

Method	*XGBoost*	*Extra Trees*	*LightGBM*
Tanpa SMOTE	91,07%	89,44%	90,84%
Dengan SMOTE	91,30%	91,30%	91,52%

Table 4 describes the testing using SMOTE, XGBoost achieved an accuracy of 91.30%, resulting in an increase of 0.23%. The testing results using SMOTE with Extra Trees yielded an accuracy of 91.30%, indicating an increase of 1.86%. Meanwhile, the testing without SMOTE using LightGBM resulted in an accuracy of 91.52%, showing an increase of 0.68%.

Table 5. Recall Comparison

Method	*XGBoost*	*Extra Trees*	*LightGBM*
Tanpa SMOTE	81,80%	70,81%	79,03%
Dengan SMOTE	89,15%	89,40%	90,49%

Table 5 describes the testing using SMOTE, XGBoost achieved a recall of 89.15%, resulting in an increase of 7.35%. In the testing using Extra Trees, a recall of 89.40% was obtained, indicating an increase of 18.59%. Meanwhile, in the testing using SMOTE with LightGBM, a recall of 90.49% was achieved, showing an increase of 11.48%.

Table 6. Precision Comparison

Method	*XGBoost*	*Extra Trees*	*LightGBM*
Tanpa SMOTE	86,76%	83,90%	84,60%
Dengan SMOTE	88.32%	89,07%	88,34%

Table 6 describes the testing using SMOTE, XGBoost achieved a precision of 88.32%, resulting in an increase of 1.56%. In the testing using SMOTE with Extra Trees, a precision of 89.07% was obtained, indicating an increase of 5.17%. Meanwhile, in the testing using SMOTE with LightGBM, a precision of 88.34% was achieved, showing an increase of 3.74%.

Table 7. F1-Score Comparison

Method	*XGBoost*	*Extra Trees*	*LightGBM*
Tanpa SMOTE	82,85%	75,30%	79,70%
Dengan SMOTE	88,32%	88,84%	89,12%

Table 7 describes the testing using SMOTE, XGBoost achieved an F1-Score of 88.32%, resulting in an increase of 5.47%. In the testing using SMOTE with Extra Trees, an F1-Score of 88.84% was obtained, indicating an increase of 13.54%. Meanwhile, in the testing using SMOTE with LightGBM, an F1-Score of 89.12% was achieved, showing an increase of 9.42%.

Table 8. AUC Comparison

Method	*XGBoost*	*Extra Trees*	*LightGBM*
Tanpa SMOTE	0.9764	0.9753	0.9761
Dengan SMOTE	0.9775	0.9800	0.9770

Table 8 describes the testing using SMOTE, XGBoost obtained an AUC of 0.9775, resulting in an increase of 0.0011. In the testing using SMOTE with Extra Trees, an AUC of 0.9800 was obtained, indicating an increase of 0.0047. Meanwhile, in the testing using SMOTE with LightGBM, an AUC of 0.9770 was achieved, showing an increase of 0.0009.

3.3 Evaluation

To visualize the results of the testing, ROC curves, confusion matrices, precision-recall curves, and feature importance should be displayed to understand the attributes that most influence the classification model results.

3.1 XGBoost Evaluation

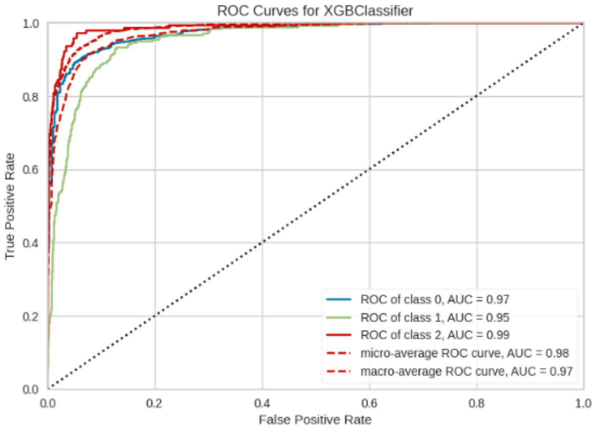


Figure 3. XGBoost ROC Curve

Figure 3 describes the XGBoost ROC curve, the ROC result for the normal class is 0.97, for the suspect class is 0.95, and for the pathological class is 0.99. Meanwhile, the average micro ROC for all three classes is 0.98 and the average macro ROC is 0.97.

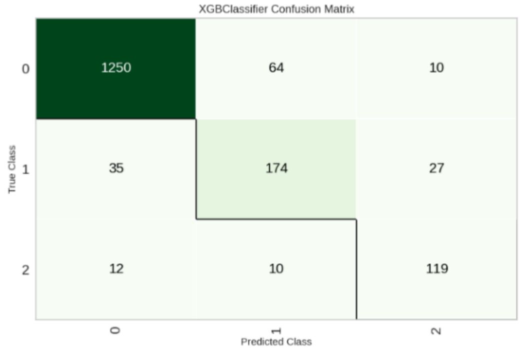


Figure 4. XGBoost Confusion Matrix

In the above figure 4, based on the evaluation results using the confusion matrix for predicting the fetal health status, it can be concluded that out of 1250 instances predicted as 'Normal', the classification correctly predicted the fetal health status. However, 64 instances predicted as 'Normal' actually turned out to be 'Suspect', and 10 instances predicted as 'Normal' actually turned out to be 'Pathological'. Regarding the prediction of 'Suspect', out of 174 instances, the classification showed consistency between the prediction and the actual result. However, 35 instances predicted as 'Suspect' actually turned out to be 'Normal', and 27 instances predicted as 'Suspect' actually turned out to be 'Pathological'. Lastly, in the prediction of 'Pathological', out of 119 instances, the classification showed consistency between the prediction and the actual result. However, 12 instances predicted as 'Pathological' actually turned out to be 'Normal', and 10 instances predicted as 'Pathological' actually turned out to be 'Suspect'.

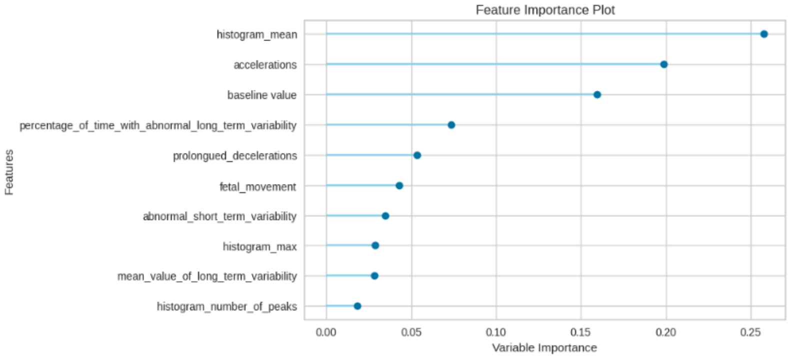


Figure 5. XGBoost *Feature Important*

Based on Figure 5, the attributes that have the most influence on classifying the fetal health status using XGBoost are histogram_mean, accelerations, baseline value, percentage_of_time_with_abnormal_long_term_variability, prolonged_decelerations, fetal_movement, abnormal_short_term_variability, histogram_max, mean_value_of_long_term_variability, and histogram_number_of_peaks.

3.2 Extra Trees Evaluation

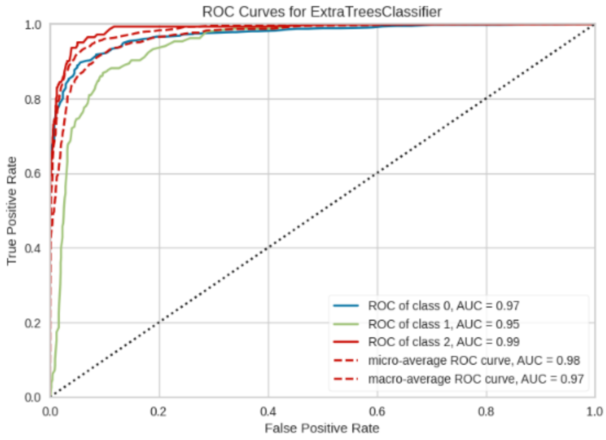


Figure 6. Extra Trees ROC Curve

Figure 6 describes the Extra Trees ROC curve, the ROC result for the normal class is 0.97, for the suspect class is 0.95, and for the pathological class is 0.99. Meanwhile, the average micro ROC for all three classes is 0.98 and the average macro ROC is 0.97.

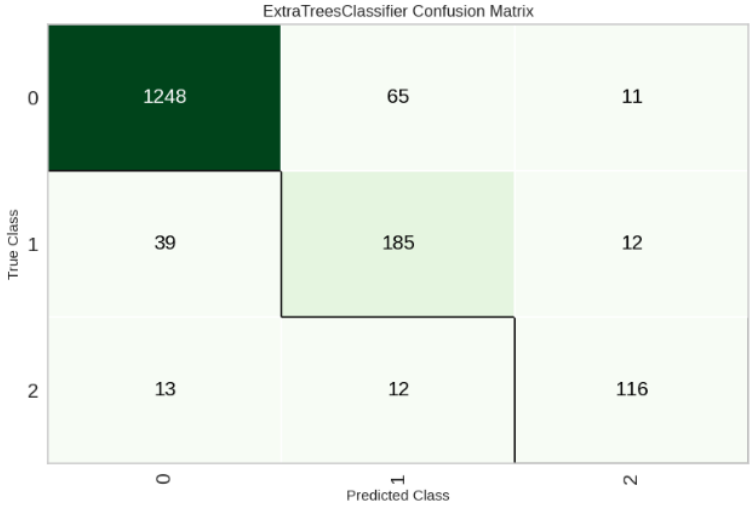


Figure 7. Extra Trees Confusion Matrix

In the above Figure 7, based on the evaluation results using the confusion matrix for predicting the fetal health status, it can be concluded that out of 1248 instances predicted as 'Normal', the classification correctly predicted the fetal health status. However, 65 instances predicted as 'Normal' actually turned out to be 'Suspect', and 11 instances predicted as 'Normal' actually turned out to be 'Pathological'. Regarding the prediction of 'Suspect', out of 185 instances, the classification showed consistency between the prediction and the actual result. However, 39 instances predicted as 'Suspect' actually turned out to be 'Normal', and 12 instances predicted as 'Suspect' actually turned out to be 'Pathological'. Lastly, in the prediction of 'Pathological', out of 116 instances, the classification showed consistency between the prediction and the actual result. However, 13 instances predicted as 'Pathological' actually turned out to be 'Normal', and 12 instances predicted as 'Pathological' actually turned out to be 'Suspect'.

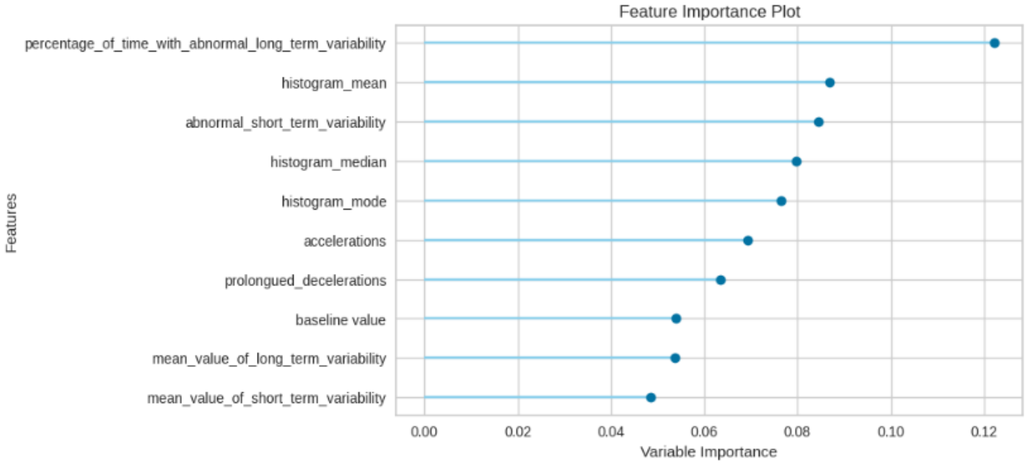


Figure 8. Extra Trees *Feature Important*

Based on the above Figure 8, the attributes that have the most influence on classifying the fetal health status using Extra Trees are percentage_of_time_with_abnormal_long_term_variability, histogram_mean, abnormal_short_term_variability, histogram_median, histogram_mode, accelerations, prolonged_decelerations, baseline value, mean_value_of_long_term_variability, and mean_value_of_short_term_variability.

3.3 LightGBM Evaluation

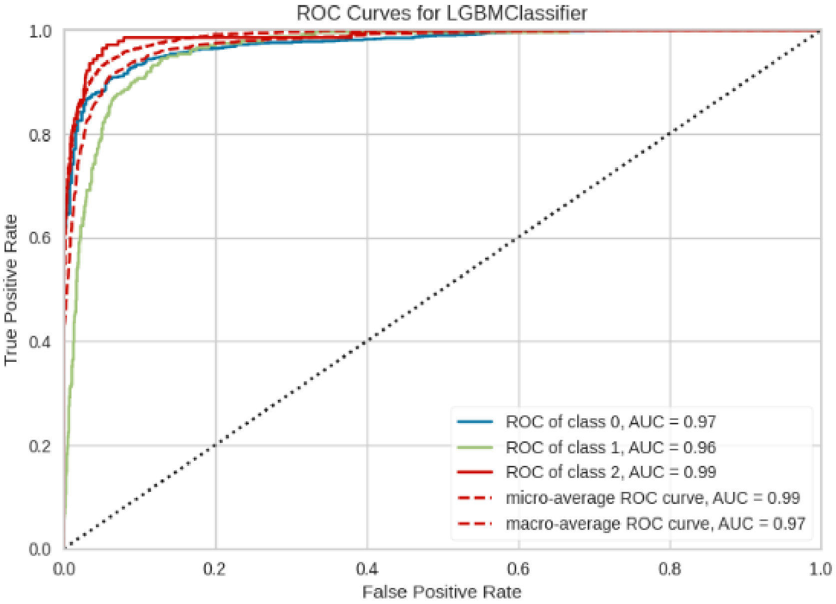
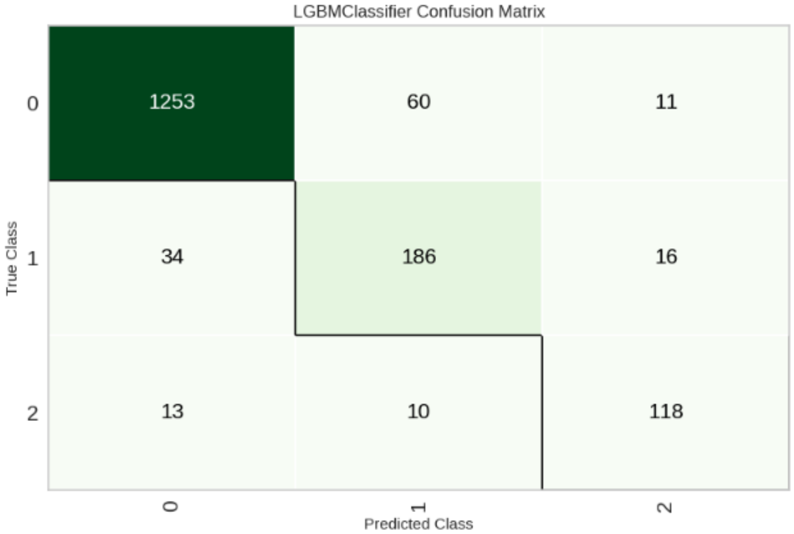


Image 9. LightGBM ROC Curve

Figure 9 describes the LightGBM ROC curve, the ROC result for the normal class is 0.97, for the suspect class is 0.96, and for the pathological class is 0.99. Meanwhile, the average micro ROC for all three classes is 0.99 and the average macro ROC is 0.97.



True Class \ Predicted Class	0	1	2
0	1253	60	11
1	34	186	16
2	13	10	118

Figure 10. LightGBM Confusion Matrix

In the above Figure 10, based on the evaluation results using the confusion matrix for predicting the fetal health status, it can be concluded that out of 1253 instances predicted as 'Normal', the classification correctly predicted the fetal health status. However, 60 instances predicted as 'Normal' actually turned out to be 'Suspect', and 11 instances predicted as 'Normal' actually turned out to be 'Pathological'. Regarding the prediction of 'Suspect', out of 186 instances, the classification showed consistency between the prediction and the actual result. However, 34 instances predicted as 'Suspect' actually turned out to be 'Normal', and 16 instances predicted as 'Suspect' actually turned out to be 'Pathological'. Lastly, in the prediction of 'Pathological', out of 118 instances, the classification showed consistency between the prediction and the actual result. However, 13 instances predicted as 'Pathological' actually turned out to be 'Normal', and 10 instances predicted as 'Pathological' actually turned out to be 'Suspect'.

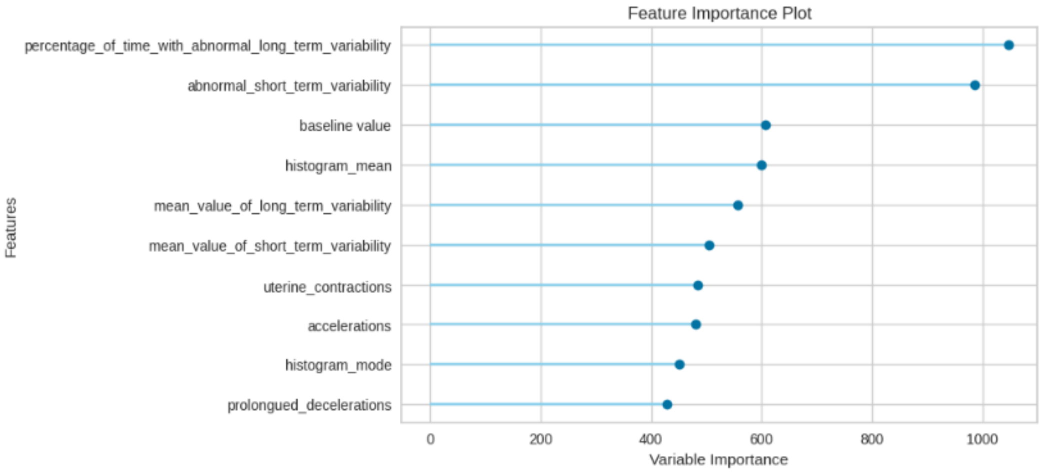


Figure 11. LightGBM *Feature Important*

Based on the above Figure11, the attributes that have the most influence on classifying the fetal health status using LightGBM are percentage_of_time_with_abnormal_long_term_variability, abnormal_short_term_variability, baseline value, histogram_mean, mean_value_of_long_term_variability, mean_value_of_short_term_variability, uterine_contractions, accelerations, histogram_mode, and prolonged_decelerations.

5 Conclusion

In this study, we compare XGBoost, Extra Trees, and LightGBM in classifying fetal health status. To address the issue of imbalanced data, SMOTE is used. This study compares the results before and after using SMOTE. From the test results, the use of SMOTE in classifying fetal health status can improve the results. The dataset used in this study is sourced from the Kaggle repository, which includes classes of normal, suspect, and pathological. Testing XGBoost, Extra Trees, and LightGBM using SMOTE yielded the best results with accuracy of 91.52%, recall of 90.49%, and F1-score of 89.12% produced by LightGBM. Meanwhile, the best precision of 89.07% and AUC of 0.9800 were obtained by Extra Trees.

Reference

[1] 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. Iciccs, pp. 1799–1803, 2019, doi: 10.1109/ICCONS.2018.8663047.

[2] 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.

[3] A. Batra, A. Chandra, and V. Matoria, “Cardiotocography Analysis using Conjunction of Machine Learning Algorithms,” *Proc. - 2017 Int. Conf. Mach. Vis. Inf. Technol. C. 2017*, pp. 1–6, 2017, doi: 10.1109/CMVIT.2017.27.

[4] S. Das, H. Mukherjee, S. M. Obaidullah, K. Roy, and C. K. Saha, “Ensemble based Technique for the Assessment of Fetal Health using Cardiotocograph – a Case Study with Standard Feature Reduction Techniques,” *Multimed. Tools Appl.*, vol. 79, no. 47–48, pp. 35147–35168, 2020, doi: 10.1007/s11042-020-08853-2.

[5] J. R. Chun, “Design of Fetal Health Classification Model for Hospital Operation Management,” *J. Digit. Converg.*, vol. 19, no. 5, pp. 263–268, 2021, [Online]. Available: https://www.koreascience.or.kr/article/JAKO202116739401225.page%0Ahttps://www.koreascience.or.kr/article/JAKO202116739401225.pdf.

[6] A. Fernández, S. García, F. Herrera, and N. V. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” *J. Artif. Intell.*

Res., vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.

[7] I. L. Cherif and A. Kortebi, "On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification," *IFIP Wirel. Days*, vol. 2019-April, pp. 1–6, 2019, doi: 10.1109/WD.2019.8734193.

[8] S. F. N. Islam, A. Sholahuddin, and A. S. Abdullah, "Extreme Gradient Boosting (XGBoost) Method in Making Forecasting Application and Analysis of USD Exchange Rates Against Rupiah," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012016.

[9] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of Tree-based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement," *Inf.*, vol. 11, no. 6, 2020, doi: 10.3390/info11060332.

[10] F. Ernawan, K. Handayani, M. Fakhreldin, and Y. Abbker, "Light Gradient Boosting with Hyper Parameter Tuning Optimization for COVID-19 Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, pp. 514–523, 2022, doi: 10.14569/IJACSA.2022.0130859.

[11] F. Alzamzami, M. Hoda, and A. El Saddik, "Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation," *IEEE Access*, vol. 8, pp. 101840–101858, 2020, doi: 10.1109/ACCESS.2020.2997330.

[12] A. Saiful, "Deteksi Mobil Dari Citra CCTV menggunakan Fuzzy Morphology dan Gradient Boosting," 2016.

[13] J. Piri, P. Mohapatra, and R. Dey, "Fetal Health Status Classification using MOGA - CD Based Feature Selection Approach," *Proc. CONECCT 2020 - 6th IEEE Int. Conf. Electron. Comput. Commun. Technol.*, 2020, doi: 10.1109/CONECCT50063.2020.9198377.

[14] G. Xue, "The Application of Machine Learning Models in Fetal State Auto-Classification Based on Cardiotocograms," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 310, no. 5, 2019, doi: 10.1088/1755-1315/310/5/052007.

[15] 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.

[16] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, "Data Sampling Methods to Dealwith the Big Data Multi-Class Imbalance Problem," *Appl. Sci.*, vol. 10, no. 4, 2020, doi: 10.3390/app10041276.

[17] X. Li and Q. Zhou, "Research on Improving SMOTE Algorithms for Unbalanced Data Set Classification," *Proc. - 2019 Int. Conf. Electron. Eng. Informatics, EEI 2019*, pp. 476–480, 2019, doi: 10.1109/EEI48997.2019.00109.