Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 1, No. 2, 2023

Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach

Teuku Rizky Noviandy [1], Sarah Ika Nainggolan [2], Raihan Raihan [3], Isra Firmansyah [3] and Rinaldi Idroes [4]

1 Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
2 Department of Obstetrics and Gynecology, Faculty of Medicine, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; ikasarah@yahoo.com (S.I.N.)
3 Department of Child Health, dr. Zainoel Abidin General Hospital, Banda Aceh 24415, Indonesia; raihan_rais@yahoo.com (R.R.); israfirmansyahhasny@gmail.com (I.F.)
4 School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 12 October 2023
Revised 21 November 2023
Accepted 29 November 2023
Available Online 2 December 2023

Keywords:
Health informatics
Classification
LightGBM

Abstract

Maternal health risk detection is crucial for reducing morbidity and mortality among pregnant women. In this study, we employed the Light Gradient Boosting Machine (LightGBM) model to identify risk levels using data from rural healthcare facilities. The dataset included key health indicators aligned with the United Nations Sustainable Development Goals. The LightGBM model underwent rigorous optimization through hyperparameter tuning and 10-fold cross-validation. Its predictive performance was benchmarked against other algorithms using accuracy, precision, recall, and F1-score, with feature importance assessed to identify critical risk predictors. The LightGBM model demonstrating the highest performance across all metrics. The results underscore the value of advanced machine learning techniques in public health. Future research directions include expanding the demographic scope, incorporating temporal data, and enhancing model transparency. This study highlights the transformative potential of machine learning in maternal healthcare, providing a foundation for improved risk detection and proactive healthcare interventions.

1. Introduction

Maternal health, which focuses on the health of women during pregnancy, childbirth, and after giving birth, is an essential part of public health [1]. Despite advancements in healthcare, maternal mortality and morbidity rates persist, underscoring the importance of early detection and effective management of maternal health risks [2, 3]. Identifying maternal health risks is a complex task, as numerous factors contribute to complications during pregnancy and childbirth, making it imperative to explore innovative approaches to enhance risk detection and intervention.

Traditionally, the detection of maternal health risks has relied on manual assessments and periodic check-ups, often resulting in delayed identification and response to potential complications [4, 5]. This conventional approach is inherently limited, as it heavily depends on the subjective judgment of healthcare professionals and may not adequately capture the nuances of evolving risk factors [6]. The consequences of delayed or inaccurate risk detection can be severe, leading to adverse outcomes for both mothers and infants [7]. Hence, there is a pressing need for more sophisticated and data-driven methodologies to improve the accuracy and timeliness of maternal health risk detection.

In recent years, the field of machine learning has witnessed significant advancements, offering promising avenues for enhancing healthcare practices [8–10]. Machine learning algorithms can analyze vast datasets and identify complex patterns that may elude traditional methods [11–14]. Leveraging these capabilities, machine learning models have the potential to transform the landscape of maternal health risk detection by providing more accurate and timely predictions [15]. This study explores the application of machine learning, specifically focusing on the Light Gradient Boosting Machine (LightGBM) approach, as a viable solution to address the limitations of traditional risk detection methods in the context of maternal health.

LightGBM, a powerful gradient boosting framework, has gained popularity for its efficiency and effectiveness in handling large datasets with high dimensionality [16]. Its ability to optimize training speed and model performance makes it well-suited for healthcare applications, where the need for real-time predictions and interpretability is paramount.

The aims of this study are to evaluate the feasibility and effectiveness of utilizing the LightGBM algorithm for maternal health risk detection. By harnessing the power of machine learning, this research seeks to contribute to the advancement of proactive and personalized healthcare interventions, ultimately reducing maternal mortality and morbidity rates. The outcomes of this study have the potential to inform future developments in the application of machine learning in maternal healthcare, fostering a paradigm shift towards data-driven and precision medicine approaches.

2. Materials and Methods

2.1. Dataset

The datasets utilized in this study were obtained from a comprehensive study conducted by Ahmed et al. [17]. The data collection process involved collaboration with various healthcare institutions, including hospitals, community clinics, and maternal health care facilities located in rural areas of Bangladesh. The data acquisition was facilitated through an Internet of Things (IoT) based risk monitoring system, ensuring real-time and continuous monitoring of maternal health parameters.

The dataset encompasses a range of vital health indicators identified as crucial contributors to maternal mortality, aligning with the priorities outlined in the Sustainable Development Goals (SDGs) set by the United Nations [18]. The target variable in this dataset is maternal risk level, categorized into three levels: low risk (406 samples), medium risk (336 samples), and high risk (272 samples). The features in this dataset include age, systolic blood pressure, diastolic blood pressure, blood sugar, body temperature and heart rate. Each of these variables plays a significant role in assessing maternal health risks, and their inclusion in the study reflects a holistic approach to risk detection. The explanation of features used in this study are presented in Table 1.

2.2. LightGBM Model

The LightGBM is a sophisticated yet efficient tool in the field of machine learning. It has proven its effectiveness across a wide range of machine learning tasks [19–21], making it a robust choice for processing complex datasets, such as those encountered in healthcare research. Its design allows for rapid processing of data, a vital attribute in medical applications where prompt and accurate results are crucial. LightGBM utilizes advanced techniques such as gradient-based one-side sampling and exclusive feature bundling [16]. These techniques enhance its ability to swiftly and accurately navigate through extensive data, identifying pertinent patterns and relationships. This capability is particularly beneficial in healthcare, where data is often intricate and multifaceted.

To optimize the performance of the LightGBM model in this study, hyperparameter tuning was employed. Five key hyperparameters were adjusted, and the details of their tuning ranges were presented in Table 2. The training process incorporated 10-fold cross-validation, a method that splits the dataset into ten parts, using nine for training and one for testing, and iterates this process ten times. This approach ensures a thorough evaluation of the model's performance across different segments of the data [22, 23]. Additionally, the training involved 100 iterations, allowing for extensive exploration of the hyperparameter space to find the optimal configuration for the LightGBM model. This meticulous training and tuning process aims to enhance the model's ability to accurately predict maternal health risk levels, thereby contributing to more effective risk detection and management in maternal healthcare.

2.3. Model Evaluation

The effectiveness of the LightGBM model in this study was rigorously evaluated using four essential metrics: accuracy, precision, recall, and the F1-score. These metrics were calculated using a weighted average approach, which is particularly suitable for multiclass

Table 1. Features in the dataset.

Feature Name	Description	Units
Age	Any ages in years during pregnant.	years
SystolicBP	Upper value of blood pressure	mmHg
DiastolicBP	Lower value of blood pressure	mmHg
BloodSugar	Blood glucose levels	mmol/L
BodyTemp	Body Temperature	F
HeartRate	A normal resting heart rate	bpm

Table 2. Tuning ranges of key hyperparameters for the LightGBM model.

Hyperparameter	Tuning Range
n_estimators	Random integer from 50 to 500
max_depth	Random integer from 3 to 15
learning_rate	Uniformly spaced values from 0.01 to 0.3 (100 values)
subsample	Uniformly spaced values from 0.2 to 1.0 (100 values)
colsample_bytree	Uniformly spaced values from 0.2 to 1.0 (100 values)

problems, ensuring a balanced evaluation across different classes [24, 25].

In addition, the LightGBM model's performance was compared with that of four other established machine learning models: Random Forest, AdaBoost, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). This comparison is pivotal to understand LightGBM's relative effectiveness and to determine the most suitable approach for maternal health risk detection. Each of these models underwent the same evaluative metrics to maintain consistency and fairness in the comparison, thereby providing a comprehensive view of the most effective machine learning strategy for this application.

3. Results and Discussion

3.1. Exploratory Data Analysis

We explored the distributions of various health parameters as depicted in Figure 1. The age distribution was slightly skewed to the right, indicating a predominantly younger population with fewer older individuals. Systolic blood pressure showed a pronounced peak around the normotensive range, with extended tails suggesting variations towards hypertensive and hypotensive ranges. Diastolic blood pressure displayed a multimodal distribution, potentially indicating subpopulations within the dataset or sample size effects. Blood sugar levels were predominantly on the lower end, with a rapid decline towards the higher levels, highlighting fewer instances of high blood sugar values. Body temperature was concentrated around the normal range, with a steep peak suggesting a largely healthy sample in terms of temperature. Heart Rate showed a multimodal pattern, suggesting a mix of resting and active states in the population at the time of measurement. These patterns underscored the general healthiness of the sample population while also highlighting areas with significant variability that warranted further investigation.

The correlation matrix, depicting the pairwise correlation coefficients between each variable in the dataset, was presented in Figure 2. The strongest positive correlation was between Diastolic and Systolic Blood Pressure (0.79), as expected, given that these measurements often increased in tandem. Age showed a notable positive correlation with both Systolic (0.42) and Diastolic Blood Pressure (0.4), as well as Blood Sugar (BS) (0.47), suggesting that higher values in these parameters were associated with older age groups. Interestingly, Body Temperature was negatively correlated with Age (-0.26), SystolicBP (-0.29), DiastolicBP (-0.26), and BS (-0.1), indicating that higher temperatures were less common in individuals with higher age or higher blood pressure and sugar levels. Heart Rate did not show a strong correlation with any other parameter, with the highest being a weak positive correlation with Body Temperature (0.14). Risk Level was positively correlated with all parameters except for Body Temperature and Heart Rate, with the strongest correlation being with BS (0.57), suggesting that higher blood sugar levels had a significant impact on the overall risk profile. These correlations provided insights into how these variables might have interacted and contributed to the overall health risk assessment.

To observe the distribution of risk level, we illustrated the two-dimensional feature space resulting from a Principal Component Analysis (PCA) applied to the features in Figure 3. The scatter plot was color-coded to represent three categories of risk: high risk (dark blue), mid risk (light blue), and low risk (grey). The distribution showed that the high-risk individuals were somewhat clustered, predominantly appearing on the positive side of PC-1, though there was no clear boundary separating the three risk levels. This overlap indicated that while PCA had

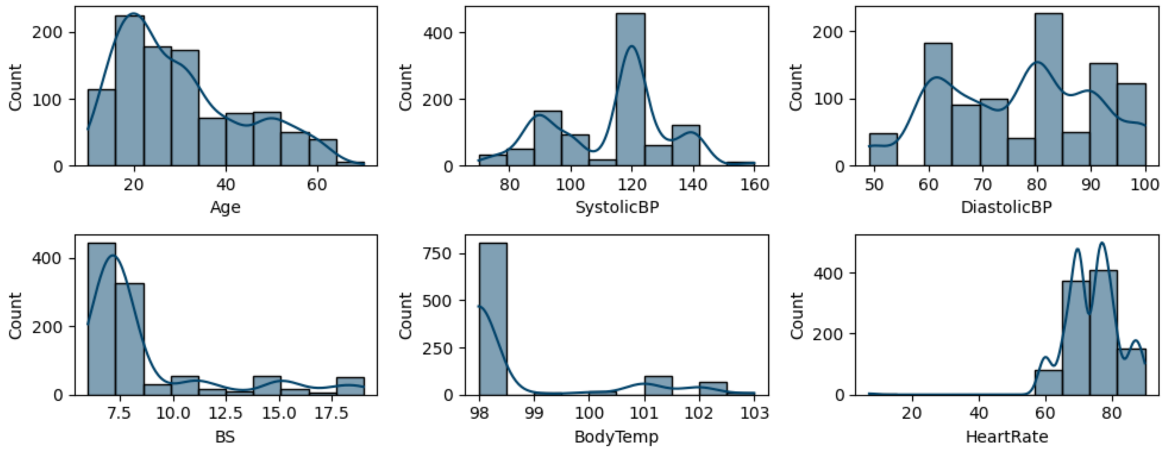

Figure 1. Histograms overlaid with kernel density estimates for each feature.

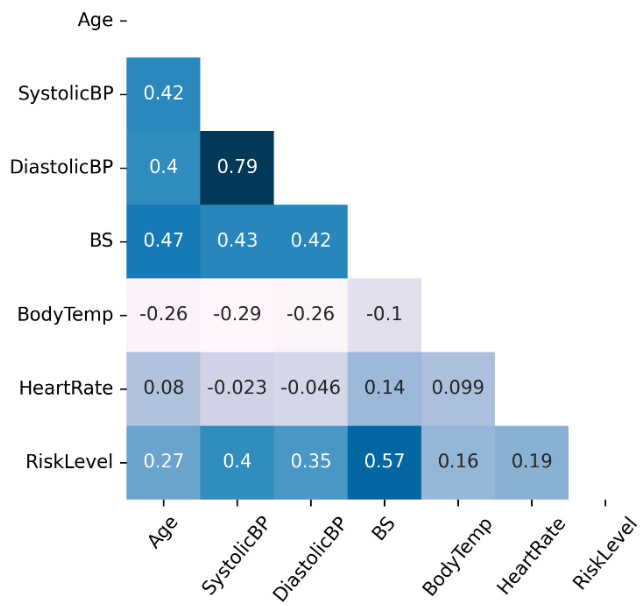

Figure 2. Correlation matrix of the dataset with the pairwise correlation coefficients.

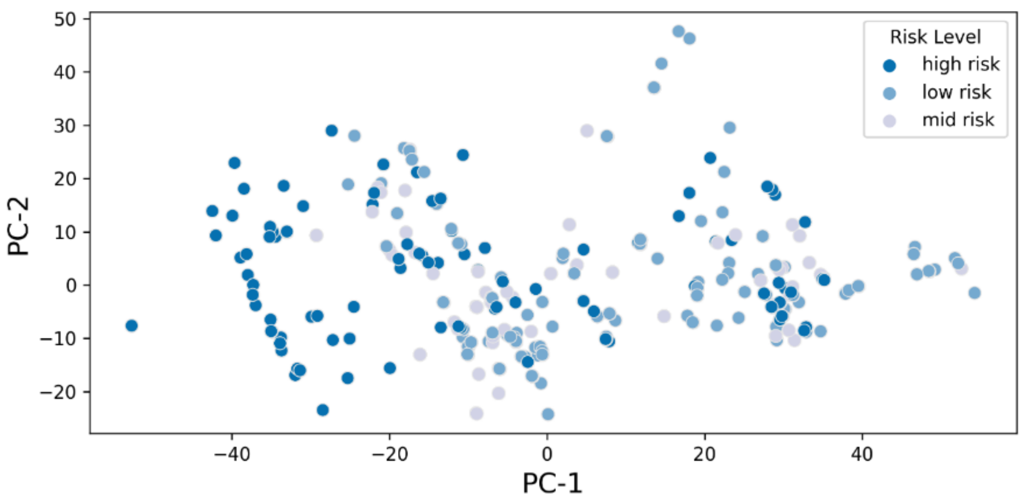

Figure 3. PCA scatter plot illustrating the spread of high, mid, and low risk levels across the first two principal components.

Table 3. Performance of the model on the testing set.

Model	Accuracy	Precision	Recall	F1-Score
LightGBM	**84.73**	**84.75**	**84.73**	**84.73**
Random Forest	81.28	81.85	81.28	81.34
AdaBoost	69.46	70.36	69.46	69.64
K-NN	66.50	67.09	66.50	66.39
SVM	59.61	64.54	59.61	56.31

Note: Bold values indicate best results

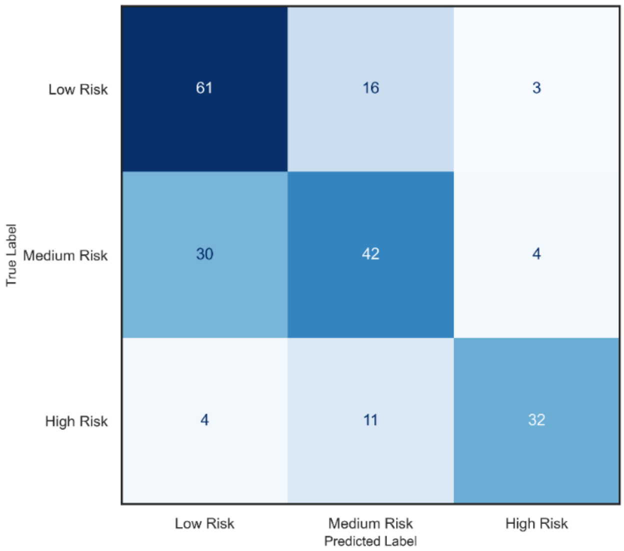

Figure 4. Confusion matrix of the LightGBM model on the testing set.

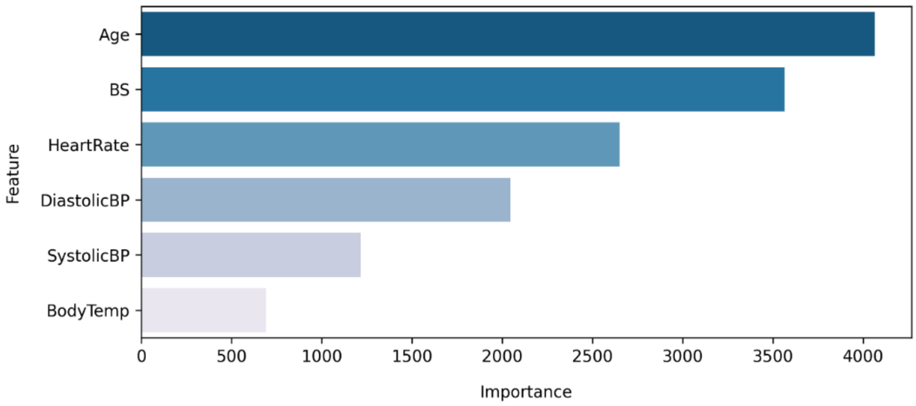

Figure 5. Feature importance of the LightGBM model on the testing set.

captured a significant amount of the variability in risk levels, there was still complexity in the data that was not completely captured by the first two principal components. The mid and low-risk individuals were more dispersed across the feature space, suggesting that these groups were less distinct based on the principal components derived. The overlap between risk categories in the PCA plot suggested that for modeling, it was challenging to achieve optimal results using a linear model alone. Given the intricate patterns and non-linear relationships apparent in the data, our LightGBM approach benefited from this complexity.

The exploratory data analysis provided crucial insights into the underlying structure and relationships within the data, identifying trends and patterns that justified the use of LightGBM. The correlation matrix, in particular, helped to understand the bivariate relationships and potential multicollinearity between variables. Additionally, PCA was employed to visualize the distribution of risk levels. This foundational understanding ensured that the subsequent modeling was based on informed decisions, thereby enhancing the interpretability and reliability of the predictive outcomes.

3.2. LightGBM Model

The LightGBM model was successfully trained for maternal risk detection in this study. The dataset was divided into two parts, with 80% used for training and the remaining 20% for testing. The training data were further utilized in a 10-fold cross-validation process, combined with random search hyperparameter tuning. The optimal hyperparameters were identified as follows: colsample_bytree 0.59, learning_rate 0.27, max_depth 13, n_estimators 219, and subsample 0.94. The performance metrics of the model are detailed in Table 3.

For the LightGBM model, the achieved accuracy of 84.73% indicates that the model correctly predicted the maternal risk level in approximately 84.73% of the cases. The precision score of 84.75% suggests that when the model predicts a specific risk level, it is correct about 84.75% of the time. The recall of 84.73% indicates that the model successfully identified 84.73% of all actual risk levels. Finally, the F1-score, which balances precision and recall, was also 84.73%, showing a harmonious balance between precision and recall in the model's predictions. These results were compared with four other machine learning models: Random Forest, AdaBoost, K-NN, and SVM. The LightGBM model outperformed these models in all metrics, demonstrating its superior effectiveness in predicting maternal health risks.

To assess the performance of the LightGBM model on distinct risk levels, the confusion matrix was visualized in Figure 4. The matrix depicted a concentration of predictions along its diagonal, indicating accurate classifications with 67 'Low Risk,' 64 'Medium Risk,' and 41 'High Risk' correct predictions. Misclassifications were relatively few, with the majority occurring between 'Low Risk' and 'Medium Risk' categories. There were minimal misclassifications between 'Low Risk' and 'High Risk', suggesting the model's effectiveness in differentiating between the most contrasting risk categories.

To discern which features were most influential in the LightGBM model's decision-making process, the feature importance was visualized in Figure 5. The visualization elucidates the relative importance of each feature in the model's predictions. Age emerged as the most significant feature, indicated by the longest bar in the chart. Blood sugar was identified as the second most influential feature. Features such as heart rate and diastolic blood pressure exhibited moderate importance, while systolic blood pressure and body temperature were the least impactful on the model's predictions.

3.3. Discussion

The implications of this study are significant for the field of maternal health. The LightGBM model's robust performance, as evidenced by its predictive accuracy, precision, recall, and F1-score, indicates that machine learning can play a pivotal role in early detection and intervention in maternal health risk. The model's ability to effectively differentiate between low, medium, and high-risk levels, with minimal misclassification as demonstrated by the confusion matrix suggests its potential in assisting healthcare professionals to prioritize and tailor care for expectant mothers. LightGBM also known for its less computational resources, which is important for practical applications in resource-constrained settings like rural healthcare centers.

The finding that age and blood sugar are the most predictive features could influence how healthcare providers assess risk factors during pregnancy. These indicators could be emphasized in screening processes, leading to earlier interventions for at-risk populations. The study also highlights the capability of IoT-based systems to support continuous and real-time monitoring, which is crucial for timely and accurate risk detection. The exploration of feature importance leads to a better understanding of the underlying dynamics that contribute to maternal health risks. Such insights could be valuable in refining data collection strategies and focusing on the most relevant parameters, thereby improving model performance and the efficiency of maternal healthcare delivery.

The adaptability of the LightGBM model to handle complex and high-dimensional data implies that it could be extended to other areas of healthcare beyond maternal risk detection. Its application might be explored in managing chronic diseases, predicting hospital readmissions, or in personalized medicine, where similar data-rich environments are common.

3.4. Limitation and Future Directions

This study, while demonstrating the utility of the LightGBM model in maternal health risk detection, is not without limitations. One of the primary constraints is the reliance on a dataset from a single geographical area, which may not encapsulate the diversity of maternal health profiles worldwide. This regional focus limits the generalizability of the findings, and the model's applicability might vary across different populations with diverse ethnic, socioeconomic, and genetic backgrounds. Another limitation is the static nature of the dataset. Maternal health is dynamic, with risk factors and individual health profiles changing throughout the course of pregnancy. The study's cross-sectional design does not account for these temporal variations, which could affect the model's predictive accuracy over time.

For future studies, expanding the dataset to include a wider range of demographics and geographic locations would be valuable to increase the robustness and generalizability of the model. Incorporating longitudinal data could provide insights into how risk factors evolve over time, enhancing the model's predictive performance throughout the different stages of pregnancy. Further research could also explore the integration of additional types of data, such as lifestyle, dietary information, or genetic markers, to enrich the model's understanding of risk factors. There is also a potential to investigate the use of sequential models or time-series analysis to predict changes in maternal health risk over time.

Lastly, prospective studies could assess the real-world impact of implementing such a model in clinical workflows, examining how it affects decision-making, resource allocation, and ultimately, patient outcomes. The feasibility of deploying these models in low-resource settings, where healthcare infrastructure may be lacking, also warrants examination. Through these future endeavors, the full potential of machine learning in improving maternal health outcomes can be realized.

4. Conclusions

This study has demonstrated the potential of the LightGBM model to effectively detect maternal health risks, leveraging a dataset of health indicators from rural Bangladesh. Despite its regional focus and the inherent complexities of predictive modeling, the LightGBM model outperformed other models, offering an efficient and powerful tool for early intervention in maternal healthcare. Future studies are encouraged to expand the dataset, enhance model interpretability, and explore real-world implementation to fully harness the capabilities of machine learning in improving maternal health outcomes globally.

Author Contributions: Conceptualization, T.R.N., S.I.N. and R.I.; methodology, T.R.N.; software, T.R.N.; validation, R.R., I.F. and R.I.; formal analysis, T.R.N.; investigation, T.R.N., S.I.N.; resources, R.R., I.F., and R.I.; data curation, R.R., I.F. and R.I.; writing—original draft preparation, R.I.; writing—review and editing, S.I.N., R.R., I.F. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset related to maternal health risk used in this study is accessible at the following link: https://archive.ics.uci.edu/dataset/863/maternal+health+risk.

Acknowledgments: The authors would like to express gratitude to their respective institutions for the support and resources provided throughout the completion of this work.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Cheng, C.-Y., Fowles, E. R., and Walker, L. O. (2006). Postpartum Maternal Health Care in the United States: A Critical Review, *Journal of Perinatal Education*, Vol. 15, No. 3, 34–42. doi:10.1624/105812406X119002.
2. Gingrey, J. P. (2020). Maternal Mortality: A US Public Health Crisis, *American Journal of Public Health*, Vol. 110, No. 4, 462–464. doi:10.2105/AJPH.2019.305552.
3. Sasmita, N. R., Ramadeska, S., Utami, R., Adha, Z., Putri, U., Syarafina, R. H., Reskiaddin, L. O., Kamal, S., Yarmaliza, Y., Muliadi, M., and Saputra, A. (2023). Maternal and Child Healthcare Services in Aceh Province, Indonesia: A Correlation and Clustering Analysis in Statistics, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 22–31. doi:10.60084/ijds.v1i1.88.
4. Dietz, P., Bombard, J., Mulready-Ward, C., Gauthier, J., Sackoff, J., Brozicevic, P., Gambatese, M., Nyland-Funke, M., England, L., Harrison, L., and Taylor, A. (2014). Validation of Self-reported Maternal and Infant Health Indicators in the Pregnancy Risk Assessment Monitoring System, *Maternal and Child Health Journal*, Vol. 18, No. 10, 2489–2498. doi:10.1007/s10995-014-1487-y.
5. Frick, A. P. (2021). Advanced Maternal Age and Adverse Pregnancy Outcomes, *Best Practice & Research Clinical Obstetrics & Gynaecology*, Vol. 70, 92–100. doi:10.1016/j.bpobgyn.2020.07.005.
6. Meyer, A. N. D., Giardina, T. D., Khawaja, L., and Singh, H. (2021). Patient and Clinician Experiences of Uncertainty in the Diagnostic Process: Current Understanding and Future Directions, *Patient Education and Counseling*, Vol. 104, No. 11, 2606–2615. doi:10.1016/j.pec.2021.07.028.
7. Heslehurst, N., Ngongalah, L., Bigirumurame, T., Nguyen, G., Odeniyi, A., Flynn, A., Smith, V., Crowe, L., Skidmore, B., Gaudet, L., Simon, A., and Hayes, L. (2022). Association between maternal adiposity measures and adverse maternal outcomes of pregnancy: Systematic review and meta-analysis, *Obesity Reviews*, Vol. 23, No. 7. doi:10.1111/obr.13449.
8. Maulana, A., Faisal, F. R., Noviandy, T. R., Rizkia, T., Idroes, G. M., Tallei, T. E., El-Shazly, M., and Idroes, R. (2023). Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 1–7. doi:10.60084/ijds.v1i1.72.
9. Atallah, R., and Al-Mousa, A. (2019). Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method, *2019 2nd International Conference on New Trends in Computing Sciences (ICTCS)*, IEEE, 1–6. doi:10.1109/ICTCS.2019.8923053.
10. Noviandy, T. R., Alfanshury, M. H., Abidin, T. F., and Riza, H. (2023). Enhancing Glioma Grading Performance: A Comparative Study on Feature Selection Techniques and Ensemble Machine Learning, *2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, IEEE, 406–411. doi:10.1109/IC3INA60834.2023.10285778.
11. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

12. Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., Richard-Nnabu, N. E., and Louni, A. (2022). A Classification Algorithm-Based Hybrid Diabetes Prediction Model, *Frontiers in Public Health*, Vol. 10. doi:10.3389/fpubh.2022.829519.

13. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.

14. Suhendra, R., Suryadi, S., Husdayanti, N., Maulana, A., Noviandy, T. R., Sasmita, N. R., Subianto, M., Earlia, N., Niode, N. J., and Idroes, R. (2023). Evaluation of Gradient Boosted Classifier in Atopic Dermatitis Severity Score Classification, *Heca Journal of Applied Sciences*, Vol. 1, No. 2, 54–61. doi:10.60084/hjas.v1i2.85.

15. Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., Reza, M. T., and Khan, M. R. H. (2022). 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, 1–17.

16. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, Vol. 30.

17. Ahmed, M., Kashem, M. A., Rahman, M., and Khatun, S. (2020). Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT), 357–365. doi:10.1007/978-981-15-2317-5_30.

18. Callister, L. C., and Edwards, J. E. (2017). Sustainable Development Goals and the Ongoing Process of Reducing Maternal Mortality, *Journal of Obstetric, Gynecologic & Neonatal Nursing*, Vol. 46, No. 3, e56–e64. doi:10.1016/j.jogn.2016.10.009.

19. Chen, C., Zhang, Q., Ma, Q., and Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion, *Chemometrics and Intelligent Laboratory Systems*, Vol. 191, 54–64. doi:10.1016/j.chemolab.2019.06.003.

20. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.

21. Rufo, D. D., Debelee, T. G., Ibenthal, A., and Negera, W. G. (2021). Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM), *Diagnostics*, Vol. 11, No. 9, 1714. doi:10.3390/diagnostics11091714.

22. Simeon, S., and Jongkon, N. (2019). Construction of Quantitative Structure Activity Relationship (QSAR) Models to Predict Potency of Structurally Diversed Janus Kinase 2 Inhibitors, *Molecules*, Vol. 24, No. 23, 4393. doi:10.3390/molecules24234393.

23. Noviandy, T. R., Idroes, G. M., Maulana, A., Hardi, I., Ringga, E. S., and Idroes, R. (2023). Credit Card Fraud Detection for Contemporary Financial Management Using XGBoost-Driven Machine Learning and Data Augmentation Techniques, *Indatu Journal of Management and Accounting*, Vol. 1, No. 1, 29–35. doi:10.60084/ijma.v1i1.78.

24. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.

25. Maulana, A., Noviandy, T. R., Suhendra, R., Earlia, N., Sofyan, H., Subianto, M., and Idroes, R. (2023). Performance Analysis and Feature Extraction for Classifying the Severity of Atopic Dermatitis Diseases, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, 226–231. doi:10.1109/COSITE60233.2023.10249760.