NU LAGUNA

MATERNAL HEALTH RISK PREDICTION USING XGBOOST CLASSIFIER

A Thesis Presented to the
Faculty of College of Computer Studies
NU Laguna

In Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science in Computer Science

Submitted by

Manjares, Laurence EJ C.
Abas, Vincent A.
Encarnacion, Lord Stephen C.

June 2024

APPROVAL SHEET

This thesis titled **MATERNAL HEALTH RISK PREDICTION USING XGBOOST CLASSIFIER** has been prepared and submitted by **ABAS, VINCENT A., ENCARNACION, LORD STEPHEN C., MANJARES, LAURENCE EJ C.** in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science.**

JOSEPH L. DOMINGO, MIT
Research Adviser

Approved in partial fulfillment of the requirements for the degree of **Bachelor of Science in Computer Science** by Oral Examination Committee.

MARLON A. DILOY, Ph.D. C.S. (CAR)
Panel Chair

MARC ACE J. LEGASPI
Panel Member

VINCENT S. RIVERA, MSCS
Panel Member

Accepted in partial fulfillment of the requirements for the degree of **BACHELOR OF SCIENCE IN COMPTER SCIENCE.**

____
(Date)

MARLON A. DILOY, Ph.D. C.S. (CAR)
Dean, School of Computer Studies

ABSTRACT

In the Philippines, the maternal mortality rate stands at 78 deaths per 100,000 live births, primarily due to complications arising during pregnancy and childbirth. The primary objective of this study is to develop an XGBoost model capable of predicting maternal health risks, identifying whether a patient is at risk, and providing insights into the leading contributing factors.

The model utilized a dataset with 1014 records from Kaggle and is achieving results with an accuracy of 93% and a Macro average of 93% on Precision, 92% on Recall, and 93% on F1 score. The weighted average score is 93% on Precision, Recall is 93%, and F1 score is 93%. The AUC-ROC score of the model is 97% on all classes with a micro-average ROC curve, AUC of 98%, and macro-average ROC curve, AUC of 97%. The results indicate that age is the most critical feature identified by the trained model, followed by blood sugar levels, heart rate, and both diastolic and systolic blood pressure. Body temperature was found to be the least influential factor. Furthermore, using the GridSearchCV approach optimized the model's performance seamlessly, enhancing its predictive accuracy.

Keywords: Feature Importance, Maternal Health, Risk Prediction, XGBoost Classifier

DEDICATION

We want to start by thanking God for helping us with this research. We believe that His wisdom and care have been with us every step of the way. He's given us the knowledge we needed, helped us stay calm when things got stressful, and protected us from harm. We're also grateful for the talents He's given us which we've used to do this work.

Next, we want to give a big thank you to our parents. They've been our biggest supporters since day one. Whenever we felt tired or unsure, they were there to cheer us on and give us the strength to keep going. Their love and encouragement mean everything to us, and we couldn't have done this without them.

And finally, we want to acknowledge our hard work. It hasn't always been easy, and there were times when we faced challenges and setbacks. But we didn't give up. We kept pushing forward, learning from our mistakes along the way. This research is a testament to our determination and resilience, and we're proud to dedicate it to everyone who helped us make it happen.

ACKNOWLEDGEMENT

The completion of this study was a reflection of the guidance and support we have received from God and our professors. We would like to express our deepest gratitude to everyone who supported us throughout this project. We would like to acknowledge with sincere appreciation and deep gratitude the persons who are as follows:

Mr. Joseph Domingo, our thesis advisor, for his thorough technical advice, skills, knowledge, ideas, and expertise he shared with us during the model’s development, and for the experiences he shared with us. These experiences are unquestionably now a part of this work.

Mr. Vincent Rivera for persistent reminders to finish the duties assigned to them so that the research can be finished. The one provided us with suggestions, insights, and proper direction for advancing and completing the work.

Mr. Marlon A. Diloy and **Mr. Leandro De Luna** for sharing their knowledge with us through useful suggestions and knowledgeable feedback to improve the project's end product.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Background of the Study

The delicate stage of pregnancy necessitates utmost care and medical attention to ensure the health of pregnant women. Maternal health issues can arise during pregnancy, childbirth, and the postnatal period. According to B.N. Lakshmi, T.S. Indumathi, and Nandini Ravi (2016), and further supported by Finlayson K, Crossland N, Bonet M, and Downe S (2020), these periods are critical for maternal health. Aligning with the Sustainable Development Goals of the World Health Organization (WHO, 2015), specifically SDG Target 3.1, the aim is to reduce the global maternal mortality ratio (MMR) to less than 70 per 100,000 live births by 2030. The United Nations strives to lower mother and infant deaths and improve maternal health; however, the rate of decline is insufficient, with a dramatic increase in maternal mortality observed in sub-Saharan African countries, including Ethiopia. Maternal mortality remains a significant healthcare issue, especially in undeveloped countries (WHO).

The Philippines is no exception to this problem. According to the Philippines - World Bank Gender Data Portal, “78 women die per 100,000 live births due to pregnancy-related causes in the Philippines” which is almost the same average in the region. These maternal deaths were mainly caused by complications in pregnancy, childbirth and puerperium, eclampsia, pre-eclampsia, and hemorrhage, Lee-Brago, P. (2023). Dr. Leila Saiji Joudane, the country representative of the United Nations Population Fund (UNFPA) in the Philippines, highlighted the severity of maternal deaths, stating that "around six to seven Filipino women die daily due to childbirth." She noted that during emergencies when access to maternal health services is disrupted, the number of maternal deaths increases. This trend was evident in the past two years, with maternal deaths rising from 1,458 in 2019 to 2,478 in 2021. Dr. Joudane emphasized that many of these deaths are preventable with accessible medical interventions and robust healthcare systems resilient to emergencies.

Several factors come into play in determining the health risk level of a pregnant woman, such as age, blood glucose levels, blood pressure, heart rate, and body temperature, which play significant roles in assessing pregnancy-related risks. The National Institute of Child Health and Human Development (NICHD) emphasizes that age, lifestyle, and pre-pregnancy health conditions can influence pregnancy risk, with risks varying from one pregnancy to another. High blood glucose levels, as noted by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), can lead to gestational diabetes, impacting both the mother and baby, potentially causing birth problems, organ abnormalities, and even miscarriage. Additionally, according to the Centers for Disease Control and Prevention (CDC) High blood pressure during pregnancy, a serious condition, can result in pre-eclampsia, eclampsia, and stroke, posing risks to vital organs.

Furthermore, WHO (World Health Organization, 2024) concluded that most of the risks encountered in pregnancy are preventable or treatable especially when attended by skilled health professionals. According to their latest available joint data with the United Nations International Children's Emergency Fund (UNICEF), specifically the database on SDG 3.1.2 Skilled Attendance at Birth, wealthy countries, and somewhat well-off countries, about 99% of births have a trained midwife, doctor, or nurse to help. But in poorer countries, only 68% in very poor countries and 78% in somewhat less poor countries get assistance from these trained healthcare workers. This simply means that if provided with proper maternal health care for every expectant mother, the sustainable development goals related to issues in maternal morbidity and mortality will be resolved. Some studies employed machine learning techniques that predict factors that contribute to the health of a Pregnant woman, such as the study of Wu, Y., Zhang, C., Mol, B. W. J., Kawai, A., Li, C., Chen, L., Wang, Y., Sheng, J., Fan, J., Shi, Y., & Huang, H. (2020), entitled "Early Prediction of Gestational Diabetes Mellitus in the Chinese Population via Advanced Machine Learning" they achieved high accuracy in predicting Gestational Diabetes Mellitus in early pregnancy. They were also able to investigate the relationship of GDM with thyroxine and BMI in the Chinese Population.

In the study of Betts, K., Kisely, S., & Alati, R. (2019) entitled "Predicting common maternal postpartum complications: leveraging health administrative data and machine learning" they were able to predict the risk of common maternal postpartum complications with the use of Gradient boosted trees achieving a good discrimination as it had effectively distinguished between individuals who developed postpartum hypertensive disorders and those who did not, with an AUC (Area Under the Curve) of 0.879, it has a 95% CI (Confidence Interval) of 0.846–0.912. It was also able to distinguish individuals with obstetric surgical wound infection gaining an AUC of 0.856, it has a 95% CI of 0.838–0.873. On the other hand, postpartum sepsis and hemorrhage obtained poor discrimination. These are just examples of studies that use machine learning, it is also evident in other studies that used machine learning to gain high accuracy in predicting health issues in pregnant women.

Research Objectives

The general objective of this study is to develop a model that will predict if a pregnant woman is at risk and identify the leading causes by employing the XGBoost algorithm.

The specific objectives are as follows:

1. identify primary variables that contribute to the predicted result.
2. train a model with tuned hyperparameters that could predict maternal health risk using the XGBoost algorithm.
3. test the model using precision, accuracy, recall, and F1 score.
4. analyze feature importance, interpret results, and discuss the practical implications of the model's predictions, including potential interventions or strategies for reducing maternal health risks based on the model's insights.

Theoretical Framework

This section outlines the application of machine learning using the XGBoost algorithm to predict maternal health risk, integrating machine learning in healthcare will be a significant approach to harness it. This research initiative emphasizes evidence-based maternal health risk assessment and acknowledges the potential of machine learning to make a significant positive impact on the lives of families.

Understanding Extreme Gradient Boosting

Due to the significant increase in data collection and storage that has been steadily growing over the years, huge data applications are rapidly becoming the center of attention. Machine learning technology has been proven to be an effective tool for classification problems with multi-parameters. Extreme Gradient Boosting is a new ensemble learning algorithm that builds upon the principles of Gradient Boosting Machines (GBM) and introduces several enhancements and optimizations (Jafarzadeh, H., et al., 2021). XGBoost operates by training multiple decision trees, with each tree being trained on a specific subset of the data. The predictions generated by each individual tree are then aggregated to produce the final prediction (Verma N., 2022).

XGBoost is a highly advanced machine learning technique that demonstrated remarkable proficiency in modeling structured data sets. Moreover, XGBoost boasts swift computation and improved design features to effectively tackle overfitting which helps in the model's generalization in the process (Khandelwal, N., 2020). This algorithm is recognized for its computational efficiency and notable success in the field of machine learning (Bakouregui et al., 2021).

Research Methodological Approach

The foundational ideas for this research are drawn from the comprehensive work of Raza A., Siddiqui HUR, Munir K, Almutairi M, Rustam F, Ashraf I. (2022), titled "Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction." This framework outlines a well-organized plan for predicting health risk levels in pregnant women, structured through a series of interconnected stages that systematically guide the research process.

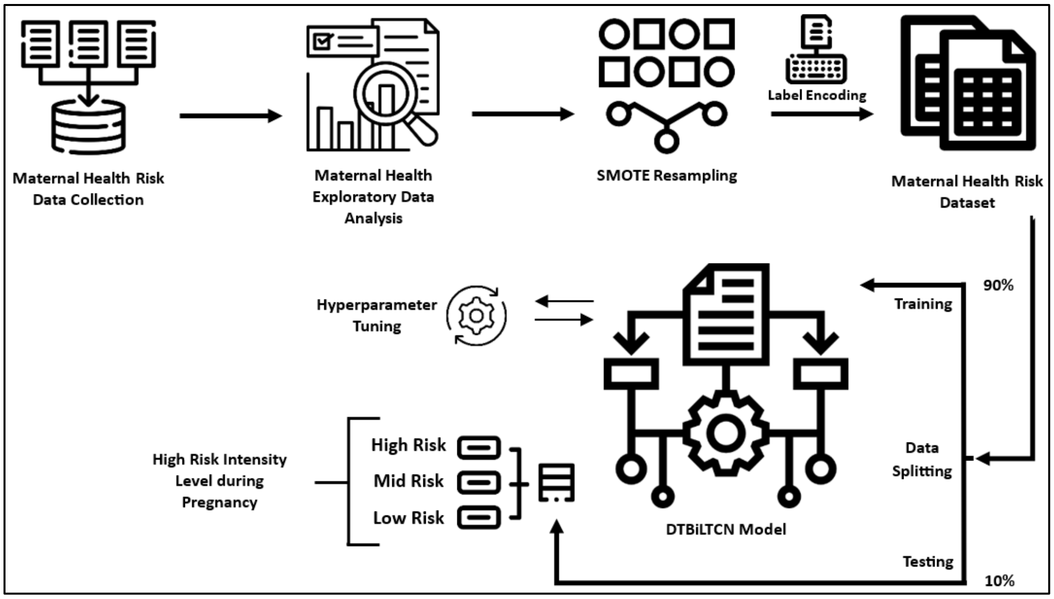

Figure 1. Methodological approach of Raza A, Siddiqui HUR, Munir K, Almutairi M, Rustam F, Ashraf I. (2022).

As shown in Figure 1, the research begins with the critical phase of data collection and preprocessing. This involves gathering a maternal health risk dataset, addressing outliers, and exploring relationships between variables to ensure the quality and relevance of the data. According to Sandbhor, S., and Chaphalkar, N.B. (2019), outliers can significantly affect the quality of a dataset and subsequently the performance of machine learning models. To mitigate these issues, feature engineering techniques are employed to handle class imbalance using z-scores, which measure the number of standard deviations a data point is from the mean.

Following preprocessing, the data is split into training and testing sets to provide flexibility in the choice of split ratios. This step is crucial for assessing model performance under various conditions and ensuring the model's generalizability.

The model selection phase involves exploring various machine learning and statistical models, ultimately concluding with the implementation of the XGBoost classifier. This decision is based on the model's ability to handle large datasets and its superior performance in predictive accuracy.

Next, the model undergoes training and tuning, which includes hyperparameter optimization. This process enhances the model's accuracy and reliability, as demonstrated in the study by Raza et al. (2022), where hyperparameter tuning significantly improved the predictive outcomes.

Model evaluation is a vital step in this framework. Performance metrics such as accuracy, precision, recall, F1 score, and AUC are used to assess the model's predictive capabilities. These metrics provide a comprehensive understanding of the model's strengths and limitations, ensuring a robust evaluation process.

Finally, the trained machine learning model is utilized to make predictions regarding maternal health risks. This application of the model aims to provide accurate and reliable health risk assessments for pregnant women, contributing to better health outcomes.

In summary, this theoretical framework is built on a structured approach involving data collection and preprocessing, model selection, training and tuning, and rigorous evaluation, culminating in the practical application of the machine learning model for maternal health risk prediction.

Conceptual Framework

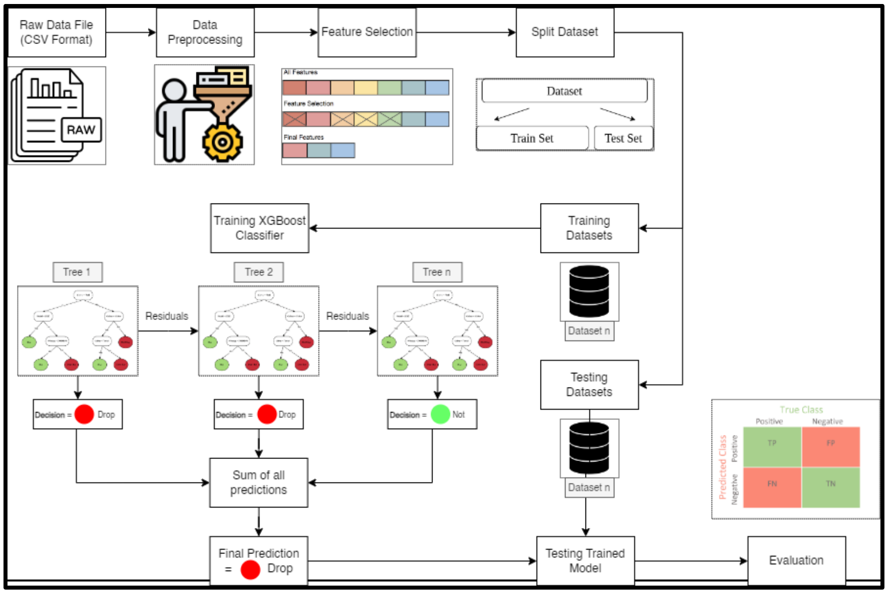

Figure 2. Diagram of the Conceptual Framework

The process starts with an extensive data collection process, specifically gathered from Kaggle and used in the study of M. Ahmed and M. A. Kashem (2020) entitled “IoT Based Risk Level Prediction Model for Maternal Health Care In The Context Of Bangladesh.” This is crucial in making sure that the data comes from reliable and relevant sources.

This is followed by the stage of data preprocessing, where the data is cleaned and refined. This step, which involves eliminating improbable values and dealing with outliers to guarantee the model's accuracy, is crucial for guaranteeing the data's integrity and usefulness.

Feature engineering is employed to refine the dataset further. Techniques are applied to eliminate outliers, thereby preventing unwanted results on the model's statistical assessments. This step is crucial for maintaining the reliability of the predictive outcomes.

Subsequently, the dataset is split into training and training subsets using an 80:20 split ratio. The training set is used to build the model, while the testing set is reserved for evaluating the model's performance.

The XGBoost classifier is then introduced as the predictive model for the study. Renowned for its efficiency in handling large datasets and its exceptional predictive accuracy, the XGBoost classifier undergoes hyperparameter optimization. This optimization is pivotal for enhancing the model's performance metrics to produce the most accurate and reliable predictions which is crucial in the medical setting.

Drawing inspiration from Raza et al. (2022), the research adheres to a structured approach for predicting health risk levels. The process incorporates feature engineering strategies to address class imbalances, utilizing z-scores to normalize the data distribution.

The model evaluation phase is integral to the framework, employing a suite of performance metrics—accuracy, precision, recall, F1 score, and AUC, including the micro and macro metrics—to rigorously assess the model's predictive capabilities. These metrics furnish a holistic view of the model's performance, highlighting its strengths and areas for improvement.

This conceptual framework encapsulates the essence of the theoretical framework by Raza et al. (2022), ensuring a coherent and systematic approach to the research study.

Scope and Limitations of the Study

Scope

The study focuses on the use of the XGBoost algorithm for predicting maternal health risks, utilizing a dataset from Kaggle. This dataset has 1,014 records with 7 features: Age, SystolicBP (Systolic Blood Pressure), DiastolicBP (Diastolic Blood Pressure), BS (Blood Sugar), BodyTemp (Body Temperature), HeartRate, and Risk. Data records with impossible values will be removed to ensure accuracy. Machine learning processes such as pre-processing data, model training, and evaluation will be conducted using the Jupyter Notebook.

Limitation

There are certain limitations to consider. The study will solely rely on the XGBoost algorithm for predicting maternal health risk and only the dataset from Kaggle will be used to train the model. Other features that have relevance in the field of maternal health and may contribute to the performance of the model that are not included in the dataset will not be used.

Significance of the Study

Promotion of Value and Social Relevance. This research holds substantial social significance as it directly enhances the quality of maternity healthcare services. By improving the accuracy of health risk assessments, this study increases the social relevance of healthcare, making it more effective and accessible for pregnant women.

Contribution to Nation Building. In the context of this research, it contributes to the development of the healthcare system which is crucial for the nation's progress.

Aligning healthcare interventions with individual health risks empowers healthcare professionals to meet the diverse needs during maternity and foster the national development and growth in maternal healthcare.

Contribution to existing body of knowledge (Computer Science). This research contributes to the growing body of knowledge in Computer Science by exploring how these techniques can be applied to predict maternal health risks. It can also help in providing valuable insights for the further development of the tool in healthcare and personalized risk assessment systems.

Continuous improvement of the teaching-learning process. This study aims to enhance the maternal healthcare process by implementing machine learning which has the potential to lead to more tailored and effective healthcare strategies that may result in the development of patient engagement and improve maternal health outcomes in healthcare domains.

Definition of Terms

Accuracy. It is the result of whether the model was able to match the predicted values to the actual values.

Boosting. It is a method to reduce errors in the weak learners and therefore will improve weak learners to have a significantly better prediction power.

Confusion Matrix. It is where you evaluate the accuracy of the trained model by using the actual data and analyzing whether the model is accurate or not based on its percentage.

Dataset. It is a collection of data that is used for analysis and machine learning.

Eclampsia. It is a severe and life-threatening complication of pre-eclampsia during pregnancy.

F1 Score. It assesses the accuracy of a model on a certain dataset and is a machine learning evaluation statistic. It is known as the harmonic mean and combines a model's precision and recall scores. It is applied to the assessment of binary categorization schemes.

Hemorrhage. It is when a person loses a lot of blood, often due to injury or a medical condition, and it can be a serious and potentially life-threatening situation.

Internet of Things. It is the integration concept of all devices that are readable, addressable recognizable, locatable, and manageable via the Internet through RFID (Radio-Frequency 2 Identification), wireless local area network, wide area network, etc. It also provides real-time information and interacts with real-time users.

Machine Learning Algorithm. It is a mathematical model or a program that can automatically learn from data that is provided and then can be used to generate predictions or decisions based on that learning.

Maternal Health Risk. It means the possible health problems or dangers that a mother might encounter during pregnancy.

Maternity. It refers to the state of being pregnant and giving birth to a child. It is also the period of motherhood and caring for the newborn.

Model. Is a machine learning algorithm, specifically XGBoost, designed to predict maternal health risks.

Overfitting. This can be called "overlearning" or "over-adapting" which occurs when a model is excessively complex relative to the amount of available training data.

Pre-eclampsia. It is a serious condition that can happen during pregnancy where a woman's blood pressure becomes very high, and it can affect the health of both the mother and the baby.

Precision. It measures how close the actual value and the expected value match up.

Pregnancy. It is when a woman has a baby growing inside her womb.

Puerperium. It is the period of time (typically last for about six weeks that involves physical and hormonal adjustments) just after childbirth when a woman's body goes through various changes as it recovers from pregnancy and childbirth.

Recall. It is the measure of whether the predicted value was able to belong to its rightful class in the confusion matrix.

Regularization. In machine learning, it refers to a set of techniques used to prevent overfitting and improve the generalization ability of a model.

Residuals. The discrepancies between the target variable's actual values and its predicted values.

Training. It is the method for a machine learning algorithm to learn.

Training Set. It is a dataset used to train a particular machine-learning algorithm.

Testing. It is the method to know if the trained machine learning is effective in terms of its accuracy in predicting.

Testing Set. It is a dataset used to test the trained machine-learning algorithm.

XGBoost. Is a gradient-boosting algorithm that has been tuned for scalability. It is an ensemble method that combines the predictions of many weak models to create a strong and precise machine-learning model.

Acronyms

The following acronyms and their meanings were used throughout the study:

ANN	Artificial Neural Network
AUC	Area Under Curve
BEmONC	Basic Emergency Obstetrics and Newborn Care
CI	Confidence Interval
DT	Decision Tree
DOH	Department of Health
FN	False Negative
FP	False Positive
GBM	Gradient Boost Machines
IDE	Integrated Development Environment
IoT	Internet of Things
KOICA	Korea International Cooperation Agency
ML	Machine Learning
MMR	Maternal Mortality Ratio
NICHD	National Institute of Child Health and Human Development
NIDDKD	National Institute of Diabetes and Digestive and Kidney Diseases
PSA	Philippine Statistics Authority
ROC	Receiver Operating Characteristic
SDG	Sustainable Development Goals

TN	True Negative
TP	True Positive
UNFPA	United Nations Population Fund
WHO	World Health Organization
XGB	Extreme Gradient Boosting

Chapter 2

REVIEW OF RELATED LITERATURE AND STUDIES

This chapter's review of related literature and studies is critical for the development and modeling of maternal health risk prediction since it provides a firm basis for the study and highlights gaps in current knowledge that must be addressed. Furthermore, the review of related literature and studies can aid in identifying potential obstacles and limitations connected to the modeling of the maternal health risk prediction model, as well as future research prospects. Furthermore, this study presents a synthesis of linked literature and studies, which may then be verified, contradicted, or enhanced by the new information offered by this study. The incorporation of related literature and studies is crucial for the development of this model since it guarantees that the framework is anchored in a solid theoretical and realistic foundation and can make significant contributions to NU Laguna and the field of machine learning.

Maternal Health in the Philippines

In a study by Cagayan, M. S. F. S., Nisperos, G. A., Facun, G.-M. G., Cagayan, B. S. S., Castro, M. C. R., & Silverio, C. E. (2022), the researchers aimed to identify the factors that obstruct the utilization of maternal care services available in Luzon which they found has the highest percentage of maternal mortality in 2020 which is 52.5% according to Philippine Statistics Authority (PSA). They identified that there is a limited-service capacity of BEmONC (Basic Emergency Obstetrics and Newborn Care) facilities which were established in the Philippines (2014) along with other community-level service providers to help restrain the high MMR. They also identified other significant factors such as financial constraints, location, and availability of transportation. In the article of Ramirez

L. (2023, August 24), the Communications Associate of WHO Philippines, it stated that "according to Field Health Service Information 2022, infant mortality rate in the country is at 10.36 (target is 15 per 1,000 live births), maternal mortality rate is at 64.68 (target is 70 per 100,000 live births) (Philippines – World Back Gender Data Portal, "78 women die per 100,000 live births due to pregnancy-related causes in the Philippines), and adolescent birth rate is at 24.36 (target is 37 per 1,000 females aged 15-19 years old)." Even though the national figures are in an improving trend compared to previous years, the sub-national level remains alarming for having MMR ranges from 29 to 152. This suggests that different regions in the country have differences in maternal health outcomes. Since it is a concern, efforts to improve healthcare services and outcomes are needed to address these variations to ensure that healthcare is evenly distributed and accessible across the country.

Maternal Health Risks

All pregnancies carry risks. According to Cleveland Clinic, their definition of a "high-risk" pregnancy is "any pregnancy that carries increased health risks for the pregnant person, fetus or both" that needs extra maternal health care during and after giving birth which can reduce the chance of complications. Maternal health risk refers to potential challenges and complications that can affect a woman's health during pregnancy, childbirth, and the postpartum period. These risks can vary based on factors such as age, pre-existing medical conditions, and lifestyle in which addressing them is crucial to ensure the well-being of both the expectant mother and the baby.

Several factors are partly responsible for maternal health morbidity and mortality. Cleveland Clinic stated on its website that a high-risk pregnancy involves increased health risks for the expectant mother, baby, or both. Having specific health issues and being either older than 35 or younger than 17 during pregnancy can make a pregnancy high risk that requires close monitoring to reduce the possibility of complications. According to the National Institute of Child Health and Human Development (NICHD), age, lifestyle, and existing pre-pregnancy health issues can contribute to pregnancy risk. Each pregnancy is different so the risks for one pregnancy may not be the same risks for another. According to the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDKD), during pregnancy, the mother experiences hormonal changes that also affect the blood glucose (blood sugar) levels. Gestational diabetes can be developed during pregnancy by having a high blood glucose level that can be harmful to both the mother and the baby which can also affect the baby's organ and can increase the chance of having birth problems or abnormalities. It can also lead to miscarriage or a stillborn baby.

On the other hand, the mother can also experience several illnesses caused by having high blood glucose levels such as eye problems and kidney disease. That's why it is important to keep in mind that the blood glucose level is at the safe level especially if the mother has existing diabetes. Another maternal risk factor is blood pressure. According to the Texas Department of State Health Services, high blood pressure or hypertension is a serious health condition that poses increased risks during pregnancy. It has the potential to lead to complications like pre-eclampsia (elevated blood pressure) and eclampsia (life-threatening seizures or coma) in expectant mothers which typically occur after 20 weeks of pregnancy and can result in damage to vital organs such as the kidney, heart, eyes, lungs, and brain. They also mentioned that "high blood pressure during pregnancy is the most common risk factor for pregnancy-related stroke". In relation to blood pressure, pregnancy requires the heart to increase its workload. This means that the heart pumps more blood each minute which increases heart rate. According to the Mayo Foundation for Medical Education and Research (2023), this results in a 30% - 50% blood volume increase to also support the growing baby. In exchange for this, the expectant mother might encounter maternal health risks such as heart rhythm issues, heart valve issues, congestive heart failure, and congenital heart defects which can pose life-threatening risks for both the mother and the growing baby.

In the 21st century, climate change is the most significant global threat that is affecting the health of many, WHO (2021). It makes extreme heat more intense, lasts longer, happens more often, and can be a threat to vulnerable groups including babies and expectant mothers, Cil, G. and Cameron, T.A. (2017). There is evidence that pregnant women are at risk of heat stress due to various factors, Di Napoli, C., Pappenberger, F., & Cloke, H. L. (2019). According to Linn B. Strand, Adrian G. Barnett, and Shilu Tong (2011), these include an increase in body weight and body fat, which can raise core body temperature and heat production. Additionally, pregnancy reduces the body's ability to cool down through sweating due to changes in the ratio of surface area to body mass.

Application of Machine Learning

Accurate prediction of postpartum complications and early gestational diabetes mellitus (GDM) during the first trimester of pregnancy is a crucial aspect of maternal healthcare. A study by Betts K, et al. (2019) focused on predicting common maternal postpartum complications necessitating inpatient care, utilizing administrative health data from over 400,000 inpatient live births in Queensland, Australia, between 2009 and 2015. This includes 20 predictors for each postpartum complication outcome (Hypertensive disorders, Sepsis, Postpartum hemorrhage, and Surgical wound infection). The following are the predictors for hypertensive disorder: Severe pre-eclampsia, Pre-eclampsia (unspecified), Gestational hypertension, Mild to moderate pre-eclampsia, Birthweight, Birth length, Allied health intervention – physiotherapy, Emergency lower section cesarean, Maternal BMI, Maternal height, Maternal weight, Maternal age, Pre-existing hypertension complicating pregnancy, Birth head circumference, Local hospital area code (confidentialised), Gestational weeks at delivery, Regular respiration at birth, HELLP syndrome, SEIFA codes, and Prophylactic antibiotic administration for cesarean.

The following are the predictors for sepsis: Postpartum hemorrhage, Maternal BMI, Private patient, SEIFA score, Maternal age, Birth length, Maternal weight, Birthweight, Maternal infections complicating pregnancy, Puerperal sepsis, Puerperal sepsis, Maternal height, Symptoms involving digestive system, Medical practitioner—private specialist, Gestation weeks at birth, General symptoms and signs, Mental/nervous system disorders of pregnancy, Month born, Allied health intervention—dietetics, and Excessive vomiting in pregnancy. The following are the predictors for postpartum hemorrhage: Postpartum hemorrhage, Maternal BMI, Birthweight, Maternal height, Administration of packed blood cells, Caesarean delivery, Allied health intervention—physiotherapy, Management of postpartum hemorrhage, Gestational weeks at delivery, Medical practitioner—private specialist, Birth length, Birth head circumference, Maternal age, Other conditions complicating pregnancy, Multiple deliveries, Maternal weight, SEIFA codes, Local hospital area code (confidentialised), Retained placenta, membranes without hemorrhage, and Postpartum manual exploration of uterine cavity.

The following are the predictors for surgical wound infection: Caesarean delivery, Emergency lower section cesarean, Maternal BMI, Maternal weight, Private patient, Medical practitioner—private specialist, Maternal age, Birth weight, Birth head circumference, Episiotomy, Maternal height, Prenatal smoking, Birth length, Elective lower section cesarean, Previous pregnancies, Month born, Obstetric surgical wound infection, SEIFA codes, Gestational weeks at delivery, and Abnormalities of forces of labor. The research employed machine learning techniques, such as gradient-boosted trees, to develop predictive models. The results showed promising discrimination for postpartum hypertensive disorders and obstetric surgical wound infections while indicating areas for improvement in data collection to enhance prediction accuracy.

In the study of Ahmed, M., Kashem, M. A., Rahman, M., & Khatun, S. (2020), they conducted a study employing machine learning models and determined that the Logistic Model Tree (LMT) classifier outperforms other models in the analysis of factors associated with maternal health. The following are the parameters in the study: Blood Pressure, Heart rate, Body Temerature, Fetal Movement, Age, BMI, Blood glucose (2-hour glucose), Blood glucose (fasting glucose), and Blood glucose (HbA1c). The data from an IoT-enabled system were gathered and applied to the LMT model, resulting in a 90% accuracy rate.

Following this study is the study of Raza et al (2022) Suggested an ensemble approach, BiLTCN, which integrates the Neural Network-based Bidirectional Long Short-Term Memory (BiLSTM), Temporal Convolutional Network, and Decision Tree as a classifier, leveraging a clinical dataset comprising 1218 instances collected through an IoT-enabled system. The researchers gathered and used a publicly available dataset from Kaggle that contains 7 features including age, SystolicBP, DiastolicB, BS, BodyTem, HeartRate, and Risklevel as target classes. The proposed system demonstrated outcomes following balancing through SMOTE, achieving an average accuracy of 88%.

Additionally, they employed feature selection methods, incorporating SVM alongside BiLTCN, and reported a 98% accuracy on the reduced feature model.

Another study by Wu Y, et al. (2020) aimed to predict early GDM and addressed the lack of accurate methods for this purpose, particularly in the Chinese population. Researchers gathered pregnancy data for 73 variables during the first trimester and selected relevant variables through machine learning-driven feature selection resulting in using 7 variables after the process this includes age, family history of diabetes in a first-degree relative, multiple pregnancies, previous GDM history, FPGa, HbA1c, and TG, and developed predictive models using advanced machine learning techniques. Notably, they found that low body mass index (BMI) and certain thyroid hormones were associated with the risk of GDM, and lipoprotein(a) showed promise as a predictive marker.

Furthermore, according to Akhan Akbulut, Egemen Ertugrul, and Varol Topcu (2018) in their findings in their study entitled “Fetal health status prediction based on maternal clinical history using machine learning techniques”, Machine learning is closely connected to medical diagnosis and prediction. It means that the use of appropriate computer programs to analyze medical information enables accurate predictions about a person's health. The researchers used a total of 23 features including Maternal age, General menstrual cyclic status of the mother, Type of pregnancy, Fetal age, Blood serotype of the mother, Past delivery number (history) of the mother, Number of abortus (history), Diabetes history of mother, Hypertension history of mother, Other significant illnesses of mother, Past-surgical operations of mother, Consanguineous marriage status, Presence of disabled children, Presence of disabled persons in mother’s family, Presence of disabled persons in father’s family, Result of double test, Result of triple test, Result of quad test, Drug-usage during the pregnancy, Alcohol taking status, Smoking status, Any existing illnesses of the mother regarding the pregnancy, and Fetal health status.

In the study of Meshram, Jagruti & Devi, Seeta & Gaikwad, Sachin & Podder, Lily & Ramachandran, and Harikrishnan. (2023), they investigated the effectiveness of using a GESTOSIS score, a validated tool for predicting Pregnancy-Induced Hypertension (PIH) risk in high-risk pregnant women with the application of machine learning algorithms. They used the features Risk_Score_Sum, Pre-eclampsia family history, GDM, Women with CAD, Increased weight gain in pregnancy, Chronic Hypertension, Age interval, Below 19, Dyslipidemia, Women with SGA, hypothyroidism, Pregnancy with IVF, Anemia, Primigravida, Pitting edema, PCOD, Obesity, and Above 35. They tested it on 70 pregnant women and found that the computer program was good at predicting the risk, with an accuracy of 97% to 99%. This means that using a computer program with the GESTOSIS score can help doctors and health workers identify women at high risk of high blood pressure during pregnancy more reliably. It makes it easier for healthcare workers in the community to use this tool without needing complicated tests.

In the study of Arrieta, Eugenia & Estrada, Francisco & Caicedo, William & Martinez Santos, Juan Carlos. (2016), entitled “Early Prediction of Severe Maternal Morbidity Using Machine Learning Techniques”, the researchers used logistic regression to predict the risk of severe maternal morbidity in pregnant women who are receiving maternal care at a clinic in Cartagena, Colombia. They used the following features Maternal parity, Pregnancy spacing less than two years, Multiple births, Prenatal care, Gestational age in first-prenatal care, Micronutrient intake, Personal history of preeclampsia, Pregnancy induced hypertension PIH, Chronic hypertension, Superimposed preeclampsia, Diabetes, Autoimmune disorders, Human immunodeficiency virus HIV, Congenital syphilis, Hepatitis B, Previous perinatal mortality, Incompatible with life VIP, Maternal causes VIP, Sexual abuse VIP, Urinary tract infection (UTI), Drinking/Smoking, Illicit and non-illicit drug use, Anemia in pregnancy, TORCH infections, Obesity in pregnancy, and Under-nutrition during pregnancy. Their model would help to provide timely and appropriate care to each patient based on their risk level after achieving some promising results.

According to Shamshuzzoha, Md., & Islam, Md. M. (2023), in their study entitled "Early Prediction Model of Macrosomia Using Machine Learning for Clinical Decision Support. Diagnostics", they tested three different machine learning algorithms for the prediction of macrosomia, they used maternal characteristics and medical history as the input data for the model and found that logistic regression yields the most accurate predictions. They also suggested that machine learning models have the potential to improve the prediction of macrosomia and can be a big help in identifying high-risk pregnancies. The study includes the following features such as maternal age, pre-pregnancy BMI, gestational age at the first prenatal visit, maternal weight gain during pregnancy, blood pressure measurements, fasting glucose levels, maternal history of gestational diabetes or diabetes, family history of diabetes, and outcome. Their study underscores the importance of leveraging machine learning algorithms in addressing macrosomia and improving decision-making for high-risk pregnancies.

In another study about the early prediction of maternal health risk factors using machine learning techniques by M. Assaduzzaman, A. A. Mamun and M. Z. Hasan (2023), they found that Random Forest algorithm provided the best results in prediction with an accuracy score of 90%, precision (90%), recall (90%), and Fl-score (90%) in maternal risk factors such as the mother's chronic condition, age, nutrition, and other medical assistance during pregnancy. The following are the variables used in the study age, Systolic BP, Diastolic BP, BS, Heart rate, and Risk level. This means that these computer programs can help us find out early what might be wrong and make sure both the mother and baby stay healthy which is very important. Akhan Akbulut, Egemen Ertugrul, and Varol Topcu (2018) stated that these predictions can be really helpful for doctors and experts to take necessary actions and better understand a problem. However, it's important to remember that the human body is very complicated, and medical decisions must not only rely on predictions. However, using machine learning in medicine can be a valuable tool to support medical professionals.

Application of XGBoost in Predicting Maternal Health

In the study conducted by Li, Y. X., Shen, X. P., Yang, C., Cao, Z. Z., Du, R., Yu, M. D., Wang, J. P., & Wang, M. (2021), the primary objective was to harness machine learning algorithms to predict the risk of pre-eclampsia (PE) among pregnant women using electronic health records. The researchers curated a dataset of 3759 cases, collected from pregnant women who received antenatal care at Xinhua Hospital Chongming Branch Affiliated with Shanghai Jiaotong University. It has eighteen binary variables and two continuous variables. Binary variables were: family history of hypertension, nulliparity, prior cesarean delivery, pregnancy interval ≥ 10 years, multifetal gestations, assisted reproductive technology, gravidity, parity, pre-gestational diabetes, heart disease, thyroid disease, renal disease, autoimmune diseases, mental disorder, uterine leiomyoma, adenomyosis, uterine malformations, history of seizure disorder. Continuous variables were maternal age and BMI. Thirty-eight clinical parameters gathered during the first antenatal care visit were utilized to construct prediction models, employing logistic regression, random forest, support vector machine, and extreme gradient boosting. Notably, the XGBoost model emerged as the standout performer, achieving an impressive accuracy of 0.920 and an area under the receiver operating curve (auROC) of 0.955. Fasting plasma glucose, mean blood pressure, and body mass index were identified as key features, with the XGBoost model showcasing superior predictive abilities, even in an easily accessible format for patient input, yielding an auROC of 0.83.

In a parallel study by Liu et al. (2020), the focus shifted to gestational diabetes mellitus (GDM) prediction in early pregnancy among Chinese women. Leveraging the XGBoost algorithm and a traditional logistic model for comparison, the researchers worked with a substantial dataset of 19,331 pregnant women. This includes records of maternal age, family history of diabetes, parity, family income, maternal education level, gravidity, pre-pregnancy body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), waist circumference (WC), hip circumference (HC), alanine aminotransferase (ALT), FPG at registration, and weight gain. The XGBoost model demonstrated a higher area under the receiver operating characteristic curve (AUC-ROC) of 0.742, surpassing the AUC-ROC of 0.663 achieved by the logistic model. This emphasized the superior predictive capability of XGBoost in identifying GDM, reinforcing its potential as a robust tool for risk assessment in early pregnancy.

Building upon these insights, Hu, X., Hu, X., Yu, Y., & Wang, J. (2023) further explored GDM prediction, specifically comparing the XGBoost machine learning model with the traditional Logistic Regression approach. With a dataset consisting of 735 women in the training set and 190 in the testing set, the XGBoost model, equipped with 20 predictors which are previous GDM, HbA1c, total protein, age, hemoglobin, plasma glucose, mean arterial pressure, fasting plasma glucose, platelet count, thyroid stimulating hormone, thyroid hormone T4, urine specific gravity, total bilirubin, ketone in urine, aspartate aminotransferase, red blood cell count, thyroid hormone T3, white blood cell count, alanine aminotransferase, glucose in urine, and pre-pregnancy BMI, outperformed the logistic regression model with four predictors. The XGBoost model exhibited an AUC of 0.946 and a predictive accuracy of 0.875, whereas the logistic regression model achieved an AUC of 0.752 and a predictive accuracy of 0.786. Remarkably, both models displayed good calibration, and the Decision Curve Analysis (DCA) indicated that selectively treating women identified as at risk by the XGBoost model provided a net benefit compared to treating all or none. This highlighted the superior predictive ability of the XGBoost model in discrimination, underlining its potential for enhancing clinical decision-making in the context of GDM.

In a broader context, Bertini, A., Salas, R., Chabert, S., Sobrevia, L., & Pardo, F. (2022) conducted a systematic review analyzing 31 articles to explore the applicability and performance of machine learning methods in predicting perinatal complications. The selected studies primarily utilized electronic medical records, medical images (resonance, ultrasound, recordings, ecotomographs, etc.), biological markers, and others (fetal heart rate and sensor) as features for prediction. Notably, machine learning methods, including support vector machine and XGBoost, demonstrated notable success in predicting complications, such as achieving 95.7% accuracy in predicting prematurity and an outstanding 99.7% accuracy in predicting neonatal mortality. This comprehensive review underlines the widespread effectiveness of machine learning in diverse perinatal contexts, reinforcing its potential as a transformative tool in women's health.

Synthesis

The synthesis of the provided information underscores the critical intersection of maternal health challenges, technological advancements in predictive analytics, and the application of machine learning algorithms. Pregnancy-related illnesses including gestational diabetes and hypertension are common, and many other factors contribute to maternal health concerns, especially in the Philippines. Studies have shown several predicted features for maternal health problems. These factors include maternal age, blood pressure, BMI, and medical history in the family.

The use of machine learning technologies can change the way pregnant women receive health care. Predictive analytics can aid medical professionals, especially nurses, and doctors, in better understanding each patient's needs, identifying high-risk pregnancies, detecting anomalies, and improving outcomes for pregnant women and their babies. The usefulness of machine learning in navigating the web of maternal health concerns can be seen through years of careful research by medical professionals and computer science experts. This will enable more focused and efficient services that put the health of pregnant women and their babies first.

Chapter 3

METHODOLOGY

This chapter discusses the project design, which includes various structures and approaches, the model's development process, model training and testing, operation, and evaluation procedures. Furthermore, previous methods from other research and literature related to the topic are regarded as an essential guide for the entire process.

Project Design

This study adopted the Machine Learning Building Process as it fits the research objectives and the most common data mining approach. Figure 3 shows the project development cycle of the ML Process.

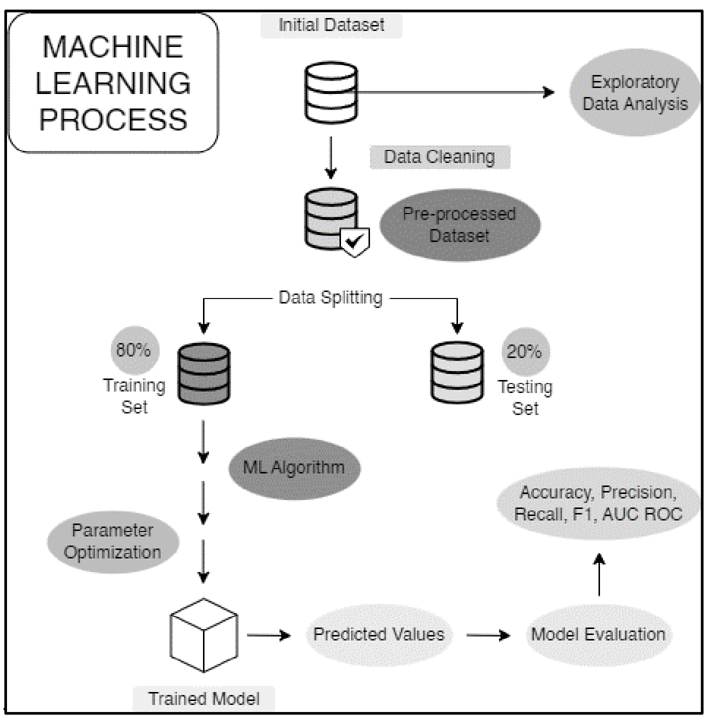

Figure 3. Machine Learning Building Process

The design phase began with the implementation of the suggested solution based on the identified problems and primary purpose. It started with the initial dataset and exploratory data analysis (EDA). The exploration of data unveiled its intrinsic characteristics, facilitating the identification of patterns, detection of outliers, exploration of correlations between variables, and identification of errors. Visualizations such as bivariate and multivariate scatterplots, charts, and heat maps were employed for EDA.

This is followed by the pre-processing stage encompassing data cleaning and preparation procedures. It included removing outliers using medical references and z-scores. The values under the RiskLevel column were changed to a numerical value to enable the use of a machine learning algorithm.

The modeling sections follow which include algorithm selection, model building, and hyperparameter tuning. The XGBoost classifier is the machine learning algorithm that was employed in this study. An 80:20 ratio was used to divide the processed dataset into training and testing sets. Initially, a basic XGBoost model was built to compare the standard approach method with GridSearchCV. Following comparison, the approach with the best outcome was applied. A GridSearchCV-based modified XGBoost classifier was employed.

If the accuracy of the used model is not met, the process will be repeated from the start. The model has the potential to fail during training and testing, which is why the process must be reevaluated with the help of this cycle to prepare for and prevent any potential difficulties that may develop later on.

Once the trained model has achieved a satisfactory degree of accuracy and precision, the procedure will proceed with the implementation of the algorithm for validation using new data. The findings of other studies will serve as the foundation for establishing the basis for determining the accuracy and precision of the trained model. Below are the steps that must be fulfilled to accomplish this study:

Exploratory Data Analysis

The initial phase of the ML Process methodology involves exploratory data analysis (EDA), which entails thoroughly examining and analyzing a dataset through visualizations. The primary goal is to comprehensively understand and gain insights into the dataset. This research concentrates on detecting and predicting the risk levels associated with pregnant women.

Data Collection

The dataset that will be used in this study is obtained from Kaggle which was used in the study of M. Ahmed and M. A. Kashem (2020) entitled “IoT Based Risk Level Prediction Model for Maternal Health Care In The Context Of Bangladesh”. According to them, the dataset was created by gathering health information from different hospitals in Dhaka and maternity clinics in Khulna, Bangladesh recorded in IoT devices, Web portals, and hospitals and then stored on a local server and cloud server in CSV format. The Internet of Things (IoT) was used for its capabilities to transfer data over a network without requiring human-to-human or human-to-computer interaction and long-distance data transfer Amala, S. & Mythili, S. (2018). The dataset contains the records of different pregnant women including the Age, SystolicBP, DiastolicBP, BS, BodyTemp, HeartRate, and RiskLevel. Each health information contributes to maternal health risk factors. The first 6 features are independent variables and the RiskLevel is the only dependent variable.

The dataset contains a total of 1014 records, of which 406 were classed as low-risk levels, 336 as mid-risk levels, and 272 as high-risk levels, each record representing a pregnant woman. It consists of 7 attributes or variables, making it suitable for benchmarking the performance of different algorithms when addressing similar problems. Additionally, the dataset can be used for training purposes in the field of machine learning.

Data Description

The dataset shown in Table 1 consists of maternal health risk factors. The sources of information are as follows:

1. Hospitals
2. Maternity Clinics

Table 1.
Hospitals and Maternal Clinics Table

SN#	Hospital Name	Address and Contact	Date
1.	Maternity Clinic	Chuadanga, Khulna, Bangladesh	26-01-2018 27-01-2018 03-04-2020
2.	Kurmitola General Hospital	Tongi Diversion Rd, Dhaka 1206	11-02-2019 12-02-2019
3.	Aichi Medical College & Hospital	Plot- 35 & 37, Sector-08, Abdullahpur Mohasorok, Dhaka 1230	12-02-2019 14-02-2019
4.	Uttara Adhunik Medical College Hospital (BMSRI)	H No 34, Road No.4 Janapath Sonargaon, Road, Dhaka 1230	18-02-2019
5.	East-West Medical College & Hospital	Aichi Nagar, JBCS Sarani P.O Khayartek, Horirampur, Dhaka 1711	18-06-2019 12-02-2020
6	CARe Medical College & Hospital	Dhaka, Bangladesh	18-06-2020

The data is stored in a comma-separated values (CSV) file and pertains to health institution records in the rural areas of Dhaka, Bangladesh using IoT (Internet of Things) based risk monitoring systems from 2018-2020. There are 1014 records of pregnant women in total, each with 7 attributes, and there are no missing values. Figure 4 shows the distribution of data based on the risk levels.

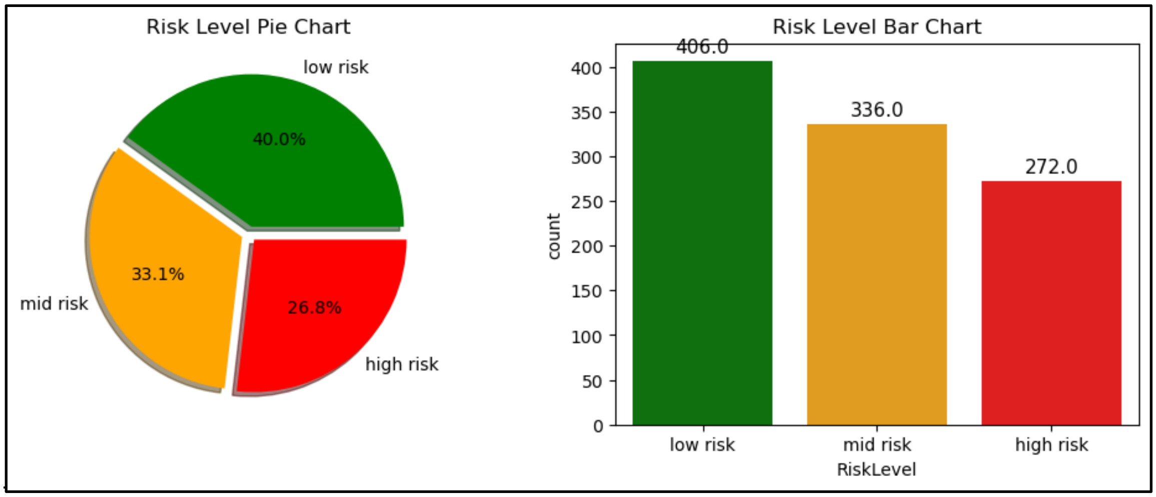

Figure 4. Distribution of different health risk levels

Table 2 describes each variable or feature used in the study. This includes the description, data type, and unit of measurement.

Table 2.
Variable Table

Variable Name	**Role**	**Type**	**Demographic**	**Description**	**Units**	**Missing Values**
Age	Feature	Integer	Age	Period of human life measured by years from birth	Years	no
SystolicBP	Feature	Integer		Upper value of blood pressure in mmHg, another significant attribute during pregnancy	mmHg	no
DiastolicBP	Feature	Integer		Lower value of blood pressure in mmHg, another significant attribute during pregnancy	mmHg	no
BS	Feature	Integer		Blood glucose levels are in terms of a molar concentration	mmol/L	no
BodyTemp	Feature	Integer		Fahrenheit	F	no
HeartRate	Feature	Integer		A normal resting heart rate	bpm	no
RiskLevel	Target	Categorical		Predicted Risk Intensity Level during pregnancy considering the previous attribute		no

Table 3 shows the different ranges of each feature that tell if a record is at risk or no risk (Ahmed et al., 2020; Cleveland Clinic, n.d.).

Table 3.
Pregnancy-related medical parameters with corresponding values and their weights

Hyperparameter	No Risk	At Risk
Age	20-29	30-45 or ≤17
SystolicBP	less than or equal to 120 mmHg	120 mmHg or greater
DiastolicBP	less than or equal to 80 mmHg	80 mmHg or greater
BS	<7.8 (<140) mmol/l(mg/dl)	≥7.8(≥2140) mmol/l(mg/dl)
BodyTemp	averages about 98.6F (37C)	98.7F (37C) or higher And (>35C or >95F) = Hypothermia
HeartRate	Heartbeat 75-80 bpm	<70 and >90bpm

Data Preprocessing

Data Cleaning

In the data cleaning process, steps were taken to ensure the integrity and quality of the dataset. Firstly, any implausible or inconsistent data points were systematically removed to maintain the reliability of the information. Subsequently, to align the dataset with the research objective, the target values of "low risk" were transformed to "0" denoting no risk, while "mid risk" and "high risk" were converted to "1" indicating at risk. Furthermore, the dataset underwent comprehensive outlier detection utilizing the z-score method, facilitating the identification and elimination of outliers that could potentially impact the accuracy of the model.

Feature Selection

After Cleaning the data, the feature selection procedure was employed using the embedded method to ensure the model's efficiency and effectiveness. The initial dataset comprised six key features: Age, Diastolic Blood Pressure (DiastolicBP), Systolic Blood

Pressure (SystolicBP), Blood Glucose (BS), Body Temperature (BodyTemp), and Heart Rate (HeartRate). Each of these features was carefully chosen based on their clinical relevance and potential influence on maternal health according to various studies.

Modeling

Algorithm Selection

XGBoost is one of the most popular in solving prediction problems, it is used in many Kaggle machine learning competitions as it is highly flexible, portable, and efficient. It has also been used in several studies in predicting in terms of maternal health risks which typically performs significantly better than the rest of the algorithms used in each study. Additionally, the researchers have tested the most common algorithms used such as Random Forest, MLP, SVC, Decision Trees, and KNN. The result is that XGBoost is the better predictor for this study. XGBoost achieved an accuracy of 92% and an AUC-ROC score of 97%, followed by Random Forest with an accuracy of 92% and an AUC-ROC score of 96%. The Decision Tree model attained an accuracy of 91%, while KNN, MLP, and SVC achieved accuracies of 79%, 65%, and 64%.

Its process is not that far from Gradient boosting but it's far better because Gradient boosting is prone to overfitting and XGBoost is not, it uses advanced regularization which is a method to control over-fitting which gives it better performance.

The following steps are how XGBoost works:

1. First, enter the data. The researchers will pick an example of a simple dataset, as shown in Table 4, that will be used in this study to show and simulate how XGBoost works.

Table 4.
Sample Dataset.

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
25	130	80	15	98	86	1
30	120	85	7	98	70	0
29	90	70	8	100	80	1
35	140	90	13	98	70	1
31	122	85	7.2	99	70	0
32	124	85	7	96	70	0

There are 6 sample people here in this dataset. They all have a class that they belong to, which is at risk = 1 and not at risk = 0. There are features here such as Age, SystolicBP, DiastolicBP, BS, BodyTemp, HeartRate, and RiskLevel.

2. Creating an initial base model starts with a weak learner in this case decision tree as a base model, typically a shallow tree with few levels. Note that the base model is always just a leaf. Since this is a binary classification, it will only have two classes to belong to which is At Risk or Not at Risk. The initial prediction probability of the base model is 0.5.
3. Compute the residuals: Calculate the residuals, which are the differences between the actual values and the probability predictions given by the current model. The formula for residual is = (actual value – prediction probability). Table 5 shows the individual record and their residual values.

Table 5.
Sample Dataset with residuals.

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel	Residuals
25	130	80	15	98	86	1	0.5
30	120	85	7	98	70	0	-0.5
29	90	70	8	100	80	1	0.5
35	140	90	13	98	70	1	0.5
31	122	85	7.2	99	70	0	-0.5
32	124	85	7	96	70	0	-0.5

4. After getting the residuals XGBoost will now construct a tree. For the example it will take the SystolicBP as the root node and then using the residuals it is now able to compute the similarity score using the formula below.

$$Similarity\ score = \frac{(\sum residuals)^2}{\sum_{i=1}^{N}[P(1-P)] + \lambda}$$

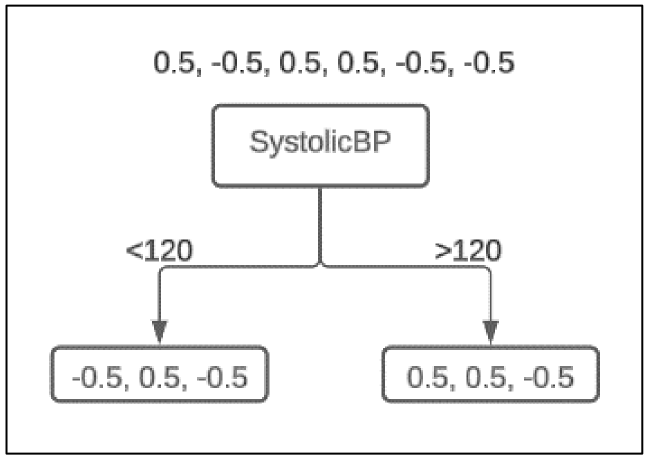

Figure 5. Initial Decision Tree

Using the formula shown in Figure 5, the similarity score on the left leaf will be 0.33 the similarity score of the right leaf will be 0.33 and lastly the similarity score of the root node is 0. Now it will need to get the **Gain (**similarity score (left node) + similarity score (right node) - similarity score (root node)) which will be 0.66. P is equal to the predicted probability and will change for each iteration, the higher the prediction probability the higher the accuracy of the model is. Then the same process will be reiterated for the other features and then whichever has the highest **Gain,** XGBoost will use that split. That is how XGBoost constructs a tree. **Pruning** is where it limits the depth of the decision tree. It is computed with the formula (P(1-P)) which equals to cover value which would have a value of 0.25. With this it can compare the branch to the cover value, if it's less than 0.25 the branch will be cut off because it is likely that the decisions in that branch have no predicting power to add.

5. Computing the prediction probability. For testing input data, the researchers will be taking the data from the first row. First is to find the output of the base model with respect to the probability using the formula below:

   P = the prediction probability.

$$\log(odds) = \log\left(\frac{P}{1-P}\right)$$

   Following the formula above the researchers will get the output from the base model as log(1) which is 0 and then fill in all the necessary data to the formula below.

$$\sigma\big(output\ of\ initial\ guess\ from\ previous\ step + learning\ rate(similarity\ score\ of\ current\ last\ node)\big)$$

$$\sigma(0 + 0.1(0.33))$$

   The learning rate can be between 0 and 1 it is dependent on the user which learning rate to use for this example the researchers will use 0.1 as the learning rate and then 0.33 because it is the similarity score of the first row's leaf node which is 130 SystolicBP which will be the right leaf node. Then this would result with $\sigma(0.033)$ The answer to this will become the new prediction probability for the first person or the inputted data. To know other prediction probabilities for the other entry, use the formula above.

6. Iteration of process. Now that new prediction probabilities are present, XGBoost will repeat the process from 3-6 until the stopping criterion is met. Then the computation would look something like the equation below as this is where the

final prediction will be. The weighted summation of all the predictions made by the trees which will be the output of each tree (T_n) and then multiplied by a learning rate (α) and then added to the initial prediction:

$$\sigma(0 + \alpha(T_1) + \alpha(T_2) \dots \alpha(T_n)$$

7. At the end it should give a lower residual the closer it is to 0 the better and it should have a good prediction probability value that will make a better prediction. Note that the sign λ or **lambda** has been set to 0 throughout the example. Lambda is a regularization technique used to reduce overfitting in the model. It will be counted as a hyperparameter, and it is advised to tune the value and experiment with it until the necessary results are achieved.

Model Training

Model training is a determining step in developing machine learning models. This involves teaching the model to recognize patterns and make accurate predictions. The researchers utilized the dataset and divided it into two sets. One is for the training set and the other is for the testing set (80:20). The selection of the training and testing dataset is randomized as shown in Figure 6.

```
X_proc = data_proc.drop("Risk", axis=1)
y_proc = data_proc.Risk
x_train_proc, x_test_proc, y_train_proc, y_test_proc = split(X_proc, y_proc, train_size=0.8, test_size=0.2, random_state=1)
```

Figure 6. Training and Testing split

With this, it allows the researchers to assess the generalization capabilities of the trained model. The training dataset serves as the foundation of the model's development process. The researchers opted to utilize Jupyter Notebook, a widely used open-source web application known for its interactive environment in code execution, data visualization, and result analysis. This choice proved ideal for conducting the model training experiment.

Jupyter Notebook supports multiple programming languages, including Python, and its extensive functionalities allow for the utilization of various libraries and frameworks, thereby expediting the model's development process. Access to these resources maximizes the effectiveness and predictive capabilities of the model.

Hyperparameter Tuning

Hyperparameters are specific variables or weights that control how an algorithm learns. In tree-based models, the learnable parameters are the choice of decision variables at each node. XGBoost is well known for its ability to automatically tune hundreds of learnable parameters to identify patterns and regularities in the data. Because of this, there will be more design choices and consequently more hyperparameters. Table 6 shows some of the main important hyperparameters used by the XGBoost algorithm.

Table 6.
XGBoost Main Hyperparameters

Hyperparameter	Description
learning_rate	It controls the step size or shrinkage applied at each boosting iteration.
gamma	Minimum loss reduction is required to make a further partition on a leaf node of the tree.
max_depth	It determines the maximum depth of each tree in the boosting process.
subsample	It determines the fraction of the sample to be randomly selected for each boosting iteration.
colsample_bytree, colsample_bylevel, colsample_bynode	A family of parameters for subsampling of columns. Bytree is the subsample ratio of columns when constructing each tree. Bylevel is the subsample ratio of columns for each level. Bynode is the subsample ratio of columns for each node (split).
reg_alpha and reg_lambda	Both are responsible for controlling the L1 and L2 regularization process. It prevents overfitting and creates simple models by adding penalties to the loss function.
tree_method	The tree construction algorithm used in XGBoost.
scale_pos_weight	Control the balance of positive and negative weights, useful for unbalanced classes.

n_estimators	It specifies the number of boosting iterations of trees to build.
objective	The objective parameter is the loss function to be minimized.
eval_metric	The eval_metric parameter is the metric used for monitoring performance during training and for early stopping.

Evaluation

Model Testing and Evaluation

Employing the trained model and the predict() function, predictions will be generated for the testing set. Through a comparison of these predicted values with the actual values from the testing set, the researchers will analyze the model's performance. Utilizing functions from scikit-learn, evaluation metrics including accuracy, confusion matrix, precision, recall, F1-score, and AUC-ROC will be computed. The evaluation results will be thoroughly examined to assess the model's performance and identify areas for further enhancement. Subsequent adjustments to the model will be made iteratively until the highest level of accuracy is attained.

Project Development

This phase discusses the requirements that must be met to build the project's functionality, together with the specifics for its execution. This contains a list of tools that will be used to build the model and the software that will be used to test it. The Python programming language and Microsoft Excel were the only tools utilized during the whole development process. Furthermore, the availability of numerous machine learning libraries has proven to be relevant. Tables 7 and 8 show all the hardware and software specifications used in the development of the model.

Table 7.
Hardware Specifications

Components	Descriptions
x86 64-bit CPU (Intel / AMD architecture)	The minimum requirement to run any Python IDE.
Minimum of 4 GB RAM	Recommended RAM is 8 GB needed to run Python and ensure that the IDE won't crash when building the model.
5 GB Free Disk Space (HDD/SSD)	This pertains to the installation of the Python IDE and other libraries used, as well as handling the raw data files.

Table 8.
Software Specifications

Components	Model Development
Operating System (OS)	At least Windows 7 / Mac OS X 10.11
Programming Language	Python 3.3 up to latest
IDE	Jupyter Notebook

Additionally, Table 9 shows the following Python libraries used in predicting maternal health risk.

Table 9.
Python Libraries

Libraries	Descriptions
Scikit-Learn	It provides the necessary machine learning tools such as data splitting and model evaluation.
XGBoost	Used for building the model.
NumPy	Used to perform a wide variety of mathematical operations on arrays.
Pandas	Allows us to analyze and manipulate data
Matplotlib	It offers a thorough library that may be used to build static, animated, and interactive visualizations for better comprehension.
Seaborn	Gives users a high-level interface for creating visually appealing and educational statistical visuals.
Yellowbrick	Extends the Scikit-Learn API to make model selection and hyperparameter tuning easier. Under the hood, it's using Matplotlib.

Operation and Testing Procedure

First, the researchers will start by preprocessing the data and then load the dataset containing Maternal Health information in CSV format into Python and perform necessary cleaning steps, such as handling missing values, outliers, and inconsistencies. Split the dataset into features (X) which are the demographic and health factors, and the target variable (y), the target variable will be the classes 0 (not at risk) and 1 (at risk). Use the train_test_split function in scikit-learn to divide the dataset into training and testing set. Then move on to the training of the XGBoost model. The researchers will import the XGBoost library as xgb. Then the researchers will create an instance of the XGBoost classifier (xgb.XGBClassifier) and set the necessary combination hyperparameters. These hyperparameters will include learning rate, number of estimators, maximum depth, etc. apply the XGBoost model to the training data using the fit method.

Once the model is trained, the model's performance will be evaluated using the testing set. Using the trained model to make predictions on the testing set and calculate evaluation metrics to assess its performance. An example of this is calculating the accuracy using the accuracy_score function from scikit-learn and other evaluation metrics will be applied as well like precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

To interpret the XGBoost model, retrieve the feature importance scores using the feature_importances_ attribute. The researchers will analyze these scores to identify the most influential features for predicting maternal health risk. Then the researchers will perform hyperparameter tuning until satisfied with the result. The researchers will also utilize techniques like grid search so that the researchers can explore different combinations of hyperparameters to achieve an accuracy that the researchers will be satisfied with. This resulted in these two experiments:

1. **GridSearchCV** – It is a systematic approach that exhaustively searches through a predefined set of hyperparameter combinations to find the best performing combination. It involves specifying a grid of hyperparameter values for each hyperparameter of the model. The grid search algorithm then uses cross-validation or a different validation set to assess the model's performance for each set of hyperparameters. The combination that achieves the highest performance metric will be selected as the optimal set of hyperparameters for the model. Figure 7 shows the process of GridSearchCV in Python.

```
param_grid = {
    'learning_rate': [0.01, 0.1, 0.2],
    'max_depth': [3, 4, 5],
    'n_estimators': [100, 500, 1000],
    'subsample': [0.5, 0.7, 0.9],
    'colsample_bytree': [0.5, 0.7, 0.9],
    'gamma': [0, 0.1, 0.2],
    'reg_alpha': [1, 0.1, 0.01]
    'reg_lambda': [1, 0.1, 0.01]
}

grid_search = GridSearchCV(estimator=xgb_classifier4, param_grid=param_grid, cv=3, n_jobs=-1, verbose=1)
grid_search.fit(X_train, y_train)
```

Figure 7. Image of GridSearchCV Coding

2. **Standard Approach** – This method uses the default hyperparameters of the classifier. This means that the researchers did not set any hyperparameters for the training and testing of the model. This is because it is appropriate to take into account the existence of a small number of predetermined factors. With this, it is possible for the performance of the model to be accurate and on par with GridSearchCV. The only difference is that setting up the hyperparameters was not included. However, the process still involves separating the data and cross validating the model. Figure 8 shows the implementation of the standard approach of the XGBoost Classifier in Python.

```
In [47]: # Calling the model
         xgb_classifier = xgb.XGBClassifier(random_state=42)

         # Fit the training data to the model
         xgb_classifier.fit(x_train_proc, y_train_proc, verbose=0)

         # Predict test values
         y_pred_proc = xgb_classifier.predict(x_test_proc)

         # Applying k-Fold Cross Validation
         accuracies = cross_val_score(estimator = xgb_classifier, X = x_train_proc, y = y_train_proc, cv = 10)
         print("Mean Accuracy: ", accuracies.mean())
         print("Standard Deviation of Accuracy: ", accuracies.std())

         # Print the accuracy and the classification report of the model
         print(f"Accuracy: {accuracy_score(y_test_proc, y_pred_proc)}")
         print(classification_report(y_test_proc, y_pred_proc))
```

Figure 8. Image of Standard Approach Coding

By employing cross-validation, the researchers will assess the performance of the model. The best set of hyperparameters will be chosen by the researchers to create a more accurate model up till they are satisfied with the outcomes.

Evaluation Procedure

Once the trained model has been generated with the use of a data set and XGBoost algorithm, it will undergo the process of evaluation. To generate an accurate and reliable model, the researchers will test the model by taking the total number of patients whose Risk Level is low. The model with the best degree of performance and accuracy will be able to be identified after a sequence of simulation tests on the models. A software system for the evaluation technique will subsequently be created using the implemented and included successful model.

The researchers will perform a multi-class confusion matrix to evaluate the performance and accuracy of the trained model. True Positive (TP) is acquired when the true value “high risk” is predicted as high risk. True Negative (TN) is when a true value “low risk” is predicted as low risk. On the other hand, a False Positive (FP) is when a true value “low risk” is wrongly predicted as high risk or mid risk. False Negative (FN) is when a true value “high risk” is wrongly predicted as low risk or mid risk.

The idea of an accuracy score is the same for both binary and multi-class classification, while the methods used to compute it vary slightly. In multi-class classification, there are two main ways to calculate accuracy score; macro-averaged accuracy and micro-averaged accuracy. Macro-averaged accuracy is calculated by calculating the accuracy for each class and then averaging the results, while Micro-averaged accuracy is calculated by dividing the total number of correct predictions by the total number of predictions. Accuracy can be a less informative measure of performance, especially if the classes are not evenly distributed (Evidently AI, n.d.).

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision indicates the accuracy rate of a model in predicting the intended class. The Precision of a label is calculated by dividing the total number of true positive results by the total number of predicted positive results. The precision score in multi-class classification can be computed using a micro-averaged or macro-averaged method for each class separately or for the entire classifier. Micro-averaged precision is calculated by dividing the total number of true positives by the total number of predicted positives.

$$\boldsymbol{Precision} = \frac{TP}{TP + FP}$$

$$\boldsymbol{Macro - averaged\ Precision} = \frac{sum\ (Precision)}{number\ of\ classes}$$

Recall is the ratio of true positives to the total number of actual positives for a given class. It is a valuable metric when false positives are less concerning than false negatives. In medical cases, it is often acceptable to sound a false alarm because it is crucial to identify all real positive cases (Agrawal, 2023). Recall scores for multiple classes can be determined using a macro- or micro-averaged method, either for the classifier as a whole or for each class separately. The micro-averaged recall is calculated by dividing the total number of true positives by the total number of actual positives. Macro-averaged recall is calculated by averaging the recall scores for each class.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN}$$

$$\boldsymbol{Macro - averaged\ Recall} = \frac{sum\ (Recall)}{number\ of\ classes}$$

The F1 score is the harmonic mean of precision and recall. F1 score can be calculated for each class individually or for the entire classifier using a micro-averaged or macro-averaged approach. Micro-averaged F1 score is calculated by averaging the F1 scores for each class. Macro-averaged F1 score is calculated by averaging the precision and recall scores for each class and then calculating the F1 score from the averaged precision and recall scores.

$$\boldsymbol{F1\ Score} = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

$$\boldsymbol{Macro - averaged\ F1\ Score} = \frac{sum\ (F1\ Scores)}{number\ of\ classes}$$

Another popular and primary method used for diagnostic tests called the Receiver Operating Characteristic (ROC) Curve, will be used as an evaluation metric. A ROC curve represents the relationship between a diagnostic test's sensitivity and 1 − specificity. It plots the true positive rate (TPR) versus the false positive rate (FPR) for different thresholds of classification scores. Various cutpoints are employed to ascertain whether test results are positive, and these correspond to the various points on the curve. An individual's average test sensitivity across all possible specificity values, or vice versa, can be represented by a ROC curve. A broader interpretation is that the test results indicate the likelihood that, of two patients chosen at random, one will have the disease or condition and the other will not and that the patient with the disease or condition will have a result indicating greater suspicion (Hanley & McNeil, 1982).

$$\boldsymbol{Sensitivity \mid True\ Positive\ Rate \mid Recall} = \frac{TP}{TP + FN}$$

$$\boldsymbol{False\ Negative\ Rate} = \frac{FN}{TP + FN}$$

$$\boldsymbol{Specificity \mid True\ Negative\ Rate} = \frac{TN}{TN + FP}$$

$$\boldsymbol{False\ Positive\ Rate} = \frac{FP}{TN + FP} = 1 - Specificity$$

The Area under the ROC Curve (AUC) is a useful metric for condensing the test's total diagnostic accuracy. It accepts values between 0 and 1, where 0 represents a test that is perfectly inaccurate and 1 represents a perfectly accurate test. AUCs of 0.5 generally indicate no discrimination (i.e., the ability to diagnose patients with and without the disease or condition based on the test); values between 0.7 and 0.8 are regarded as acceptable; values between 0.8 and 0.9 are regarded as excellent; and values exceeding 0.9 are regarded as remarkable (Nahm, 2022).

In binary classification, where the TPR and FPR can be clearly defined, ROC curves are commonly employed. When multiclass classification is used, binarizing the output is the first step towards obtaining an idea of TPR or FPR. The researchers applied the One-vs-Rest method. The OvR is a technique for evaluating multiclass models that compare each class simultaneously to all other classes. In this case, the researchers designate one class as the "positive" class and refer to the remaining classes (the rest) as the "negative" classes.

For balanced datasets, metrics like accuracy, precision, and recall are useful for assessing classification models; however, in cases when the data is unbalanced, alternative techniques such as AUC-ROC yield superior evaluations of the model's performance (Agrawal, 2023).

Furthermore, the finished program will be tested by a data science/statistics professional who will make major comments, suggestions, and recommendations on the methodology employed and the general application of the model to the software.

Chapter 4

RESULTS AND DISCUSSION

This chapter presents and explains this study’s findings. It comprises the project's capabilities and limitations, description, structure, and evaluation. In accordance with the project's goals, it offers visualization and discussion through computations, figures, and presentation tables.

Project Description

This project focuses on utilizing the XGBoost classifier to predict whether a pregnant woman is at risk. Maternal health is a critical concern, and timely risk assessment can significantly impact outcomes for both mothers and their babies. The project involves collecting a comprehensive dataset with key maternal health indicators, including age, blood glucose, blood pressure, heart rate, and body temperature. After data preprocessing and feature selection, an XGBoost classifier will be trained and tuned to predict maternal health risk. The goal is to develop an accurate and interpretable model that can aid healthcare professionals in identifying pregnant women who are at risk, allowing for more targeted care and support during pregnancy. The project's success will be assessed through various evaluation metrics such as confusion matrix, f1 score, precision score, recall score and, ultimately, aims to contribute to improved maternal healthcare.

Project Structure

The study's structural components will be discussed in this section. Figure 9 illustrates the project structure of the study.

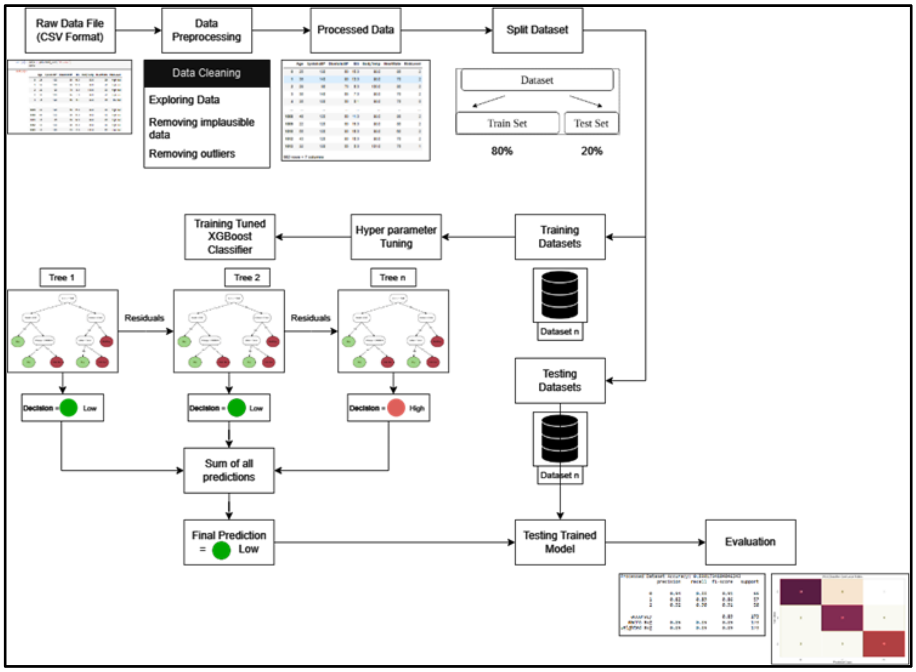

Figure 9. Project Structure

The initial step involved importing the dataset Following this, exploratory data analysis was conducted to understand the dataset's characteristics. During this phase, the implausible data was removed, for example, if a patient has a body temperature of greater than 100 and the risk level is low, which is not supposed to be as it should be at least mid or high-risk level. The outliers were also removed using the z score. If the standard deviation of any data point is less than -3 or greater than 3 was classified. After that the processed data was split into a training set and a testing set with a ratio of 80:20. The researchers allocated 80% to training and 20% to testing because it is commonly used in the study that the researchers have gathered and produced a great accuracy, and it focuses on training the model but leaving a good amount of data for testing its capabilities. The XGBoost model's hyperparameter was then tuned through GridsearchCV using the processed dataset, after generating the best hyperparameters to use, it was then applied to the model and made a prediction again using the processed data, producing results that were evaluated using various evaluation metrics such as accuracy, f1 score, precision score, recall score, and confusion matrix.

Project Capabilities and Limitations

This section provides the list of capabilities and limitations of the model:

Capabilities of the Model

1. The model can predict maternal health risk based on input data, given that the values inputted are in the correct unit/s measurement.
2. The model can show the possible leading causes and the normal values for each feature if the user is predicted to be at risk.

Limitation of the Model

1. The model does not provide recommended actions or possible treatments to the user.

Project Evaluation

Initial Testing Results

The quantity of information about women with mid and high risks is also crucial to the training and testing procedure. There should be an equal number of records for women with low, mid, and high risks. Since the dataset is imbalanced, there is more bias for low risk than the high-risk records. The total number of data used for training and testing is 1014. Then, it was split into an 80:20 ratio resulting in 811 records of training data and 203 records of testing data.

	precision	recall	f1-score	support
0	0.87	0.87	0.87	77
1	0.92	0.92	0.92	126
accuracy			0.90	203
macro avg	0.90	0.90	0.90	203
weighted avg	0.90	0.90	0.90	203

Figure 10. Initial Classification Report of the Trained Model using Standard Approach

	precision	recall	f1-score	support
0	0.88	0.88	0.88	77
1	0.93	0.93	0.93	126
accuracy			0.91	203
macro avg	0.91	0.91	0.91	203
weighted avg	0.91	0.91	0.91	203

Figure 11. Initial Classification Report of the Trained Model using GridSearchCV

Figures 10 and 11 showed that GridSearchCV produced better results with an accuracy of 91% than the Standard Approach's 90%. Though it's only a small improvement, it's worth noting that in the medical field, every improvement is significant.

Model Evaluation Results

The research presented its findings in this section after using the filtered dataset to evaluate the model. The investigators employed a multi-class classification and confusion matrix at this phase of the investigation to visually represent the results. According to the results of the classification report obtained from the initial testing, the GridSearchCV approach has the best performance, and this method was utilized for prediction. According to the first testing results, this study's methodology involved using sorted and cleaned data, which produced 862 records in total. Of those records, 173 records (or 20% of the total data) were used for the testing process utilizing the trained model. Table 10 shows the hyperparameters that were changed to improve the model, while Table 11 shows the default hyperparameters that could be changed to improve the model but weren't due to hardware limitations.

Table 10.
XGBoost Hyperparameters Set

Name	**Result**
objective	binary: logistic
learning_rate	0.1
max_depth	5
gamma	0.1
reg_alpha	0.1
reg_lambda	0.5
n_estimators	1000
subsample	0.7
colsample_bytree	0.9
random_state	1

Table 11.
XGBoost Default Hyperparameters Set

Name	**Result**
colsample_bylevel	None, default = 1
colsample_bynode	None, default = 1
max_delta_step	None, default = 0
max_leaves	None, default = 0
min_child_weight	None, default = 1
num_parallel_tree	None, default = 1
sampling_method	None, default = uniform
scale_pos_weight	None, default = 1

The summary of the predicted results presented by the confusion matrix is shown in Figure 12. Table 12 showed that the generated results for Not at Risk (0) are True Positive (TP) is 58, False Positive (FP) is 4, False Negative (FN) is 8, and True Negative (TN) is 103; And for At Risk (1), TP is 103, FP is 8, FN is 4, and TN is 58.

Table 12.
Predicted Results for Maternal Health Risk

	True Positive	False Positive	False Negative	True Negative	Total
0	58	4	8	103	173
1	103	8	4	58	173

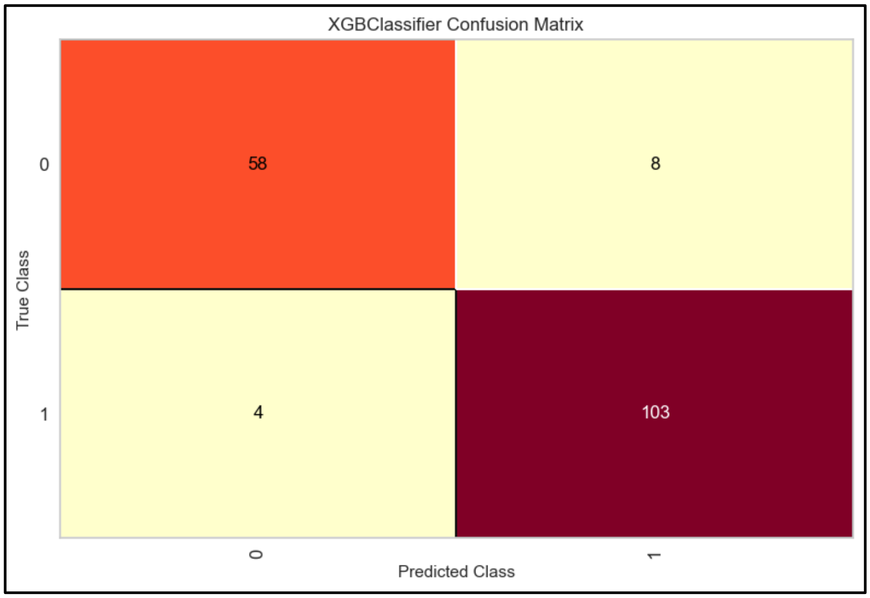

Figure 12. Confusion Matrix of Predicted Results

For displaying the trained model's accuracy, precision, recall, and F1 score, the classification matrix formula was used to compute each metric.

$$\boldsymbol{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{161}{173} = 0.9306 * 100 = \mathbf{93.06}\%$$

Class 0: Not At Risk

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} = \frac{58}{58 + 4} = \frac{58}{62} = 0.9355 * 100 = \mathbf{93.55}\%$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} = \frac{58}{58 + 8} = \frac{58}{66} = 0.8788 * 100 = \mathbf{87.88}\%$$

$$\boldsymbol{F1\ Score} = \frac{2 * (Precision * Recall)}{Precision + Recall} = \frac{2 * (0.9355 * 0.8788)}{0.9355 + 0.8788} = 0.9062 * 100$$

$$= \mathbf{90.62}\%$$

Class 1: At Risk

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} = \frac{103}{103 + 8} = \frac{103}{111} = 0.9279 * 100 = \mathbf{92.79}\%$$

$$\boldsymbol{Recall} = \frac{TP}{TP + FN} = \frac{103}{103 + 4} = \frac{103}{107} = 0.9626 * 100 = \mathbf{96.26}\%$$

$$\boldsymbol{F1\ Score} = \frac{2 * (Precision * Recall)}{Precision + Recall} = \frac{2 * (0.9279 * 0.9626)}{0.9279 + 0.9626} = 0.9449 * 100$$

$$= \mathbf{94.49}\%$$

Macro-averaging results:

$$\boldsymbol{Precision} = \frac{sum\ (Precision)}{number\ of\ classes} = \frac{0.9355 + 0.9279}{2} = \frac{1.8634}{2} = 0.9317 * 100$$

$$= \mathbf{93.17}\%$$

$$\boldsymbol{Recall} = \frac{sum\ (Recall)}{number\ of\ classes} = \frac{0.8788 + 0.9626}{2} = \frac{1.8414}{2} = 0.9207 * 100$$

$$= \mathbf{92.07}\%$$

$$\boldsymbol{F1\ Score} = \frac{sum\ (F1\ Scores)}{number\ of\ classes} = \frac{0.9062 + 0.9449}{2} = \frac{1.8511}{2} = 0.9256 * 100$$

$$= \mathbf{92.56}\%$$

The summary of the classification report of the XGBoost model is shown in Figure 13. This includes all the micro- and macro-averaged metrics together with the support.

	precision	recall	f1-score	support
0	0.94	0.88	0.91	66
1	0.93	0.96	0.94	107
accuracy			0.93	173
macro avg	0.93	0.92	0.93	173
weighted avg	0.93	0.93	0.93	173

Figure 13. Classification Report of Modified XGBoost Model

Figure 14 illustrates the ROC Curves of the XGBoost Model for all classes with the micro- and macro-averaged AUC.

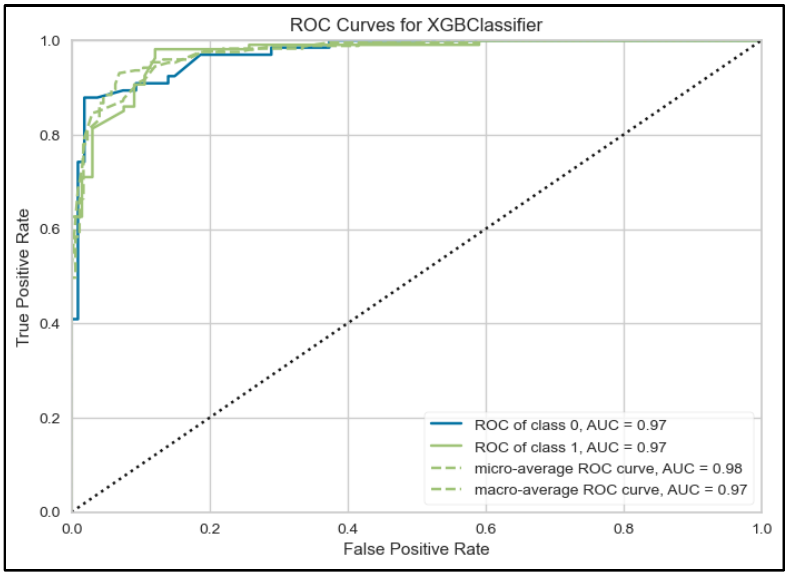

Figure 14. ROC Curves for Modified XGBoost Model

Additionally, the important features chosen by the XGBoost Classifier are plotted in Figure 15, with Age being the most important and BodyTemp the least important.

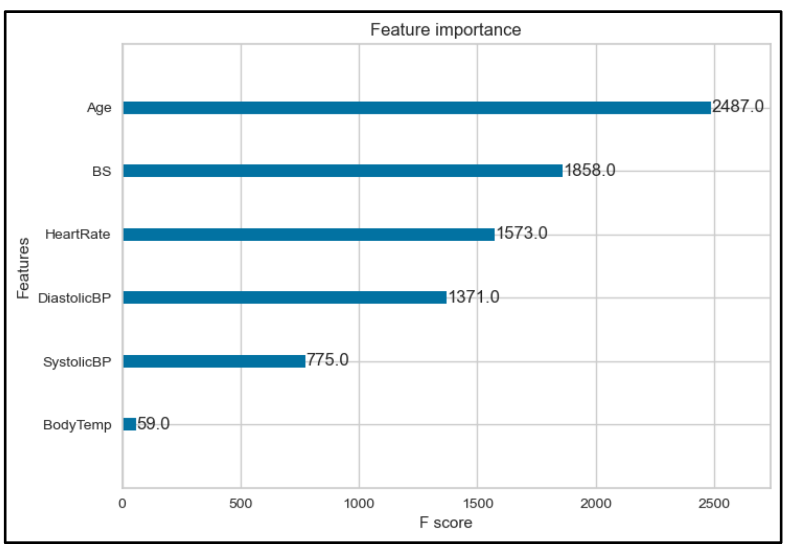

Figure 15. Important Features Based on the XGBoost Model

Predicting

Upon testing the model by inputting the researchers' values in each variable it was able to predict correctly. Figure 18 shows the Python code for the predicting part of the model. The researchers put data such as shown in Figure 16. These are all normal values for each variable so this should result as "0" or not at risk which the model was able to predict.

```
Enter values for features:
Age: 25
Systolic Blood Pressure: 90
Diastolic Blood Pressure: 70
Blood Sugar (mmol/l): 6.1
Body Temperature (°F): 98
Heart Rate: 80

Results:

You are not at risk.
```

Figure 16. Testing of the model part 1

In Figure 17, the researchers have inputted values that might lead to the patient being at risk such as the age of 10 or a fairly high systolic blood pressure for a patient of that age, which should result as "1" or at risk which the model was able to predict and the model was able to provide the possible leading causes that placed the patient in that position.

```
Enter values for features:
Age: 10
Systolic Blood Pressure: 120
Diastolic Blood Pressure: 90
Blood Sugar (mmol/l): 6.1
Body Temperature (°F): 100
Heart Rate: 70

Results:

You are at risk.

These are the possible problems you are facing that contribute to your result:

- Your age is lower than the prefered age of pregnancy. The normal age in pregnancy is 20-29
- You have a high systolic BP. The normal is less than or equal to 120 mmHg
- You have a high diastolic BP. The normal is less than or equal to 80 mmHg
- Your temperature is high. The normal is 98.6°F or 37°C
```

Figure 17. Testing of the model part 2

These tests showcase how nurses can utilize this model to oversee the health indicators of pregnant women via IoT gadgets or regular medical assessments for proactive monitoring. Recognizing heightened variables is pivotal for accurately gauging their well-being and deciding the appropriate course of action. The similarity in symptoms between pregnancy-related ailments and common illnesses complicates diagnosis, posing challenges in pinpointing the underlying issue. Consequently, employing advanced technology like IoT devices and continuous medical tests aids in early detection and intervention. By leveraging these tools, healthcare professionals can enhance their ability to provide timely and effective care to pregnant individuals. This approach underscores the importance of preventive measures in managing maternal health.

Comparison with Existing Literature

The study conducted by Ahmed M. et al. (2020), titled "Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT)," demonstrated an impressive accuracy of 90% using a Logistic Model Tree (LMT). Their research focuses on predicting the risk level of maternal health. In contrast, the current study predicts whether a pregnant woman is at risk or not, employing XGBoost, and achieved an accuracy of 93%. Notably, unlike Ahmed's study, this research meticulously addressed the presence of implausible data within the dataset. This careful data curation prevents misleading insights and addresses ethical concerns, ensuring the reliability of predictions in real-world scenarios.

In a related study by Raza A. et al. (2022), titled "Ensemble Learning-Based Feature Engineering to Analyze Maternal Health During Pregnancy and Health Risk Prediction," the authors utilized a BiLTCN approach and, similar to Ahmed's study, did not exclude implausible data. They reported a Precision, Recall, and F1 score of 84%, along with a micro-ROC of 97%. Additionally, they detailed class-specific performance metrics, indicating 95% for the low-risk class, 94% for the mid-risk class, and 99% for the high-risk class.

This study achieved a Precision, Recall, and F1 score of 93%, aligning closely with the results of Raza A. et al. The micro-ROC and macro-ROC also stood at 97%. Notably, this research has an AUC score of 97% for both the “not at risk” and “at risk” classes. As emphasized in Raza A. et al.'s study, a higher AUC ROC score signifies a superior classification model. Therefore, despite comparable micro-ROC values, the higher AUC scores in specific risk classes in this study underscore its robust performance in maternal health risk prediction.

Interpretation of Results

These results indicate that the model performs well in distinguishing between different risk levels, with a particular strength in predicting high-risk cases, which is crucial in clinical settings. The high precision values (94% and 93%) signify that the model makes few errors in classifying cases as non-risk when they are actually at risk. Similarly, the high recall values (88% and 96%) demonstrate the model's effectiveness in identifying all instances of each case. The balanced macro-average metrics further validate the model's robustness across both risk classes.

This suggests that the methodology used has produced a reliable predictive model for maternal health risk assessment. Such a model holds significant implications for risk management applications, providing a dependable basis for informed decision-making in situations where risks may vary.

In terms of medical studies, the Area under the Receiver Operating Characteristic Curve (AUC-ROC) serves as a metric to evaluate the accuracy of diagnostic tests. A high AUC indicates that the model excels in distinguishing between risk and non-risk scenarios. Therefore, the high AUC in this context suggests that the model is highly capable of accurately predicting both risk and non-risk outcomes, reinforcing its utility in clinical settings.

Chapter 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter is the conclusion of this study, it contains a discussion and a summary of the research and, this includes recommendations for future research/studies.

Summary of Findings

Based on the collected, examined, and evaluated data, the summary of the results are summed up as follows:

The XGBoost model identified several significant factors that contribute to assessing the risk of pregnant women. Among these factors, the age of the woman emerged as the most crucial feature, followed by blood sugar levels, heart rate, and both diastolic and systolic blood pressure. Interestingly, the model deemed body temperature as the least influential factor in predicting risk.

Employing a pre-determined XGBoost classifier model on a pre-processed dataset containing six key features yielded promising results. Utilizing the GridSearchCV approach optimized the model's performance, resulting in an impressive accuracy rate of 93.06% during testing.

Examining the model's performance further, it demonstrated commendable accuracy rates in both identifying individuals at risk and those not at risk. The true positive accuracy, indicating the model's ability to correctly predict risk when present, stood at 93.17%. Conversely, the true negative accuracy, representing the model's proficiency in identifying individuals not at risk, reached 96.26%.

Additionally, assessing the model's overall performance through macro-averaged metrics revealed consistent reliability. The model achieved a macro-averaged accuracy, precision, recall, and F1 score of 93.06%, 93.17%, 92.07%, and 92.56%, respectively. Furthermore, the macro-averaged area under the ROC curve (AUC-ROC) reached an impressive 97%, underscoring the model's ability to discriminate between individuals who are at risk and those who are not at risk effectively.

An additional significant capability of the model lies in its ability to offer insights to users identified as at risk, pinpointing potential leading causes based on the input data.

Conclusion

This study successfully achieved its objective of developing a predictive model for maternal health risk using the XGBoost algorithm. The comprehensive evaluation of the model demonstrated its efficacy in distinguishing between different risk levels with a notable strength in identifying high-risk pregnancies and presenting insights into the potential leading causes. The precision and recall values indicated minimal errors in classification and high sensitivity across all risk classes affirming the model’s reliability. The findings hold significant implications for maternal healthcare that provide a reliable tool for informed decision-making and risk management. In essence, the study contributes as a valuable asset to maternal care because it offers healthcare professionals the means to enhance risk level assessment that may help or potentially prevent adverse maternal health outcomes. Continuously improving and using the model can really help make maternal healthcare better which will essentially make expectant mothers and babies healthier.

Recommendations

These are the following recommendations for future studies.

1. More Dataset and Quality Dataset:

Having More Datasets and a better dataset quality will result in a more reliable prediction as this will aid in capturing complex patterns and improving its predictive capabilities. An example of better dataset quality is having more features on the dataset that can affect the result on the pregnant woman and baby's health such as BMI, and bad habits like smoking or drinking alcohol. Obesity in pregnancy can result in more pregnancy complications, such as the risk of miscarriage, fetal and congenital anomalies, thromboembolism, preeclampsia and gestational hypertension, fetal macrosomia, gestational diabetes mellitus, IUGR (intrauterine growth restriction), and stillbirth, as well as intrapartum and postpartum complications and neonatal mortality. Being underweight also provides problems such as Premature birth and low birth weight of the infants, these infants are at risk for health and development problems as they get older. Smoking during pregnancy leads to more problems in pregnancy such as miscarriage and premature labor. Furthermore, babies whose mothers are smoking during pregnancy are at a higher risk of sudden unexpected death in infancy (SUDI), having weaker lungs, and having an unhealthy low birth weight. Alcohol use during pregnancy can cause miscarriage, stillbirth, and a range of lifelong physical, behavioral, and intellectual disabilities. These disabilities are known as fetal alcohol spectrum disorders (FASDs).

2. Consider applying other advanced ensembling techniques:

Ensembling involves combining predictions from multiple models to improve overall performance. Stacking, combining multiple base models with a meta-model to make predictions. The meta-model is trained on the predictions of the base models and the meta-model combines the predictions of the base models into a final prediction. By leveraging the diversity of individual models, ensembling can enhance the robustness and generalization of the overall predictive capabilities of the model. This technique can result in improved accuracy and model stability.

REFERENCES

ACOG. (n.d.). *Preeclampsia and high blood pressure during pregnancy*. Retrieved November 11, 2023 from https://www.acog.org/womens-health/faqs/preeclampsia-and-high-blood-pressure-during-pregnancy.

Agrawal, S. K. (2023). Metrics to Evaluate your Classification Model to take the right decisions. Retrieved November 7, 2023 from https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/.

Ahmed, M., Kashem, M. A., Rahman, M., & Khatun, S. (2020b). Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT). In *Lecture notes in electrical engineering* (pp. 357–365). https://doi.org/10.1007/978-981-15-2317-5_30.

Akhan Akbulut, Egemen Ertugrul, Varol Topcu. (2018). Fetal health status prediction based on maternal clinical history using machine learning techniques. Computer Methods and Programs in Biomedicine, Volume 163, Pages 87-100, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2018.06.010.

Amala, S. & Mythili, S. (2018). IoT based health care monitoring system for rural pregnant women. *International Journal of Pure and Applied Mathematics*. 119. 837-843.

Arrieta, Eugenia & Estrada, Francisco & Caicedo, William & Martinez Santos, Juan Carlos. (2016). Early Prediction of Severe Maternal Morbidity Using Machine Learning Techniques. 10022. 259-270. DOI:10.1007/978-3-319-47955-2_22.

Bakouregui, A. S., Mohamed, H. M., Yahia, A., & Benmokrane, B. (2021). Explainable extreme gradient boosting tree-based prediction of load-carrying capacity of FRP-RC columns. *Engineering Structures,* 245, 112836. doi:10.1016/j.engstruct.2021.112836.

Bertini, A., Salas, R., Chabert, S., Sobrevia, L., & Pardo, F. (2022). Using machine learning to predict complications in pregnancy: A systematic review. *Frontiers in Bioengineering and Biotechnology, 9*. https://doi.org/10.3389/fbioe.2021.780389.

Betts, K., Kisely, S., & Alati, R. (2019). Predicting common maternal postpartum complications: leveraging health administrative data and machine learning. *BJOG: An International Journal of Obstetrics and Gynaecology*, *126*(6), 702–709. https://doi.org/10.1111/1471-0528.15607.

Bosschieter, T. M., Xu, Z., Liang, H., Lengerich, B. J., Nori, H., Sitcov, K., Souter, V., & Caruana, R. (2022). Using interpretable machine learning to predict maternal and fetal outcomes. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2207.05322.

Cagayan, M. S. F. S., Nisperos, G. A., Facun, G.-M. G., Cagayan, B. S. S., Castro, M. C. R., & Silverio, C. E. (2022). Mothers' Perspectives on Utilization of Maternal Health Services in Rural Health Units in Luzon: A Qualitative Study. Acta Medica Philippina, 56(16). https://doi.org/10.47895/amp.v56i16.5761.

Cil, G. and Cameron, T.A. (2017), Potential Climate Change Health Risks from Increases in Heat Waves: Abnormal Birth Outcomes and Adverse Maternal Health Conditions. Risk Analysis, 37: 2066-2079. https://doi.org/10.1111/risa.12767.

Di Napoli, C., Pappenberger, F., & Cloke, H. L. (2019). Verification of Heat Stress Thresholds for a Health-Based Heat-Wave Definition. *Journal of Applied Meteorology and Climatology*, 58(6), 1177-1194. https://doi.org/10.1175/JAMC-D-18-0246.1.

Eunice Kennedy Shriver U.S. Department of Health and Human Services. (n.d.). What are some factors that make a pregnancy high risk?. *National Institute of Child Health and Human Development*. https://www.nichd.nih.gov/health/topics/high-risk/conditioninfo/factors.

Evidently AI. (n.d.). Accuracy, precision, and recall in multi-class classification. Retrieved November 7, 2023 from https://www.evidentlyai.com/classification-metrics/multi-class-metrics.

Finlayson K, Crossland N, Bonet M, Downe S (2020). What matters to women in the postnatal period: A meta-synthesis of qualitative studies. *PLoS ONE* 15(4): e0231415. https://doi.org/10.1371/journal.pone.0231415.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Hu, X., Hu, X., Yu, Y., & Wang, J. (2023). Prediction model for gestational diabetes mellitus using the XG boost machine learning algorithm. *Frontiers in Endocrinology, 14*. https://doi.org/10.3389/fendo.2023.1105062.

Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). Bagging and Boosting Ensemble Classifiers for Classification of

Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. *Remote Sensing*, *13*(21), 4405. https://doi.org/10.3390/rs13214405.

Khandelwal, N. (2020, July 16). A Brief Introduction to XGBoost. *Medium*. https://towardsdatascience.com/a-brief-introduction-to-xgboost-3eaee2e3e5d6.

Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). An hybrid approach for prediction based health monitoring in pregnant women. *Procedia Technology*, *24*, 1635–1642. https://doi.org/10.1016/j.protcy.2016.05.171.

Lee-Brago, P. (2023, May 14). UN: Philippines maternal deaths on the rise. Philstar.com. https://www.philstar.com/headlines/2023/05/15/2266392/un-philippines-maternal-deaths-rise.

Li, Y. X., Shen, X. P., Yang, C., Cao, Z. Z., Du, R., Yu, M. D., Wang, J. P., & Wang, M. (2021). Novelelectronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms. Pregnancy hypertension, 26, 102–109. https://doi.org/10.1016/j.preghy.2021.10.006.

Linn B. Strand, Adrian G. Barnett, Shilu Tong (2011). The influence of season and ambient temperature on birth outcomes: A review of the epidemiological literature, Environmental Research, Volume 111, Issue 3, Pages 451-462, ISSN 0013-9351, https://doi.org/10.1016/j.envres.2011.01.023.

Liu, H., Li, J., Leng, J., Wang, H., Liu, J., Li, W., Liu, H., Wang, S., Ma, J., Chan, J. C., Yu, Z., Hu, G., Li, C., & Yang, X. (2020). Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China. *Diabetes/Metabolism Research and Reviews, 37*(5). https://doi.org/10.1002/dmrr.3397.

M. Ahmed and M. A. Kashem (2020). "IoT Based Risk Level Prediction Model For Maternal Health Care In The Context Of Bangladesh," 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2020, pp. 1-6, DOI: 10.1109/STI50764.2020.9350320.

M. Assaduzzaman, A. A. Mamun and M. Z. Hasan (2023). "Early Prediction of Maternal Health Risk Factors Using Machine Learning Techniques," 2023 International Conference for Advancement in Technology (ICONAT), Goa, India, pp. 1-6, DOI: 10.1109/ICONAT57137.2023.10080700.

Maternal health risk data. (2021, December 21). Kaggle. https://www.kaggle.com/datasets/csafrit2/maternal-health-risk-data.

Maternal risk factors. Maternal Risk Factors | Texas DSHS. (n.d.). https://www.dshs.texas.gov/maternal-child-health/programs-activities-maternal-child-health/hear-her-texas/maternal-risk-factors.

Mayo Foundation for Medical Education and Research. (2023, August 10). Managing heart conditions during pregnancy. Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/pregnancy-week-by-week/in-depth/pregnancy/art-20045977.

Meshram, Jagruti & Devi, Seeta & Gaikwad, Sachin & Podder, Lily & Ramachandran, Harikrishnan. (2023). Machine Learning Algorithm and GESTOSIS Score Assisted High Risk Pregnancy Induced Hypertension Prediction. Revue d'Intelligence Artificielle. 37. 117-128. https://doi.org/10.18280/ria.370115.

Nahm, F. S. (2022). Receiver operating characteristic curve: Overview and practical use for clinicians. *Korean Journal of Anesthesiology, 75*(1), 25–36. https://doi.org/10.4097/kja.21209.

Professional, C. C. medical. (n.d.). High-risk pregnancy: Risk factors, Complications & Treatment. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/22190-high-risk-pregnancy.

Ramirez, L. (2023, August 24). Doh, who, Koica Partnership empowers local authorities in improving the health of mothers, children, and teens in the Philippines. World Health Organization. https://www.who.int/philippines/news/detail/24-08-2023-doh--who--koica-partnership-empowers-local-authorities--in-improving-the-health-of-mothers--children--and-teens-in-the-philippines.

Raza A, Siddiqui HUR, Munir K, Almutairi M, Rustam F, Ashraf I. (2022) Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. PLOS ONE 17(11): e0276525. https://doi.org/10.1371/journal.pone.0276525.

Sandbhor, S., Chaphalkar, N.B. (2019). Impact of Outlier Detection on Neural Networks Based Property Value Prediction. In: Satapathy, S., Bhateja, V., Somanah, R., Yang, XS., Senkerik, R. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 862. Springer, Singapore. https://doi.org/10.1007/978-981-13-3329-3_45.

Shamshuzzoha, Md., & Islam, Md. M. (2023). Early Prediction Model of Macrosomia Using Machine Learning for Clinical Decision Support. Diagnostics, 13(17), 2754. https://doi.org/10.3390/diagnostics13172754.

U.S. Department of Health and Human Services. (n.d.). Pregnancy if you have diabetes - NIDDK. National Institute of Diabetes and Digestive and Kidney Diseases. https://www.niddk.nih.gov/health-information/diabetes/diabetes-pregnancy.

United Nations. (n.d.). SDG indicators - SDG indicators. United Nations. https://unstats.un.org/sdgs/metadata/.

Verma, N. (2022, September 7). *XGBoost algorithm explained in less than 5 minutes*. Medium. https://medium.com/@techynilesh/xgboost-algorithm-explained-in-less-than-5-minutes-b561dcc1ccee.

World Health Organization. (2015, September). SDG target 3.1 maternal mortality. World Health Organization. https://www.who.int/data/gho/data/themes/topics/sdg-target-3-1-maternal-mortality.

World Health Organization. (2021). COP26 special report on climate change and health: the health argument for climate action.

World Health Organization. (n.d.). Maternal mortality. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality.

WorldBank. (n.d.). Philippines. World Bank Gender Data Portal. https://genderdata.worldbank.org/countries/philippines/#:~:text=78%20women%20die%20per%20100%2C000,same%20as%20its%20regional%20average.

Wu, Y., Zhang, C., Mol, B. W. J., Kawai, A., Li, C., Chen, L., Wang, Y., Sheng, J., Fan, J., Shi, Y., & Huang, H. (2020). Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *The Journal of Clinical Endocrinology and Metabolism*, *106*(3), e1191–e1205. https://doi.org/10.1210/clinem/dgaa899.

APPENDIX A:

SCREENSHOT OF RAW DATASET

Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	Risk
25	130	80	15	98	86	high risk
35	140	90	13	98	70	high risk
29	90	70	8	100	80	high risk
30	140	85	7	98	70	high risk
35	120	60	6.1	98	76	low risk
23	140	80	7.01	98	70	high risk
23	130	70	7.01	98	78	mid risk
35	85	60	11	102	86	high risk
32	120	90	6.9	98	70	mid risk
42	130	80	18	98	70	high risk
23	90	60	7.01	98	76	low risk
19	120	80	7	98	70	mid risk
25	110	89	7.01	98	77	low risk
20	120	75	7.01	100	70	mid risk
48	120	80	11	98	88	mid risk
15	120	80	7.01	98	70	low risk
50	140	90	15	98	90	high risk
25	140	100	7.01	98	80	high risk
30	120	80	6.9	101	76	mid risk
10	70	50	6.9	98	70	low risk
40	140	100	18	98	90	high risk
50	140	80	6.7	98	70	mid risk

APPENDIX B:

CONVERSION OF VALUES

Converting High Risk, Mid Risk to 1, and Low Risk to 0

Risk	
Categorical Value	**Numerical Value**
Low Risk	0
High Risk, Mid Risk	1

APPENDIX C:

SOURCE CODE

Initialization

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

        from sklearn.model_selection import train_test_split as split, GridSearchCV
        from sklearn.metrics import accuracy_score, confusion_matrix, ConfusionMatrixDisplay, roc_auc_score, roc_curve, auc, precision_sc

In [2]: data = pd.read_csv('MH.csv')
        data
```

This is the code for importing the necessary libraries and assigning the dataset to the variable "data".

Data processing

```
In [3]: data_proc = data.replace({'high risk': 2, 'mid risk': 1, 'low risk' : 0}, regex=True)

In [4]: data_proc = data_proc[~((data_proc["BodyTemp"] > 100) & (data_proc["Risk"] == 0))]
        data_proc = data_proc[~((data_proc["BS"] > 10) & (data_proc["Risk"] == 0) | (data_proc["BS"] > 10) & (data_proc["Risk"] == 1))]
        data_proc = data_proc[~((data_proc["Age"] > 50) & (data_proc["Risk"] == 0) | (data_proc["Age"] > 50) & (data_proc["Risk"] == 1))]
        data_proc = data_proc[~(data_proc["HeartRate"] < 55)]
```

```
data_proc = data_proc.replace({2: 1}, regex=True)
data_proc
```

This is where the processing of the data will happen. First, the different class was changed to its numerical value and then as the dataset contains impossible data, this code removes all the impossible data present. Then the code shown changes all the "2" which is the "High Risk" to "1".

Removing Outliers with z-score

```
In [6]: from scipy import stats

        z_scores = np.abs(stats.zscore(data_proc))

        threshold = 3

        # Find rows where all z-scores are within the threshold of -3 to 3
        keep_rows = (z_scores < threshold).all(axis=1) & (z_scores > -threshold).all(axis=1)

        # Filter the data_proc array to keep only the rows that meet the threshold criteria
        data_proc = data_proc[keep_rows]
        data_proc
Out[6]:
```

Z score was used to remove the outliers present in the data. If the record's absolute value exceeds the threshold which is -3 and 3 it will be removed.

Splitting Dataset to Training and Testing (80:20) at Random

```
In [7]: # Processed Dataset
        X_proc = data_proc.drop("Risk", axis=1)
        y_proc = data_proc.Risk
        x_train_proc, x_test_proc, y_train_proc, y_test_proc = split(X_proc, y_proc, train_size=0.8, test_size=0.2, random_state=1)

        # print(f"Original data has {x_train.shape[0]} train data and {x_test.shape[0]} test data\n")
        print(f"Processes data has {x_train_proc.shape[0]} train data and {x_test_proc.shape[0]} test data")

        Processes data has 689 train data and 173 test data
```

The dataset was split randomly with a training set of 80% and a testing set of 20%.

Train the Model and Evaluate Performance with a Classification Report

```
In [9]: import xgboost as xgb

        xgb_classifier4 = xgb.XGBClassifier(
            learning_rate=0.1,
            max_depth=5,
            gamma=0.1,
            n_estimators=1000,
            subsample=0.7,
            colsample_bytree=0.9,
            reg_alpha=0.1,
            reg_lambda=0.5,
            random_state=42)

        xgb_classifier4.fit(x_train_proc, y_train_proc,)

        eval_set = [(x_train_proc, y_train_proc), (x_test_proc, y_test_proc)]
        xgb_classifier4.fit(x_train_proc, y_train_proc, eval_metric=["auc"], eval_set=eval_set, verbose=0)
        y_pred_proc = xgb_classifier4.predict(x_test_proc)
        y_prob = xgb_classifier4.predict_proba(x_test_proc)[:, 1]

        print(f"Processed Dataset Accuracy: {accuracy_score(y_test_proc, y_pred_proc)}")
        print(classification_report(y_test_proc, y_pred_proc))
```

This is where the model XGBoost is trained, its hyperparameters are fine-tuned first and then the model is fed with the training set to train and then fed with the testing set for the evaluation of its performance which will be outputted by the Classification report.

Evaluate with AUC-ROC metric.

```
from yellowbrick.classifier import ROCAUC
visualizer = ROCAUC(xgb_classifier4, classes=["0", "1"], support=True)

visualizer.fit(x_train_proc, y_train_proc)        # Fit the training data to the visualizer
visualizer.score(x_test_proc, y_test_proc)        # Evaluate the model on the test data
visualizer.show()
```

This part of the code shows the visualization of the AUC-ROC metric.

Predict

```
In [16]: def predict_with_user_input(model):
    print("Enter values for features:")
    age = float(input("Age: "))
    systolic_bp = float(input("Systolic Blood Pressure: "))
    diastolic_bp = float(input("Diastolic Blood Pressure: "))
    blood_sugar = float(input("Blood Sugar (mmol/l): "))
    body_temp = float(input("Body Temperature (°F): "))
    heart_rate = float(input("Heart Rate: "))

    new_data = pd.DataFrame({
        'Age': [age],
        'SystolicBP': [systolic_bp],
        'DiastolicBP': [diastolic_bp],
        'BS': [blood_sugar],
        'BodyTemp': [body_temp],
        'HeartRate': [heart_rate]
    })

    # Using the trained model to predict the outcome
    prediction = model.predict(new_data)

    print("\nResults:")

    if prediction[0] == 1:
        print("\nYou are at risk.")
        print("\nThese are the possible problems you are facing that contribute to your result:\n")

        if age < 20:
            print("- Your age is lower than the prefered age of pregnancy. The normal age in pregnancy is 20-29")
        elif age > 29:
            print("- Your age is higher than the prefered age of pregnancy. The normal age in pregnancy is 20-29")

        if systolic_bp >= 120:
            print("- You have a high systolic BP. The normal is less than or equal to 120 mmHg")

        if diastolic_bp >= 80:
            print("- You have a high diastolic BP. The normal is less than or equal to 80 mmHg")

        if blood_sugar >= 7.8:
            print("- You have a high Blood Sugar. The normal is less than 7.8 mmol/l or less than 140 mg/dl")

        if body_temp >= 98.7:
            print("- Your temperature is high. The normal is 98.6°F or 37°C")

        if heart_rate > 90:
            print("- Your heart rate is high. The normal is 75-80 bpm")
        elif heart_rate < 70:
            print("- Your heart rate is low. The normal is 75-80 bpm")

    else:
        print("\nYou are not at risk.")

predict_with_user_input(xgb_classifier4)
```

This part of the code is used for testing the model. The code will require the user to input the values necessary to make the prediction and then the user's maternal health risk will be predicted. If the user is predicted to be "at risk" it will show the possible leading causes of the result and the normal values for each feature. If the prediction is "not at risk" it will just state that the user is not at risk.

CURRICULUM VITAE

Laurence EJ C. Manjares

laurencemanjares109@gmail.com | 09666804411

PERSONAL INFORMATION

Date of Birth:	June 22, 2002
Age:	21 years old
Place of Birth:	Calamba City, Laguna
Height:	167 cm
Religion:	Catholic
Sex:	Male
Civil Status:	Single

EDUCATIONAL ATTAINMENT

Bachelor's Degree: 2020-Present

National University (NU Laguna)
Bachelor of Science in Computer Science

Senior High School: 2018-2020

Colegio de San Juan de Letran Calamba
Science, Technology, Engineering, and Mathematics (STEM)

Vincent A. Abas

abasvincent12@gmail.com | 09694770402

PERSONAL INFORMATION

Date of Birth:	October 12, 2002
Age:	21 years old
Place of Birth:	Binan City, Laguna
Height:	167 cm
Religion:	Catholic
Sex:	Male
Civil Status:	Single

EDUCATIONAL ATTAINMENT

Bachelor's Degree: 2020-Present

National University (NU Laguna)
Bachelor of Science in Computer Science

Senior High School: 2018-2020

STI College Calamba
Information and Communication Technology (ICT)

Lord Stephen C. Encarnacion

stephenencarnacion@gmail.com | 09691959196

PERSONAL INFORMATION

Date of Birth:	October 11, 2001
Age:	22 years old
Place of Birth:	Calamba City, Laguna
Height:	192 cm
Religion:	Christian
Sex:	Male
Civil Status:	Single

EDUCATIONAL ATTAINMENT

Bachelor's Degree: 2020-Present

National University (NU Laguna)
Bachelor of Science in Computer Science

Senior High School: 2018-2020

Saint Benilde International School (Calamba), Inc.
General Academic Strand (GAS)