AFRICAN CENTRE OF EXCELLENCE

IN DATA SCIENCE

APPLICATION OF MACHINE LEARNING ALGORITHMS TO PREDICT FACTORSOF CESAREAN DELIVERY AMONG WOMEN IN RWANDA

By

Rose AYINKAMIYE

Registration Number: 221000115

A dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science of Data Science in Biostatistics.

University of Rwanda, College of Business and Economics

Supervisor: Prof. François Niragire (Ph.D.)

September, 2023

DECLARATION

I declare that this dissertation entitled "**Application of machine learning algorithms to predict factors of cesarean delivery among women in Rwanda** "is the result of my own work and has not been submitted for any other degree at the University of Rwanda or any other institution.

Student Name: Rose AYINKAMIYE

Signature: 

APPROVAL SHEET

This dissertation entitled "**Application of machine learning algorithms to predict factors of cesarean delivery among women in Rwanda.**" written and submitted by Rose AYINKAMIYE in partial fulfilment of the requirements for the degree of Master of Science in Data Science in Biostatistics, is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 19% which is less than 20% accepted by African Center of Excellence in Data Science (ACE-DS).

17 September 2023

Prof. François NIRAGIRE (PH. D)

Supervisor

Dr. Ignace KABANO

Head of Training

ACKNOWLEDGMENTS

Firstly, I thank the Lord Almighty for the opportunity of unforgettable fruitful during my study.

Secondary , I would like to express my sincere gratitude to the team at the University of Rwanda's African center of Excellence in Data Science for providing me with the opportunity to gain fundamental experience in skills-based learning during my studies.

Thirdly, I would like to express my gratitude to Prof. Francois Niragire , my thesis supervisor , for guiding my research and offering essential resources and a supportive infrastructure that greatly contributed to the achievement of my study objectives.

I appreciate my loving family for their unconditional affection, support, encouragement, and always being there for the whole thing I needed during my studies.

Finally, I want to express my heartfelt gratitude to all my classmates, especially those in the biostatistics department , with whom I have relied on and shared wonderful moments throughout my academic journey

I thank you all!

ABSTRACT

Cesarean delivery is a surgical technique typically performed when vaginal childbirth would pose risk to the health of both the mother and the baby . From 1990 to 2021, the rate of cesarean deliveries increased from 7% to 21.1% of all global deliveries. Cesarean deliveries are becoming increasingly common in Sub-Saharan African , just as they are in Rwanda in 2021, they accounted for 15.6% of all births. Machine learning (ML) techniques have not yet been applied in Rwanda to predict cesarean delivery and identify the factors contributing to its increasing prevalence . As result, machine learning algorithms were employed to predict cesarean deliveries among women, which served as the project's primary objective. This initiative aims to enhance maternal and newborn safety by reducing unnecessary cesarean deliveries during labor and after.

This study used a cross-sectional secondary data from the Rwanda demographic health survey (RDHS 2019-2020) to predict cesarean deliveries. To achieve the main research goal, five ML algorithms, a namely logistic regressions, decision trees, random forests, K nearest neighbor (KNN) classifier and support vector machine (SVM) were applied. Evaluation metrics such confusion matrix, precision, accuracy, area under curve score, and F1 score were used to determine the best cesarean delivery prediction model.

Multivariable logistic regression results show that, several factors linked to higher cesarean delivery likelihood in Rwandan women. These factors include birth order, number of antenatal care visits, birth weight, mother's education level, short rapid breaths of mother during labor, mother's wealth index, place of residence, mother's current age, age of mother at first birth, number of terminated pregnancies , and the twin status of the child. In terms of predictive models for cesarean delivery created using machine learning techniques, the random forest model emerged as the most effective model achieving an accuracy rate of 83.0% and an area under the curve (AUC) score of 90.2%. The decision tree model followed closely with an accuracy rate of 79% and AUC score of 80.7%.The k-nearest neighbor model achieved an accuracy rate of 75% and AUC score of 81.4%. Meanwhile, the support vector machine model exhibited an accuracy rate of 63% and AUC score of 63%. Lastly, the logistic regression model had the lowest performance , with an accuracy of 61% and AUC score of 66.2%, when compared to the other predictive models used in the study.

Our research highlights various factors driving the increase in cesarean sections among Rwandan women. Therefore, healthcare policies should explore alternative strategies to ensure women can access essential cesarean deliveries as advised by a doctor

Keywords: Cesarean delivery, prediction, DHS, machine learning algorithms, Rwanda

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

RDHS: Rwanda Demographic Health Survey

PDHS: Pakistan Demographic Health Survey

WHO: World health Organization?

ML: Machine Learning

CS: Cesarean Section

HIV: Human Immunodeficiency Virus

AUC: Area Under Curve

SSA: Sub-Saharan Africa

KNN: K Nearest Neighbor

SVM: Support Vector Machine

RPHC: Rwanda Population Housing Census

IR: Individual Records

TN: True Negative

TP: True Positive

FN: False Negative

FP: False Positive

ROC: Receiver Operating Characteristic

STI: Sexual Transmission Infection diseases

OR: Odd Ratio

PSU: Primary Sampling Unit

LIST OF DEFINITIONS

Fetal distress: is the time that the infant may start experiencing issues that result in an irregular heartbeat [1].

Breach presentation: is when the infant is set up to come out feet or bottom first It has significant implications in terms of delivery since, it carries a higher perinatal mortality and morbidity, largely due to birth asphyxia [2].

Antepartum hemorrhage: is the blood loss in excess of 1000 ml from 24 weeks of pregnancy and prior to the birth of baby, mothers who have experienced antepartum hemorrhage must consider having their baby via cesarean section [3].

Abruptio placentae: is the period of time prior to labor and birth when the placenta begins to detach from the uterine wall, it is among the factors that contribute to bleeding in the second half of a pregnancy [4].

Transverse lie is when the baby is in a horizontal position in the uterus up to labor time, cesarean birth is always performed [5].

Cephalopelvic disproportion: is when labor fails to progress due to mechanical reasons because the fetal head's size and the mother's pelvic size are out of proportion [6].

CHAPTER ONE: GENERAL INTRODUCTION

This chapter serves as an introduction to the rising prevalence of cesarean sections in Rwanda. It delves into the historical context of this trend with investigating the various factors contributing to the increase in cesarean births. The primary objectives here is to minimize unnecessary cesarean sections within Rwanda. Additionally, the chapter offers a comprehensive analysis of cesarean section rates on a global, regional and local level, with specific focus on Rwandan's situation. It also outlines the study's problem statement, highlighting the influence of machine learning (ML) techniques on predictive models for cesarean births. Furthermore, it articulates the study's objectives, research questions, scope, significance , and organizational structure. These elements offer an overview of how the research will be conducted.

1.1. Study background

Cesarean delivery is a surgical procedure used to deliver baby and is typically performed when a vaginal delivery is not possible or safe [7].This procedure is usually chosen when delivering the baby vaginally would pose a risk to either the mother's or the infant's health. It has been increasing at an up-setting rate recently throughout the world [8].

Globally, the rate of cesarean deliveries increased from 7% to 21.0% between 1990 and 2021 for all deliveries. It is predicted that the percentage of cesarean deliveries will reach 29% by 2030[3]. In Africa, cesarean delivery rates increased by 4.5%, rising from 2.9% in 1990 to 7.4% in 2014[9]. The rate of cesarean deliveries in sub-Saharan Africa grew from 5% in 1990 to 8,6% in 2018. In contrast, it increased to 42% in America and the Caribbean. Eastern cesarean delivery rates increased by 44.9%, Western Asia by 34.7% , and Northern Africa by 31.5% [8].

The state of women's health has improved significantly due to the efforts of Rwanda's Ministry of health. Maternal mortality has decreased , reaching 203 deaths per 1,000 live births from 2014-15 to 2019-2020. This decline has been accelerated by a reduction in pregnancy related issues and increased use of and access to knowledgeable and skilled care for pregnant women [10].

According to the final report of the 2019/2020 Rwanda Demographic and Health Survey (RDHS), the cesarean delivery rate in Rwanda was 2.2% between 1992 and 2000. However, this rate significantly increased to 15.5% of all births in 2021[11].Additionally, the prevalence of cesarean sections has been on the rise in Rwandan healthcare facilities. Especially, it was four times higher in private facilities(60.6%) compared to public facilities(15.4%)[12].The purpose of the study was to utilize machine learning algorithms to predict cesarean deliveries and identify the key factors influencing cesarean delivery among women.

1.2. Problem statement

In 1985, a group of experts in reproductive health concluded that there is no justification for any nation to have a cesarean delivery rate outside of the range of 10-15%. Cesarean deliveries are considered as underutilized in nations with rates below 10% and overutilized in nation with rates above 10%. While cesarean deliveries do reduce the rates of maternal and infant mortality, they may also waste human and financial resources [13],[14].

However, despite the above recommendation, Rwanda had a cesarean birth rate nearly 16%, exceeding the recommended range of 10-15% as suggested by the World Health Organization(WHO). This was linked with additional early deliveries and rise in neonatal and maternal death rates[15].

The increase in cesarean deliveries has led to increase in maternal morbidity and mortality which was caused by hemorrhage complication, surgical site infections and blood clots in the veins that can seriously injure or kill a woman [16]. Machine learning algorithms, as a form of artificial intelligence acquired from data to improve information and to make decision and prediction. These algorithms are particularly important in healthcare for predicting the likelihood of a cesarean delivery[17].

In study conducted in republic of Korea, machine learning methods have been established to predict women at higher risk for emergency cesarean in order to reduce maternity and neonatal deaths. The logistic regression models , with an accuracy rate of 78%, exhibited the highest predictive performance among the nine machine learning algorithms evaluated[18].

Another study used machine learning techniques to classify cesarean deliveries. The results of experiment demonstrated that both random forest classifier and K nearest neighbor achieved the best accuracy rates by correctly predicting 95 cases [19].

In Rwanda , machine learning(ML) algorithms have not yet been applied to predict cesarean deliveries and determine the exposures contributing to their increase. Previously, only classical models were used to assess the prevalence of cesarean deliveries and factors associated with them. These models included bivariate and multivariable logistic regression for examining parameters related to the utilization of cesarean section in Rwandan hospitals and among the general population[12]. Thus, this project aimed to employ ML algorithms to predict cesarean birth and to identify the key predictors of cesarean births among Rwandan women.

1.3. Objectives of study

1.3.1. Overall objective

The overall objective of the project is to predict cesarean delivery among women using machine learning algorithms.

1.3.2. Specific objectives

- To build a predictive model for cesarean deliveries in Rwanda utilizing ML algorithms.
- To determine risk variables influencing cesarean delivery in Rwanda.
- To rank order the factors contributing to the rise of cesarean delivery using random feature selection algorithms

1.4. Research questions

- What is the most appropriate model to predict cesarean delivery in Rwanda utilizing ML algorithms?
- What are the key factors of cesarean delivery in Rwanda?
- What factors are the most influential for cesarean delivery using random feature selection algorithms?

1.5. Significance of the study

Machine learning techniques were utilized to evaluate cesarean deliveries and investigate the variables contributing to their rise, utilizing data from the RDHS 2019-2020 dataset. This study provides insights into the factors contributing to the rise in cesarean deliveries and presents the most accurate cesarean delivery prediction model. Policymakers and health planners can utilize the study's findings about the most significant factors in determining cesarean delivery to determine where inter that are most important in determining cesarean delivery to determine where interventions should be targeted. Furthermore, the study's results might help women and broader community in making informed decisions about cesarean sections based on medical advice .

1.6. Scope and study limitation

The goal of this project was to predict cesarean deliveries among Rwandan women using machine learning techniques and elements chosen from a sample size of 14634 women aged 15 to 49, using the RDHS 2019–20 dataset. Five ML techniques were employed to predict the likelihood of cesarean birth in Rwanda, such as logistic regression classifier, support vector machines, decision trees classifier, random forest classifier and k nearest neighbors. Metrics such as accuracy, precision, F1 score, confusion matrix, and area under the curve were used for evaluating the model's performance and logistic multivariable analysis was employed for the analysis. Since there were many missing values for the variables in the RDHS 2019–20 dataset, the approach of deleting rows and columns with missing values during data cleaning was not considered. Therefore, imputation techniques were utilized in this work.

1.7. Organization of study

This study consists of five chapters. The first chapter, which serves as a general introduction delineates the study’s background, problem statement, purpose, study questions, significance of the study, scope and limitation of the study, and the organization of the study. In the second chapter , we delve into the findings related to the system framework, which identifies the variables included in the conceptual framework and explores the factors contributing to the growth in cesarean deliveries. Chapter three expounds on the research methodology, the type of data used , and the machine learning techniques employed to predict cesarean deliveries. The fourth chapter encompasses data analysis, data quality checks, and the presentation of study outcomes. Chapter five includes a discussion of the outcomes , a general conclusion and recommendations.

CHAPTER TWO: LITERATURE REVIEW

2.1.Theoretical review

A Cesarean section (CS) is a surgical procedure employed when vaginal delivery becomes medically unfeasible, primarily to reduce the risk of maternal and neonatal mortality. A study investigating the surge in cesarean section rates in South Asian countries noted a significant increase, soaring from 7.2% in 2000 to 18.1% by 2015. This research, relying on quantitative data analysis and drawing from electronic databases like Medline, Nepjol, and Banglajol, identified several key factors contributing to this rise. These included advanced maternal age, higher maternal education, urban residency, greater economic affluence, prior cesarean deliveries, and pregnancies complicated by medical issues [20]. Understanding these factors is crucial for healthcare providers and policymakers to ensure the careful use of cesarean sections while prioritizing the well-being of both mothers and infants.

The results of a study conducted in Sub-Saharan African countries found that less than 15% of women in the region underwent cesarean sections (C-sections) during childbirth. The primary objective of the study was to investigate the impact of socio-economic factors on the likelihood of women in Sub-Saharan Africa electing for C-sections. The study focused on a target population comprising married women aged 15 to 49 years old and employed binary logistic regression for its analysis. The findings of the study revealed significant associations between the utilization of cesarean delivery and certain socio-economic factors, namely maternal education, wealth index, and place of residence, emphasizing the importance of these factors in shaping maternal healthcare choices and access to C-section deliveries [21].This information provides valuable insights for healthcare planners and policymakers seeking to improve maternal healthcare services in the region.

Research done at the St. Joseph Medical Hospital in Moshi, Tanzania, between 2009 and 2011, there were 1167 babies born, and 212 of them were delivered via cesarean section, accounting for18% of all births. The most common reasons for women to have a cesarean delivery were discovered to be malpresentation of the infant (20%), fetal distress (11%), extended labor (30%),and previous cesarean (14.7%), followed by infections, large babies, and hypertension in moms as the reasons for cesarean in mothers [22].

An analysis of the rising trends in cesarean delivery rates in Pakistan, conducted using data from 1990 to 2013 (PDHS1990-2013 dataset), revealed a substantial increase, climbing from 2.7% in 1990 to 15.8% in 2013. To gain insights into the factors associated with this rise, a multivariate logistic model was employed to examine the prevalence of cesarean births in relation to various sociodemographic factors. The study's findings shed light on significant disparities in cesarean section rates among different groups of women. Notably, women's education played a crucial role, as those with higher education levels exhibited a much higher cesarean rate, standing at 40.3%, compared to women with no formal education, who had a rate of 7.5%. Moreover, a marked urban-rural divide was observed, with women residing in rural areas experiencing a cesarean section rate of 11.5%, in contrast to urban women, amongwhom the rate was considerably higher at 25.6%. Economic disparities also had a notable impact, with cesarean birth rates ranging from 5.5% among the poorest women to 35.3% amongthe wealthiest. These findings underscore the influence of education, urbanization, and socioeconomic status on cesarean delivery trends in Pakistan [23]. Understanding thesedynamics is crucial for healthcare planners and policymakers to ensure equitable access to safechildbirth practices while addressing the factors contributing to the rising cesarean section rates[22].

Research conducted between 2012 and 2013 in Pakistan, utilizing data from the Pakistan Demographic and Health Survey (PDHS), aimed to identify the factors associated withcesarean sections. The researchers employed partial least squares regression for categorical factors and combined Pearson's Contingency Coefficient with loading weight analysis to assessmodel performance and select influential factors. The results of this study revealed several factors that were closely associated with the likelihood of cesarean sections. These factors included the region of residence, place residence (urban or rural), educational levels of both the mother and her partner, the mother's wealth index, her age at childbirth, the number of antenatal care visits, history of previous terminated pregnancies, the size of the child at birth, the time interval between preceding births, the use of contraceptive methods, and the status ofthe baby at birth [23].These findings shed light on the multifaceted determinants of cesarean section rates in Pakistan, providing valuable insights for healthcare policymakers and practitioners in the region.

In a comprehensive study conducted at Japanese perinatal centers, researchers employed a logistic regression model to investigate the trends in cesarean birth rates and identify key contributing factors. Their findings revealed a noteworthy shift in cesarean birth rates over the years, with the proportion rising significantly from 17.3% in 2022 to 23.4% in 2012. Notably, the increase was particularly pronounced in certain demographic groups. Mothers aged over 35 years old, cases of preterm or neonatal birth, low birth weight babies (weighing ≤ 2.5 kg), prior cesarean birth (with a striking 90.0% increase), and instances of breech presentation (with a substantial 94.5% surge) all exhibited a significant upswing in cesarean deliveries during the study period [17]. These findings underscore the importance of further exploring the underlying factors contributing to this trend and developing targeted strategies to optimize maternal and neonatal outcomes.

2.1. Conceptual framework of cesarean section

A Cesarean section (CS) is a critically important medical procedure that has been instrumental in reducing morbidity and mortality rates among both women and newborns. This surgical intervention is often employed to address complications such as direct hemorrhage, infection, pregnancy-related hypertension issues, and labor obstruction during childbirth. By providing a safe and effective means of delivery in these challenging situations, cesarean section plays a vital role in safeguarding the lives and well-being of both mothers and their infants, ensuring a healthier start to their journey together.

As shown in **Figure1**, the framework from Kizito and Schumacher [25], was used to conceptualize the study variables. The preferences of women for cesarean delivery were influenced by various factors, such as socio-economic factors, child characteristics, and maternal factors, all of which had an effect on the dependent variable, cesarean section.

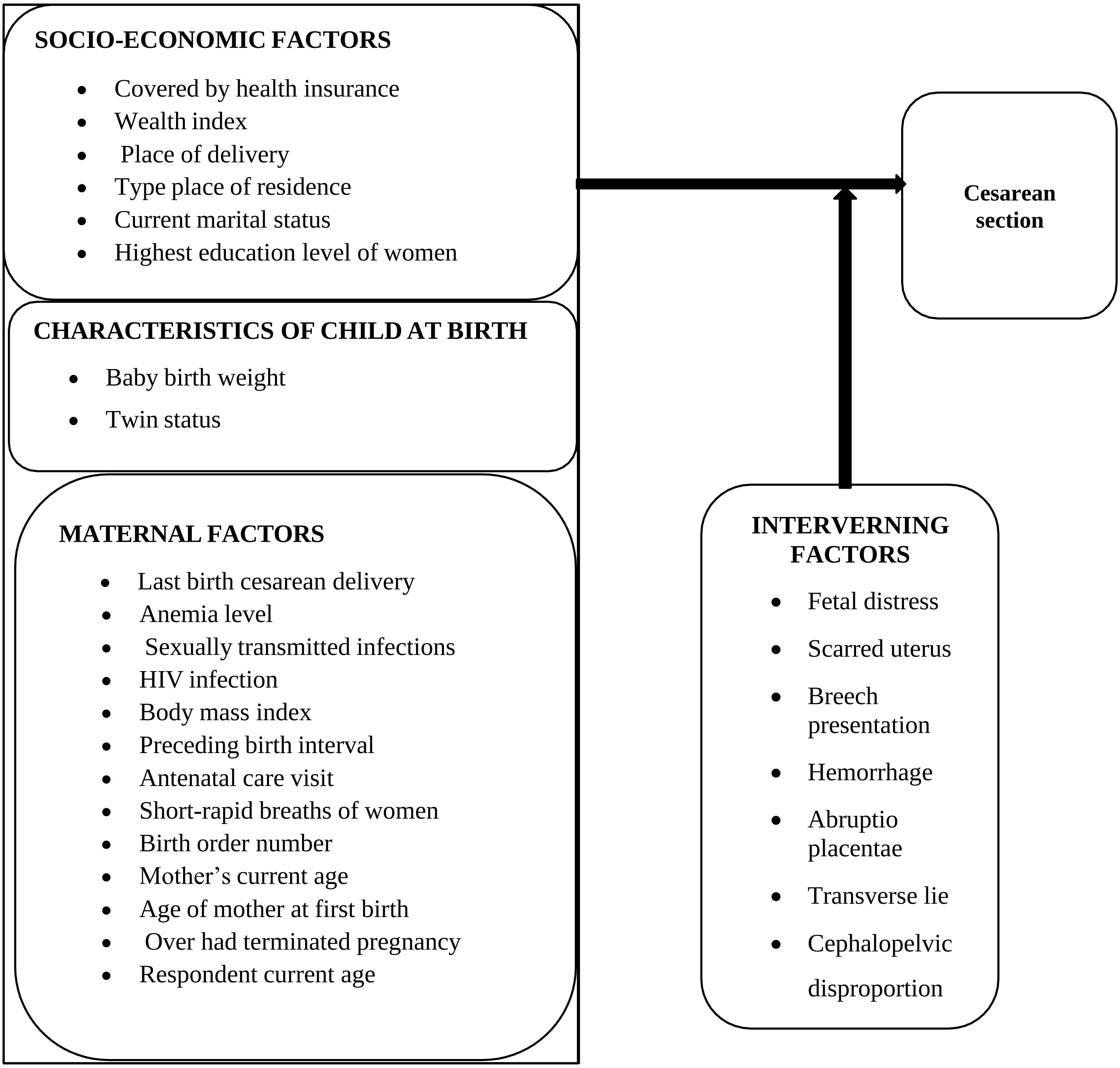

Figure 1:Conceptual framework of cesarean section use(modified from Kizito 2021)

2.2.1 Maternal factors

A woman's choice to undergo a cesarean section is primarily influenced by maternal variables, encompassing aspects related to pregnancy, labor, the mother's overall health, and the condition of her unborn child. These maternal variables include factors such as a diagnosis of hypertension or high blood pressure, the level of anemia, health insurance coverage, whether the last birth was by cesarean section, the interval between preceding births, the respondent's current age, the place of delivery, the presence of sexually transmitted infections, the occurrence of short rapid breath in the woman, and the mother's HIV infection status [26].This array of maternal variables collectively plays a significant role in shaping a woman's decision-making process regarding the mode of childbirth.

2.2.2. Socio-economic factors

Socioeconomic variables, such as a woman's marital status, employment status, wealth index, educational background, and place of residence, have been identified as key factors influencing the increased rate of cesarean deliveries among women. Drawing upon general research findings regarding socioeconomic disparities in cesarean delivery usage in Ghana, the results of a multivariable model analysis revealed several noteworthy trends. Women aged over 25 exhibited a higher likelihood of undergoing cesarean deliveries compared to their younger counterparts aged 15 to 24. Additionally, the study found that wealthier women were more likely to choose for cesarean deliveries in comparison to those with lower economic status. Furthermore, women with higher levels of education displayed a greater cesarean delivery rate, while those with lower educational attainment had a lower likelihood of choosing this mode of delivery. Lastly, cesarean delivery rates were higher among women residing in urban areas as opposed to those living in rural settings[24].These socioeconomic factors collectively contribute to variations in cesarean delivery rates among women in Ghana.

2.2.3.Characteristics of child at birth

The characteristics of the child at birth can influence whether their mother will opt for a cesarean delivery. These child-related factors encompass parameters such as the baby's birth weight and their twin status. These attributes of the child play a crucial role in determining the choice of cesarean delivery for the mother.

1.Baby birth weight

Doctors may recommend a cesarean delivery if the baby's birth weight exceeds 5000 grams, and the predicted weight is greater than 4500 grams, as this scenario carries a higher probability of challenging delivery, birth-related injuries, respiratory issues, and an extended hospital stay requiring intensive monitoring and care [28].

2.Twin status

The factor of a twin pregnancy exerts a substantial influence on the utilization of cesarean delivery among women, largely owing to its association with increased perinatal mortality rates, especially in cases of prematurity. Among planned cesarean deliveries for twin pregnancies, a significant 90.9% of cases opted for cesarean delivery, whereas among planned vaginal births, 40.9% resulted in cesarean deliveries. It is important to note that the challenging nature of twin births may contribute significantly to perinatal mortality or morbidity. To mitigate these difficulties and potential risks, the option of scheduling a cesarean delivery should be carefully considered as an alternative [29].This approach may offer a safer and more controlled birthing experience for women with twin pregnancies.

2.2.3.Intervening factors of cesarean delivery

Intervening factors, in this context, are defined as variables that can impact the relationship between dependent and independent variables but were not included in the current research. According to research conducted to identify risk factors associated with cesarean delivery among women in North Jadon, the results of a multivariable logistic regression model demonstrated that the most common reasons for planning cesarean delivery were as follows: "prolonged fetal distress (233.5%), scarred uterus (63.5%), breech presentation (7.6%), hemorrhage, abruptio placentae (0.5%), transverse lie, and cephalopelvic disproportion (0.3%)" [25].These findings shed light on the predominant factors influencing the decision to opt for a cesarean delivery among this specific group of women.

CHAPTER THREE: RESEARCH METHODOLOGY

3.0. Introduction

Within this chapter, we explain the methodologies and data sources employed in the classification process to predict cesarean deliveries among women. Our aim in this project is to achieve accurate predictions, and to accomplish this, we joined the power of various machine learning algorithms, including random forests, decision trees, k-nearest neighbors, support vector machines, and logistic regression models. These algorithms formed the core of our predictive modeling approach, allowing us to leverage data-driven insights and robust classification techniques to enhance our understanding and prediction of cesarean delivery outcomes.

3.1. Study design

This study focuses on predicting cesarean section rates using cross-sectional secondary data obtained from the Rwanda Demographic Health Survey 2019-2020 (RDHS 2019-2020). Machine learning algorithms are essential tools for this task as they facilitate the prediction and classification of output values based on input data, utilizing various variable and model selection techniques [31]. To investigate the relationship between cesarean deliveries and several explanatory variables, this research employed supervised machine learning models to forecast the likelihood of a cesarean section. In order to achieve accurate predictions, these machine learning models relied on both training and testing datasets [32].

3.2. Data source

The research utilized data from the cross-sectional and nationally representative survey RDHS 2019-2020, conducted in Rwanda.

3.3. Study population

The research was open to all women between the ages of 15 and 49, including permanent residents of the selected households as well as visitors who had spent the night before the survey began.

3.4. Sample size and sampling design

The respondents in this study were selected based on retrospective data obtained from the RDHS 2019-2020 dataset, which consists of individual files (IR) collected through a questionnaire designed to gather information from women between the ages of 15 and 49. This dataset included a total of 14,634 women who had given birth during the RDHS 2019-2020 survey period. The research employed a two-stage sample design method during the respondent selection process [33]. In the first stage, primary sample units (PSUs) were chosen, representing segments of the population from the 2012 RPHC. The second stage involved the systematic random selection of households within the selected PSUs from the first stage. Weighting factors were applied to the sample at each stage to ensure that the results would be representative at the national level [34].This robust sampling methodology allowed for a comprehensive and statistically sound analysis of the research data.

3.5. Study variables

The study utilized the RDHS 2029-20 dataset, comprising 5117 variables and 14634 records. The selection of factors for this research was based on a thorough review of the existing literature and a well-defined conceptual framework.

3.5.1. Dependent variable

The dependent variable employed in this study is the percentage of live births delivered via cesarean section within the five-year period preceding the investigation. In the Rwanda Demographic Health Survey 2019-2020 individual record dataset, the categorization of delivery by cesarean section was previously assigned a value of 0 to signify that the child was not delivered via cesarean, and a value of 1 was assigned to indicate that the infant was born through a cesarean section [35].

3.5.2. Independent variable used for the study and its coding

The explanatory variables in this study, which are associated with the increased occurrence of cesarean sections, were determined based on a comprehensive review of existing literature. In total, 21 variables were recorded and utilized in our research. To facilitate a straightforward and consistent analysis in line with prior studies, many of these variables were combined into two or more categories using specific recording techniques. For instance, we categorized women's body mass index (BMI) into three distinct values: those with a BMI less than 25 were classified as normal, individuals with a BMI ranging from 25 to 30 were considered overweight, and women with a BMI exceeding 30 were categorized as having obesity [36]. Similarly, birth weights were divided into three categories: normal birth weight (birth weight >= 2.5 kgs and < 4.0 kgs), low birth weight (birth weight < 2.5 kgs), and excessively high birth weight (>= 4.0 kgs) [37] . This systematic categorization of variables enables a more structured and comprehensible analysis of the factors contributing to the rise in cesarean sections as identified in the literature review.

Table 1:independemt variable and its coding used for this study

Code for DHS	Name of variable	Variable coding
V401	Last birth cesarean section	0=non 1=yes
B0$01	Child is twin	0=single birth 2=second multiple
V190	Wealth index	1=poorest 2=poorer 3=middle 4=richer 5=richest
M19$1	Birth weights	1=0-2.5kg 2=2.5-4 kgs 3=4^{+} kgs
V212	Age of mother at first pregnancy	1=below 20 years old 2= age between 20 and 34 3= 35 and above
V228	Over had of terminated pregnancy	0= no 1= yes
V445	Women ‘s body mass index	1=Less than 25:Normal 2=25-30: overweight 3= “>=30”: obesity
BORD$01	Birth order number	1=first birth 2=second birth 3=third birth 4= four and over in birth
M14$1	Number of antenatal visits	0=no antenatal care visit 1= 1 2=2-3 3=4^{+}
H31B$1	Short-rapid breaths of mothers	0=no 1=yes

V457	Anemia level of women	1=severe 2=moderate 3=mild 4=not anemic
V106	Mother’s highest education level	0= no education 1=primary 2=secondary 3=higher
V763A	Had any STI in the 12 months	0=no 1=yes
V774A	HIV transmitted during pregnancy	0=no 1=yes
V501	Mothers ‘current marital status	0=never union 1=married 2=living with partner 3= widowed 4=divorced 5= separated
V481	Covered by health insurance	0=no 1=yes
V025	Type place of residence	1=urban 2=rural
M15$1	Place of delivery	1=home 2=hospital 3=another place
V012	Mother’s current age	1=15-25 2=25-35 3=35+ 2=women aged above 25
B11$01	Preceding birth interval (in months)	1=Less than 24 months 2=greater or equal to 24 months

3.6. Study sample size

Based on the RDHS 2019-20 dataset, which included all women who completed the interview, the total sample size for this study comprised 14,634 women between the ages of 15 and 49. Within this research, a total of 953 women gave birth via cesarean section, while 4,837 women opted for vaginal delivery.

3.7. Importing data

The study's data were initially in SPSS format; hence, they were converted to CSV format and subsequently imported into the Python environment using the "read-csv" function. This marks the initial stage of the data analysis process

3.8. Data pre-processing

Data preprocessing is a crucial step in machine learning, aimed at organizing and cleaning data to make it suitable for the creation and training of ML models [38]. In our pursuit of developing the most high-performing cesarean delivery model, we employed various missing data imputation techniques, such as mean imputation for continuous variables and mode imputation for categorical data, as part of our data cleaning process in the current research [39]. Additionally, we identified and eliminated outliers within the dataset using interquartile range techniques, thereby enhancing the data's quality and making it suitable for classification and utilization in our analysis.

3.9. Handling imbalanced dataset

This serves as a typical instance of a classification problem characterized by an unequal distribution of target classes within the dataset. Specifically, the categorical differentiation between cesarean births and vaginal births revealed a notable imbalance, with 953 instances of cesarean section deliveries compared to a much larger count of 4837 instances of vaginal deliveries. To rectify this class imbalance issue, we employed the Synthetic Minority Oversampling Technique (SMOTE), which effectively generates new data points within the minority class. This approach proves advantageous as it helps maintain data integrity within the training dataset while addressing the disparity in class distribution [40].

3.10. Features selection importance

It is technique for minimizing number of variables used in the construction of a prediction model. The current study used random forest classifier feature importance methodology for purpose of finding the features that have the greatest impact on women's use of cesarean births.

3.11. Methods of analysis

3.11.1. Cross-tabulation analysis

This was applied to see the connection between categorical variables. By extension, it indicates whether the variables are related to each other. The chi-squared test uses a contingency table to examine data that have been categorized and organized into rows and columns, with one category's data being arranged in rows and the other in columns, and each categorical column containing at least two or more categories[41]. In this study, the cesarean delivery status was ascertained by conducting cross-tabulations between the outcome variable and predictor variables. This approach allowed us to examine the distribution of independent variables in relation to the dependent variable and assess whether any associations existed.

3.11.2. Multivariable analysis

In pursuit of the study's second objective, which is to "determine the risk variables associated with the rise in cesarean deliveries in Rwanda," we conducted a multivariable logistic regression analysis. This analysis was conducted at a significance level of 5% with a 95% confidence interval, enabling us to assess the impact of explanatory variables on cesarean delivery. Furthermore, we successfully identified and categorized the underlying causes contributing to the increase in cesarean births, providing valuable insights into this important aspect of maternal healthcare in Rwanda.

3.12. Methods of constructing predictive models

This study utilized the RDHS2019-20 dataset, encompassing individual records (IR), to construct a predictive model for cesarean deliveries. Machine learning techniques, including logistic regression, decision trees, random forests classifier, KNN classifiers, and support vector machines, were employed for this purpose. These algorithms were deemed appropriate for this study due to the binary nature of the outcome variable, cesarean delivery status, which was categorized as "yes" for live births delivered via cesarean section and "no" for other cases.

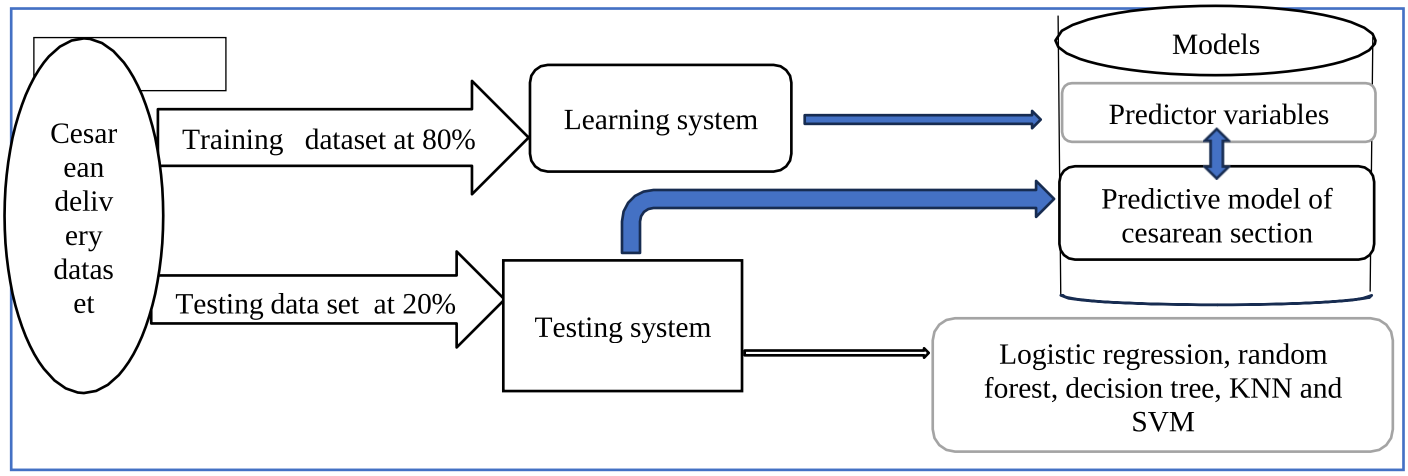

Figure 2:ALgorithm(process) for constructing predictive model of cesarean section

a) Logistic regression model

The study employed a machine learning technique known as logistic regression, specifically for the categorization of discrete outcomes into two distinct categories. Logistic regression is a valuable tool for binary classification, enabling the prediction of binary events by leveraging the logit function to address classification challenges. This method effectively establishes and estimates the relationship between a single dependent variable and a multitude of explanatory variables. Logistic regression finds widespread application in various fields, such as marketing, where it predicts whether a customer will purchase a product, in the banking sector to discern fraudulent credit card transactions, and in online education to forecast whether a student will successfully complete a course within a given timeframe [42]. In our research, we connected the power of logistic regression to construct a predictive model tailored to the context of cesarean deliveries.

b) Decision trees

Decision trees represent a type of supervised learning method applicable to both classification and regression challenges. The fundamental purpose of this algorithm is to construct a predictive model capable of forecasting the values of outcome variables. Decision trees employ a tree-like visual representation to address problems. Within this structure, decisions are made based on specific assumptions, where each leaf node corresponds to a target label, branches encapsulate decision rules, and the dataset's features are situated at the inner nodes of the tree. This method furnishes a collection of rules organized in a format resembling a flowchart, with inner nodes serving as tests or features at various stages. Each branch signifies an outcome related to the associated features, and the path from a leaf node to the root effectively explains the classification rules [43].

A tree may be understood through dividing the dataset into subsets relying on the quality of good test. This cycle is repeated on every decided subset in a recursive manner called recursive partitioning. The recursion is stated to be whole while the subset on the node is the identical cost as that of the target variable subset or while splitting does now no longer add any greater cost to the predictions made [44]. In this study, the prediction model for cesarean birth was created using decision tree classifier.

c) Random forest classifier

Random forest algorithms defined as common supervised ML techniques. It can be applied to Problems involving both classification and regression. It is entirely based on the principle of ensemble learning, which combined more than one decision tree to produce a forest of trees called a random forest in order to solve classification and regression problems and enhancing the model's performance. Random forest uses many decision trees on different subsets of the provided datasets and uses the common to increase the predicted accuracy of dataset.

The random forest builds decision trees on a specific dataset by taking the prediction from each tree, as opposed to just counting one selection tree, it predicts the outcome by producing predictions from each tree and selecting the best response based on voting winnings, more trees increase the model's accuracy and help to prevent overfitting issues in the dataset [45]. The study used this classifier to build a predictive random forest model for cesarean sections.

d) Support vector machine

In our current study, we employed the supervised machine learning technique known as Support Vector Machine (SVM) to construct a predictive model for cesarean deliveries. SVM is a widely utilized approach, particularly in classification tasks. The main objective of the SVM algorithm is to establish an optimal decision boundary, often referred to as a hyperplane, that effectively separates data points in an n-dimensional space into distinct classes. This enables us to efficiently classify new data points in future scenarios. By leveraging SVM, we aimed to develop a robust and accurate predictive model for cesarean deliveries, enhancing our understanding and decision-making capabilities in the context of maternal healthcare [46].

e) K nearest neighbor classifier

The K-nearest-neighbor (KNN) classifier is a versatile member of the supervised learning algorithm family. It is a widely employed classifier that classifies data by considering the nearest training samples within a specific neighborhood. The label most frequently represented among these neighboring data points is assigned to the data point in question. In our study, we harnessed the power of KNN to construct a predictive model for cesarean deliveries. This choice was motivated by its straightforward implementation and rapid computation, making it a practical tool for building predictive models in the context of our research [47].

3.12.1. Methods of evaluating performance of predictive models

In our study, we employed machine learning metrics to gauge the effectiveness of ourpredictive models. During the evaluation phase, we relied on several performance metrics, including the confusion matrix, precision, accuracy, area under the curve (AUC), and the F1 score. These metrics played a crucial role in assessing the predictive power and reliability of our models, allowing us to make informed judgments about their performance and suitability for the specific task at hand.

a. Confusion matrix

A classification model's overall effectiveness is assessed using a confusion matrix, which is a N x N matrix where N is the total quantity of target classes. With Algorithms, the confusion matrix compares the actual target values to those predicted values. This offers us a complete picture of the classification model's performance and the types of errors it is generating. Actual values are true values taken from specific observations, whereas predicted values are values that the model predicts. These two values are combined to generate the confusion matrix [48]. The below table shows 2x2 matrix for positive and negative classes.

Table 2:Confusion matrix

Total predictions	**Real: Positive**	**Real: Negative**
Predicted: Positive	True positive	False positive
Predicted: Negative	False negative	True negative

- **True Negative (TN)** shows how a model accurately predicted negative values as negative.
- **True Positive (TP)** shows how a model accurately identified Positive values as Positive.
- **False Negative (FN)** shows how a model wrongly predicted positive values as negative.
- **False Positive (FP)** shows how a model wrongly predicted negative values as positive

b. Precision

Precision is a crucial metric that measures the accuracy of a model's predictions among all the positive classes that it correctly identified. In our study, precision played a pivotal role in evaluating the model's performance by assessing its output through the utilization of a confusion matrix. This allowed us to gauge the model's effectiveness in correctly identifying and classifying instances within the positive classes, providing valuable insights into its predictive capabilities.

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

c. Accuracy

Accuracy is determined by calculating the proportion of correct predictions made by the model's output in relation to the total number of predictions generated by the model. This metric is employed to gauge how often the model accurately predicts the desired outcomes.

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

d. Recall

The recall score is employed to assess a model's capability to predict positive values. In this study, it was used to gauge how accurately the model predicts positive attributes and measure the frequency of such correct predictions.

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

e. F1 score

The F1 score is often described as the harmonic mean that encapsulates both recall and precision. This technique becomes particularly useful when there is a need to consider both precision and recall simultaneously. In our study, we evaluated both recall and precision concurrently to provide a comprehensive assessment of the model's performance.

$$\textbf{F1 Score} = \frac{2*recall*precision}{recall+precision}$$

f. Area under curve score

A graphical representation called a ROC curve serves to summarize the performance of a model across various threshold settings. The area under the ROC curve, often referred to as AUC, varies between 0 and 1. An AUC value of 1 signifies the optimal performance potential of our model, whereas an AUC of 0 suggests that all predictions are incorrect. In this study, we employed the AUC score to assess the performance of our prediction model.

CHAPTER FOUR: DATA ANALYSIS AND PRESENTATION OF RESULTS

4.1 Introduction

The current chapter presents the outcomes obtained through the application of machinelearning algorithms to predict the factors contributing to the rise of cesarean sections in Rwanda, utilizing data from the RDHS2019-20 dataset. In pursuit of the objective to identify the factors associated with the increase in cesarean sections, a multivariable binary logistic regression was employed. Furthermore, to achieve the primary aims of this study, a predictive modelfor cesarean deliveries was constructed using a combination of support vector machines, decision trees,random forests, and KNN classifiers. The evaluation of these models utilized various metrics such as confusion matrices, precision, accuracy, and the F1 score. These metrics were employed to ascertain the accuracy and effectiveness of the cesarean delivery prediction models.

4.2. Data quality check

Before to constructing machine learning techniques, it was essential to preprocess the data by applying reduction, data cleaning, and balancing techniques. This preprocessing was crucial to enhance model performance and ensure that the model generates accurate predictions [49].

4.2.1. Handling missing value

The original dataset comprised 1,436 records and included 23 attributes. Following data cleaning procedures, specifically addressing missing values, we expanded the dataset to encompass 6,167 records while retaining the same 23 attributes. For categorical variables with missing values, we employed the imputation mode method, which entails assigning the most frequently occurring value in the respective category to fill in the gaps. In cases of numerical variables with missing data, we opted for the imputation mean approach. This involved computing the average across all variables and substituting missing values with this calculated mean. These meticulous data cleaning steps were undertaken to ensure the dataset's completeness and integrity, laying the foundation for the development of a robust cesarean prediction model.

4.2.2. Handling outliers

Outliers are particular data points that fall outside the typical range of values for the entire dataset. These outliers often have the potential to distort the outcomes of statistical analyses conducted on the dataset. In this study, we employed a box plot as a visual tool to identify outliers in numerical variables. Additionally, when dealing with non-normally distributed data, we relied on the interquartile range to pinpoint and handle outliers [50].Following the removal of outliers from the dataset, which was achieved through addressing missing values, our final dataset consisted of 5,790 records and 23 attributes.

4.2.3. Balancing dataset used in the study

In the study, the dependent variable exhibited 953 entries in the minority class, while the majority class consisted of 4837 data points. To address the class imbalance issue, weemployed the Synthetic Minority Oversampling Technique (SMOTE), which involves oversampling the minority class by generating additional samples from it. Upon achieving a balanced dataset, the total of 4837 cesarean births and 4837 vaginal births were evenly distributed across the dataset for further analysis.

4.3. Frequencies distribution of study sample according to selected potential predictors.

The results of the univariate analysis, as presented in **Table 3**, reveal that a total of 5,790 women participated in the study. Among them, 83.54% (4,837) opted for vaginal delivery, while 16.46% (953) underwent cesarean section deliveries. It is noteworthy that the majority of women (82.04%) gave birth to babies with weights ranging from 2.5 kgs to 4 kgs, in contrast to a smaller percentage (6.86%) who delivered babies weighing over 4 kgs.

In the context of socio-economic disparities, 23.04% of the poorest women were active participants in this study. The majority of women in the sample had completed primary school, accounting for 65.8%, as opposed to those with higher education qualifications. In terms of marital status, 50.16% of the women were married, in contrast to widowed women. Health-wise, over 95.36% of the women reported no history of STIs in the past 12 months, whereas 4.64% had experienced such infections. Geographically, a significant majority hailed from rural areas (77.91%), with the remaining 22.09% residing in urban settings. A remarkable 97.01% of women gave birth in a hospital, while only 1.16% opted for home births.

The study primarily consisted of female participants, with 77.2% falling within the age range of 20 to 34 at the time of their first childbirth, while a mere 0.36% were older than 30 years. A significant portion of women attended at least four antenatal care visits (49.15%), in contrast to those who had no antenatal care visits (1.19%). Concerning body mass index (BMI), the distribution among women in the study was as follows: 49.64% were classified as thin, 35.27% as normal, 11.5% as overweight, and 3.58% as obese. Approximately 10.25% of the women reported experiencing shortness of breath, while 89.5% did not encounter any breathing difficulties.

Regarding pregnancy intervals, a higher proportion (90.22%) of deliveries occurred among mothers with an interval of 24 months or more between pregnancies, compared to 7.78% of births to mothers with shorter intervals. Cesarean sections were more frequently reported among mothers with four or more previous births (32.9%) as opposed to those with a third birth (18.95%). A majority of women in the study had a history of pregnancy complications (23.14%) compared to those who did not (16.29%). In terms of HIV transmission during pregnancy, 23.14% of women reported such transmission, while 26.29% did not experience it. Furthermore, the study found that the majority of mothers (93.94%) did not suffer from anemia, whereas only a small fraction (0.1%) had severe levels of anemia.

Table 3:Frequencies distribution of study participants according to selected potential predictors

Characteristics	**Frequencies(N)**	**Percentage (%)**
Delivery by cesarean section		
No	4837	83.54
Yes	953	16.46
Last birth of cesarean section		
No	4837	83.54
Yes	953	16.46
Age of mother at first birth		
Below 20 years old	1276	22.04
Age between 20 and 34	4470	77.2
35 and above	44	0.76
Over had terminated pregnancy		
No	4847	16.29
Yes	943	23.04
Birth weight(kgs)		
0-2.5 kgs	643	11.11
2.5-4 kgs	4750	82.04
4^{+}	397	6.86
Wealth index of women		
Poorest	1334	23.04
Poorer	1092	18.86
Middle	1117	19.29
Richer	1124	19.41
Richest	1123	19.4

Anemia level of women		
Severe	6	0.1
Moderate	95	1.64
Mild	245	4.23
Not anemic	5439	93.94
Mothers' highest education level		
No education	566	9.75
Primary	3780	65.28
Secondary	1153	19.91
Higher	291	5.03
Had any STI in the 12 months		
No	5523	95.39
Yes	267	4.61
Mother's current marital status		
Never union	624	10.78
Married	2904	50.16
Living with partner	1786	30.85
Widowed	84	1.45
Divorced	144	2.49
Separated	248	4.28
Covered by health insurance		
No	1005	17.36
Yes	4785	82.64
Place of residence		
Urban	1278	22.09
Rural	4508	77.91
Place of delivery		
Home	106	1.83
Hospitals	67	1.16
Another place	5617	97.01
HIV transmitted during pregnancy		
No	1522	**26.29**
Yes	4268	**73.71**

Mother's current age		
Women age between 15 and 25	1022	17.65
Women age between 25 and 35	2796	48.29
Women age 35 and above	1972	34.06
Number antenatal visits		
0	70	1.21
1	202	3.49
2-3	2673	46.17
4^{+}	2845	49.15
Body mass index of women		
0-18.49: Thin	2875	49.64
18.49- 24.9: Normal	2042	35.27
24.9-29.9: overweight	666	11.5
30 weight and over: obese	207	3.58
Birth order		
First birth	1433	24.75
Second birth	1355	23.4
Third birth	1097	18.95
Four and over in birth	1905	32.9
Mother's short rapid breaths		
No	5196	89.74
Yes	594	10.26
Preceding birth interval		
Less than 24 months	566	9.78
Greater or equal 24 months	5224	90.22
Child is twin		
Single birth	5704	98.51
Second multiple	86	1.49

4.4. Association of selected potential predictors with cesarean delivery using cross tabulation analysis.

The cross-tabulation findings revealed statistically significant associations between various factors and the increasing prevalence of cesarean sections among women. These factors include birth weights ($\chi2$ = 14.29, P-value < 0.001), mom's wealth index ($\chi2$ = 171.01, P-value < 0.001), mom's highest education level ($\chi2$ = 219.98, P-value < 0.001), marital status of women ($\chi2$ = 29.19, P-value < 0.001), coverage by health insurance ($\chi2$ = 13.91, P-value < 0.001), place of residence ($\chi2$ = 83.83, P-value < 0.001), place of delivery ($\chi2$ = 24.28, P-value < 0.001), mom's current age ($\chi2$ = 11.072, P-value < 0.001), number of antenatal visits ($\chi2$ = 44.51, P-value < 0.001), mom's short-rapid breaths ($\chi2$ = 4.90, P-value < 0.028), birth order ($\chi2$ = 49.86, P-value < 0.001), last birth cesarean section ($\chi2$ = 5782.72, P-value < 0.001), body mass index ($\chi2$ = 10.66, P-value < 0.015), whether the child is a twin ($\chi2$ = 15.30, P-value < 0.001), age of the mother at first birth ($\chi2$ = 64.23, P-value < 0.001), the mother's current age ($\chi2$ = 20.79, P-value < 0.001), and a history of terminated pregnancies ($\chi2$ = 19.73, P-value < 0.001). These statistical associations provide valuable insights into the factors contributing to the increasing prevalence of cesarean sections among women.

For babies born weighing 4 kg or more, the cesarean section rate was relatively higher (22.0%) than it was for babies weighing less than 2.5 kg (14.0%). Women with the richest wealth index were reported to have a higher cesarean section rate (29.0%) than other mothers. Those with higher education reported experiencing an increase in cesarean deliveries (45.0%) compared to other women with lower levels of education. Married women reported experiencing a higherrate of cesarean deliveries (19.0%) compared to those who were separated from their husbands (13.0%). It was observed that the cesarean section rate was higher for families with health insurance (17.0%) compared to those without health insurance cards (12.0%). Women from urban areas were more associated with cesarean sections (25.0%) compared to those from rural areas (14.0%).

The cesarean section rate was higher in private health facilities (53.0%) compared to 0.0% in home births and 16.0% in public health facilities. Mothers above the age of 25 years old were more associated with cesarean sections compared to those below 25 years old, with rates of 17.0% and 13.0%, respectively. It was observed that women who experienced short rapid breaths were at a higher risk of having a cesarean section, with a rate of 20.0%, compared to those without rapid breaths, who had a rate of 16.0%. It was found that mothers giving birth for the first time experienced a higher risk of having a cesarean section (19.0%) compared to those with their fourth birth and above (12.0%). Mothers who attended at least 4 antenatal visits were more likely to have a cesarean section, with a rate of 19.0%, compared to those who had no antenatal visits, with a rate of 4.0%. Similarly, mothers who were overweight were more likely to undergo a cesarean delivery, with a rate of 19.0%, compared to mothers with a normal weight, who had a rate of 14.0%.

It was observed that women who had a previous cesarean section were almost certain to have another cesarean section, with a rate of 100.0%, while mothers of single births had a cesarean section rate of 16.0%. Twin mothers were even more likely to undergo a c-section, with a rate of 33.1%. Mothers who gave birth for the first time at an age above 35 years old were more likely to have a cesarean section, with a rate of 45%, compared to those below the age of 20 years old, who had a rate of 11%. Additionally, women who had experienced previous problems with abortion were more likely to have a cesarean delivery, with a rate of 21%, compared to women without such issues, who had a rate of 15% (**Table 4**).

Table 4:Distribution of cesarean delivery according to potential predictors and its association

	No cesarean section (NO)	Cesarean section (Yes)		
Characteristics	N (%)	N (%)	Chi-square	P-value
(Birth weights/size of child at birth			14.29	**<0.001****
0-2.5 kgs	556(86)	87(14.0)		
25-4 kgs	3973(84)	777(16)		
4^{+} kgs	308(78)	89(22)		
Over had terminated pregnancy				
No	4096(85.0)	751(15.0)	19.73	**<0.001****
Yes	741(79.0)	202(21.0)		
Age of mother at first birth				
Below 20 years old	1140(89.0)	136(11.0)	64.23	**<0.001****
Age between 20 and 34	3673(82.0)	797(18.0)		
35 and above	24(55.0)	20(45.0)		
Mother's wealth index				
Poorest	1173(88.0)	161(12.0)	171.01	**<0.001****
Poorer	958(88.0)	134(12.0)		
Middle	971(87.0)	146(13.0)		
Richer	938(83.0)	186(17.0)		
Richest	797(71.0)	326(29.0)		
Anemia level of women				
Severe	4(67.0)	2(33,0)	1.48	0.823
Moderate	80(84.0)	15(16.0)		
Mild	207(84.0)	38(16.0)		
Not anemic	4539(84.0)	896(16.0)		
Mother's highest education level				
No education	509(90.0)	57(10.0)	219.98	**<0.001****
Primary	3254(86.0)	524(14.0)		
Secondary	912(79.0)	241(21.0)		

Higher	161(55.0)	130(45.0)		
Had any STI in the 12 months				
No	4621(84.0)	902(16)	1.23	**0.268**
Yes	216(81.0)	51(19.0)		
Mother's marital status				
Never union	525(84.0)	99(16.0)	29.19	**<0.001****[**]**
Married	2354(81.0)	550(19.0)		
Living with partner	1542(86.0)	244(14.0)		
Widowed	73(87.0)	11(13.0)		
Divorced	123(85.0)	21(15.0)		
Separated	220(89.0)	28(11.00)		
Place of residence				
Urban	960(75.0)	318(25.0)	83.83	**<0.001****
Rural	3877(86.6)	635(14.0)		
Place of delivery				
Home	106(100.0)	0(0.0)	24.28	**<0.001****
Hospital	61(91.0)	6(9.0)		
Another place	4670(83.0)	947(17.0)		
Mother's current age				
Women age between 15 and 25	887(87.0)	135(13.0)	20.79	**<0.001****
Women age between 25 and 35	2274(81.0)	521(19.0)		
Women age 35 and above	1676(85.0)	296(15.0)		
Number antenatal visits				
0	66(96.0)	3(4.0)	44.01	**<0.001****
1	189(94.0)	13(6.0)		
2-3	2282(85.0)	391(15.0)		
4^+	2282(85.0)	391(15.0)		
Mother's short-rapid breaths				
No	4360(84.0)	836(16)	4.90	**0.028****
Yes	474(80.0)	117(20.0)		
Covered by health insurance				
No	880(88.0)	125(12)	13.91	**<0.001****
Yes	3954(83.0)	827(17.0)		
HIV transmitted during pregnancy				
No	1266(83.0)	260(17.0)	0.44	**0.503**
Yes	3571(84.0)	693(16.0)		
Birth order				
First birth	1154(81.0)	279(19.0)	49.86	**<0.001****
Second births	1111(82.0)	244(18)		
Thirds birth	888(81.0)	209(19.0)		
Fourth births and over	1684(88.0)	221(12.0)		
Body Mass index of women				
0-18.49: Thin	83(85.0)	15(15.0)	10.66	**0.015****
18.49- 24.9: Normal	1645(85.0)	279(15.0)		
24.9-29.9: overweight	550(81.0)	132(19)		
30 weight and over: obese	2559(83.0)	527(17.0)		
Last birth cesarean section				

No	4837(100.0)	0(0.0)	5782.72	**<0.001****
Yes	0(0.0)	953(100.0)		
Child is twin				
Single birth	4779(84.0)	925(16)	15.30	**<0.001****
Second multiple	58(67.0)	28(33)		
Preceding birth interval				
Less than 24 months	482(85.0)	84(15.0)	1.06	**0.301**
Greater or equal 24 months	4355(83.0)	869(17.0)		

4.5. Predictors of the cesarean delivery in Rwanda

Multivariable logistic regression analysis is the process of employing regression techniques with a binary outcome variable and multiple explanatory factors to discern the most effective model for predicting the likelihood of success or failure in the binary response variable, given the values of numerous predictors [41]. In pursuit of the second objective of this study, multivariable logistic regression analysis was employed to identify the factors linked to the increasing incidence of cesarean deliveries, utilizing both a full and reduced model. As per the results of the multivariable analysis, factors such as maternal 'short rapid breath,' birth order, birth weight, maternal wealth index, number of antenatal care visits, maternal age at first birth, number of terminated pregnancies, place of residence, highest education level attained by women, maternal current age, and the presence of twins were identified as factors associated with the rising rate of cesarean deliveries among women.

The results of the study revealed several significant findings related to the likelihood of cesarean delivery. The results revealed that the odds of having a cesarean delivery were 0.44% less likely for women solely with their fourth or later pregnancy than for those with their first birth (OR=0.55, 95% CI [0.41-0.75], P<0.001).The odds of having a cesarean delivery were 90.4% higher in those who gave birth to a baby weighing 4 kgs or more than those who gave birth to a baby weighing less than 2.5 kgs (OR=1.904, 95% CI [1.34 – 2.69], P<0.001). Mothers who attended at least 4 antenatal care visits were 3.73 times more likely to have a cesarean section than those who did not attend any antenatal care visits (OR=3.73, 95% CI [1.15–12.11], P<0.028). Mothers from the richest wealth index categories were 49% more likely to give birth via cesarean section than the poorest women (OR=1.49, 95% CI [1.12–1.98], P<0.005).

Mothers with higher education were 2.72 times more likely to give birth by cesarean delivery than mothers with no education (OR=2.72, 95% CI [1.78–4.17], P<0.001), and mothers with secondary education were 49% more likely to undergo a cesarean delivery than non-educated mothers (OR=1.49, 95% CI [1.054-2.108], P<0.024). The odds of having a cesarean were 19.3% less among mothers from rural areas than mothers from urban areas (OR=0.807, 95% CI [0.66–0.98], P<0.031). Women with short rapid breaths were 1.307 times more likely to undergo a cesarean section than women without short rapid breaths (OR=1.305, 95% CI [1.04–1.63], P<0.020). Women who gave birth to twins were 94% more likely to give birth via cesarean than women who gave birth to a single baby (OR=1.94, 95% CI [1.51–2.49], P<0.001). Mothers aged between 25 and 35 years old were 30% more likely to have a cesarean section than mothers aged between 15 and 25 (OR=1.3, 95% CI [1.008–1.677], P<0.043). Mothers with their first birth at the age of 35 and above were 4.67 times more likely to have a cesarean section than mothers who gave birth for the first time below the age of 20 (OR=4.36, 95% CI [2.13–8.88], P<0.001). Although mothers who had their first born between the ages of 20 and 34 were 1.34 times more likely to give birth via cesarean section than those who did so before the age of 20 (OR=1.34, 95% CI [1.07-1.66], P<0.009). Mothers who had a history of pregnancy complications, such as previous abortions, were 48% more likely to give birth via cesarean section than mothers who did not have such complications (OR=1.48, 95% CI [1.23-1.78], P<0.001).

Table 5:Predictors of the cesarean delivery in Rwanda

	Reduced model				**Full model**			
Characteristics /variables	**Odd ratio (OR)**	**95 % of Confidence interval**		**P-value**	**Odd ratio (OR)**	**95% confidence interval**		**p-value**
		Lower	Upper			Lower	Upper	
Body Mass index								
0-18.49: Thin	1				1			
18.49- 24.9: Normal	0.84	0.47	1.51	0.557	0.83	0.46	1.502	0.557
24.9-29.9: Overweight	0.93	0.51	1.71	0.802	0.92	0.506	1.69	0.802
30 weight and over: obese	0.88	0.49	1.56	0.648	0.87	0.49	1.55	0.648
Birth order number								
First birth	1				1			
Second births	0.89	0.71	1.08	0.22	0.81	0.65	1.02	0.076
Thirds birth	1.007	0.80	1.26	0.951	0.92	0.72	1.19	0.561
Fourth births and over	0.59	0.47	0.75	**<0.001****	0.55	0.41	0.75	**<0.001****
Covered by health insurance								
No	1				1			
Yes	1.024	0.82	1.27	0.829	1.025	0.82	1.27	0.822
Birth weight /size child at birth								

0-2.5 kgs	1				1			
25-4 kgs	1.27	0.99	1.64	0.064	1.27	0.98	1.64	0.066
4^{+} kgs	1.89	1.33	2.68	**<0.001****	1.904	1.34	2.69	**<0.001****
Number of antenatal visits								
0	1				1			
1	1.47	0.40	5.42	0.558	1.49	0.41	5.47	0.547
2-3	3.15	0.97	10.21	0.056	3.16	0.98	10.25	0.055
4^{+}	3.72	1.14	12.08	**<0.029****	3.73	1.15	12.11	**<0.028****
Mother's wealth index								
Poorest	1				1			
Poorer	0.936	0.72	1.201	0.608	0.93	0.72	1.203	0.606
Middle	0.97	0.75	1.24	0.799	0.96	0.75	1.24	0.786
Richer	1.091	0.85	1.4	0.488	1.08	0.84	1.39	0.509
Richest	1.49	1.13	1.98	**<0.005****	1.49	1.12	1.98	**<0.005****
Mother's highest education leve l								
No education	1				1			
Primary	1.21	0.90	1.64	0.197	1.23	0.91	1.66	0.172
Secondary	1.46	1.04	2.07	**<0.030****	1.49	1.054	2.108	**<0.024****
Higher	2.77	1.81	4.23	**<0.001****	2.72	1.78	4.17	**<0.001****
Type place of residence								
Urban	1				1			
Rural	0.80	0.66	0.97	**<0.027****	0.807	0.66	0.98	**<0.031****
Mother's current age								
Women age between 15 and 25	1				1			
Women age between 25 and 35					1.300	1.008	1.677	**<0.043****
Women age 35 and above					1.22	0.88	1.706	0.226
Mother's marital status								
Never union	1							
Married	1.08	0.82	1.42	0.57	1.04	0.78	1.37	0.78
Living with partner	0.88	0.67	1.16	0.37	0.86	0.65	1.12	0.27
Widowed	1.075	0.53	2.16	0.84	1.03	0.51	2.08	0.92
Divorced	1.022	0.60	1.73	0.933	0.99	0.58	1.68	0.98
Separated	0.72	0.45	1.14	0.16	0.69	0.43	1.105	0.124
Age of mother at first birth								
Below 20 years old	1							
Age between 20 and 34	1.43	1.16	1.75	**<0.001****	1.34	1.07	1.66	**<0.009****

35 and above	4.67	2.38	9.18	**<0.001****	4.36	2.13	8.88	**<0.001****
Over had terminated pregnancy								
No	1				**1**			
Yes	1.48	1.23	1.78	**<0.001****	0.673	0.559	0.810	**<0.001****
Mothers 'short rapid breaths								
No	1				**1**			
Yes	1.307	1.04	1.64	**<0.019****	1.305	1.04	1.63	**<0.020****
Child is twin								
Single birth	1				1			
Second multiple	1.96	1.53	2.51	**<0.001****	1.94	1.51	2.49	**<0.001****

4.6. Importance features selection

As shown in **Figure 3**, the top ten reasons for the increased use of cesarean births among Rwandan mothers were identified using random feature selection approaches. These factors include wealth index, birth order number, body mass index, current marital status, mother's highest education level, number of antenatal care visits, coverage by health insurance, type of place of residence, respondent age, and birth weight.

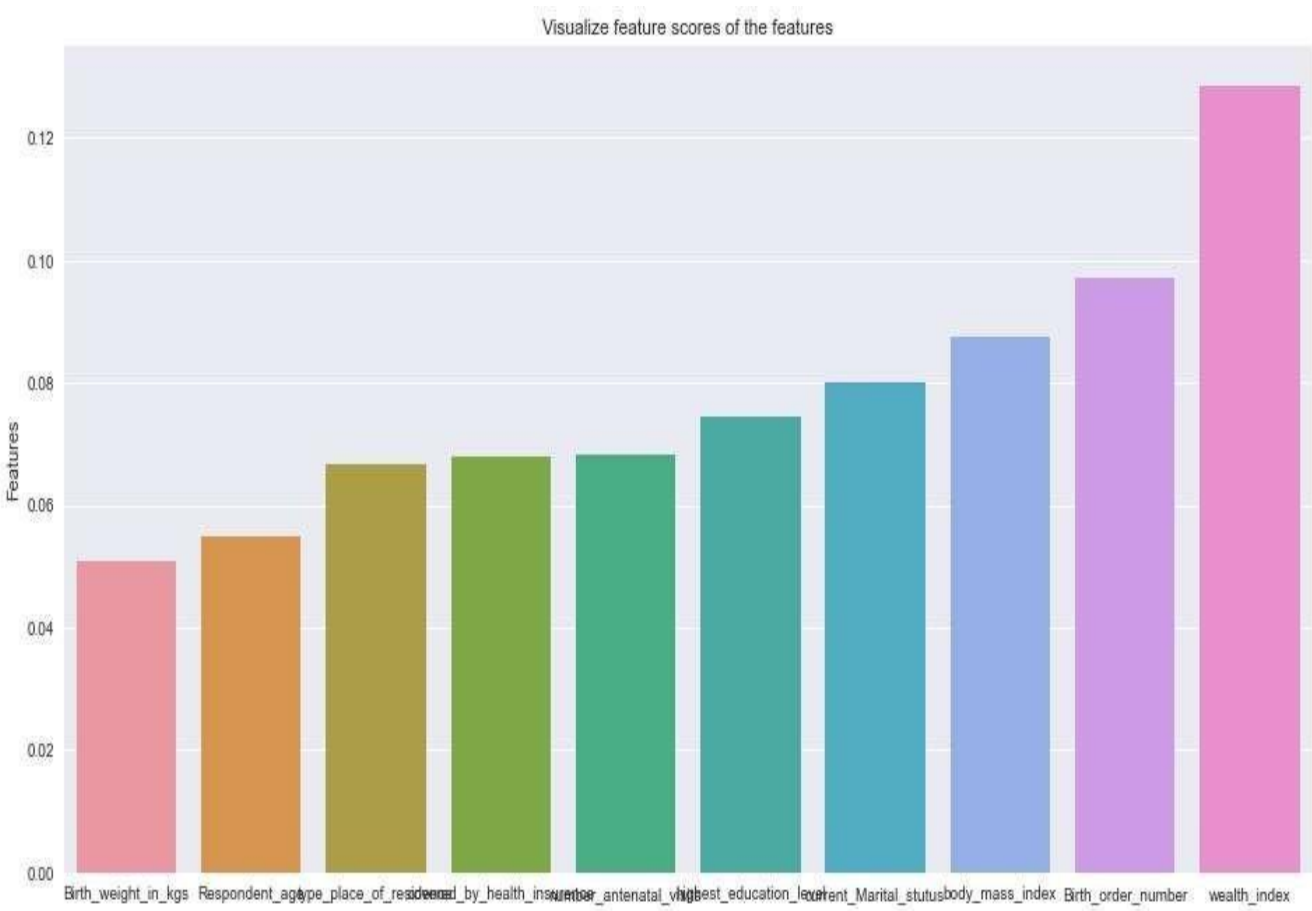

Figure 3:Random features selection importance

4.7. Predictive model of cesarean section

A prediction model for cesarean birth was developed using various machine learning algorithms, including logistic regression, decision trees, random forests, K-nearest neighbor classifiers, and support vector machines. The dataset was split into two parts, with 30% of the data reserved for testing and 70% of the data used for training. All machine learning models were trained on the training dataset, and the model's performance was evaluated using a set of metrics, including the confusion matrix, precision, recall score, F1 score, accuracy, and AUC score. These metrics were employed to assess how effectively the predictive model for cesarean delivery performed.

4. 8. Performance of predictive models of cesarean delivery

4.8.1. Logistics regression model

The predictive results indicate that the logistic regression model accurately predicted that 885 women would give birth through cesarean section, while 887 women would not. However, it incorrectly predicted that 557 women would undergo cesarean sections, when in fact 564 women would not. According to the research, the logistic regression model achieved aprecision of 61.0%, a recall score of 61.0%, an accuracy of 61.0%, and an area under the curve score of 66.2% for predicting the rate of cesarean sections.

Table 6:Evaluation metrics for logistic regression model

Evaluation metrics for logistic regression	Confusion matrix		Precision	F1 score	Recall score	Accuracy	UAC Score
	Predicted value		0.61	0.61	0.61	0.61	0.662
Actual value	Cesarean	No cesarean					
Cesarean	885	564					
No cesarean	557	887					

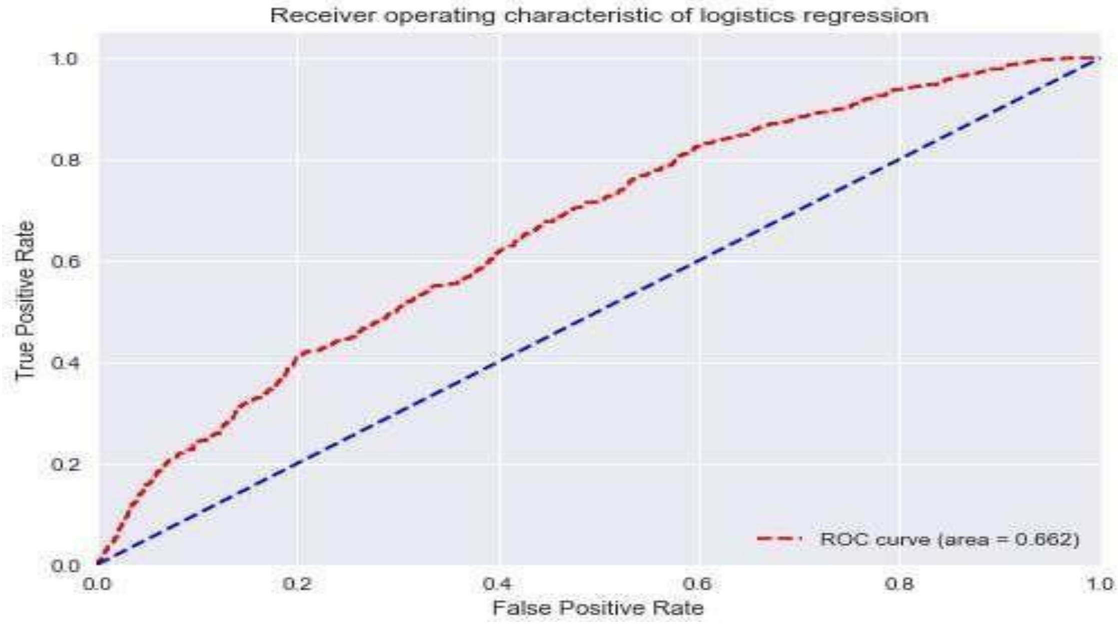

Figure 4:Receiver operating characteristics of logistic regression

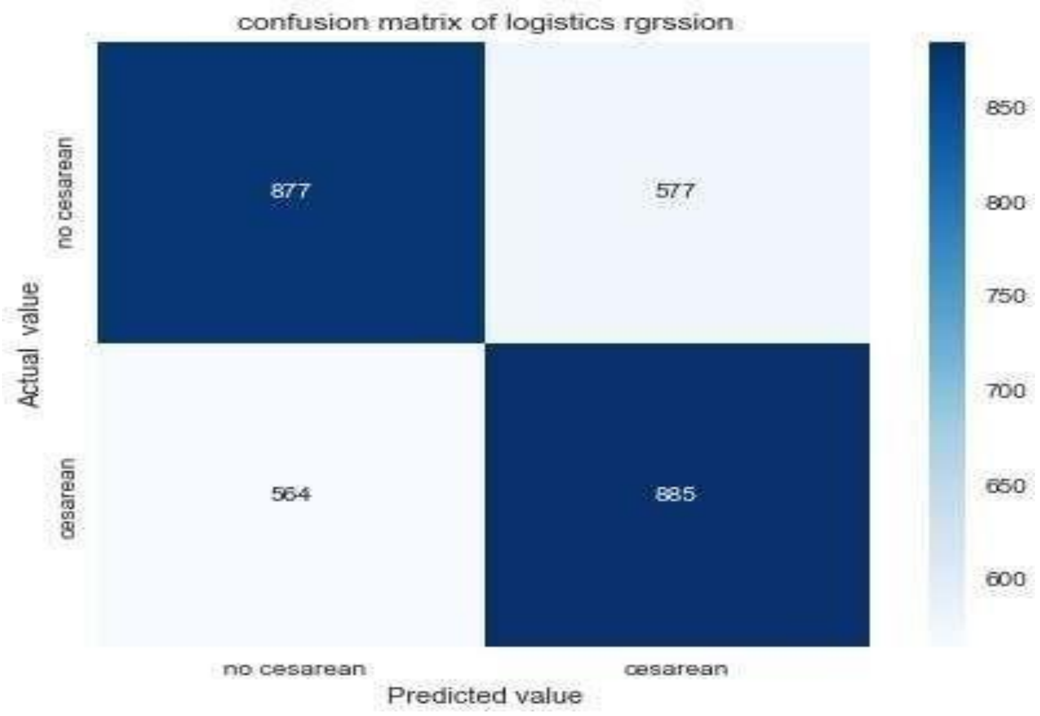

Figure 5:Confusion matrix of logistic regression model

4.8.2. Random forest classifier

The random forest model correctly predicted that 1236 women gave birth through cesarean delivery, while it correctly predicted that 1163 women did not. However, it made incorrect predictions for 291 women, falsely classifying them as having given birth by cesarean section when they did not, and for 947 women, incorrectly classifying them as not having given birth by cesarean section when they did. Additionally, the random forest model achieved the following performance metrics: accuracy: 83.0%, precision: 83.0%, AUC score: 83.0%, F1 score 90.2% and recall score 83.0%

Table 7:Evaluation metrics for random forest classifier

<table>
<tr><td rowspan="2">Evaluation metrics for random forest</td><td colspan="2">Confusion matrix</td><td>Precision</td><td>F1 score</td><td>Recall score</td><td>Accuracy</td><td>AUC Score</td></tr>
<tr><td colspan="2">Predicted</td><td rowspan="4">0.83</td><td rowspan="4">0.83</td><td rowspan="4">0.83</td><td rowspan="4">0.83</td><td rowspan="4">0.902</td></tr>
<tr><td>Observed</td><td>cesarean</td><td>No cesarean</td></tr>
<tr><td>Cesarean</td><td>1236</td><td>947</td></tr>
<tr><td>No cesarean</td><td>291</td><td>1163</td></tr>
</table>

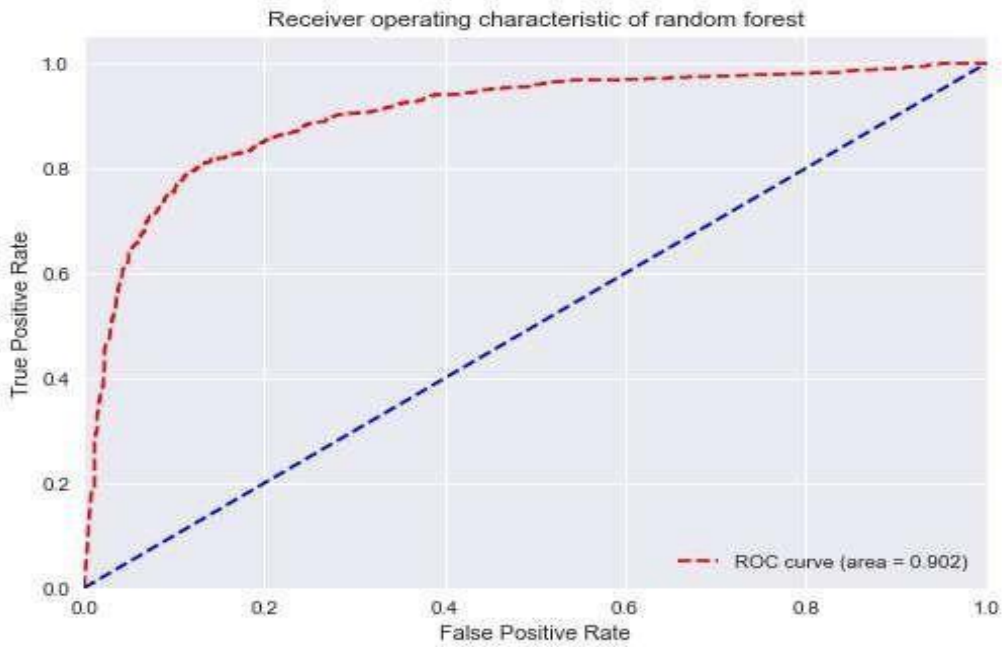

Figure 6:Receiver operating characteristic of random forest model

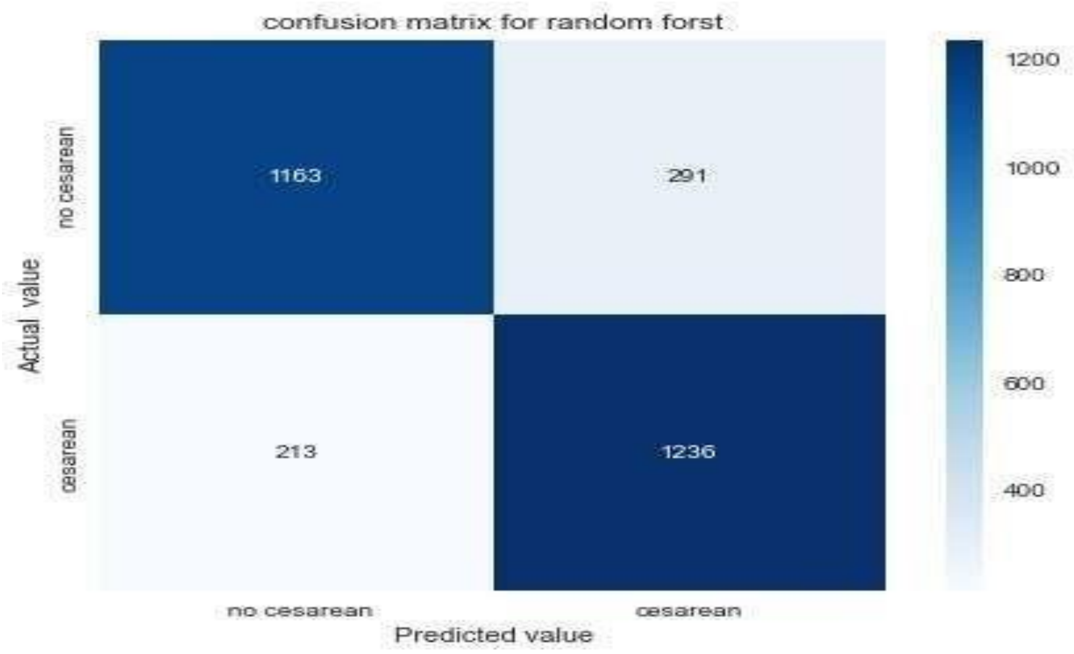

Figure 7:Confusion matrix of random forest model

4.8.3.Support vector machine

The results from the predictive model demonstrate that the support vector machine accurately predicted that 1001 women would give birth via cesarean section, while it correctly predicted that 827 women would not. However, the support vector machine made the mistake of predicting that 627 women would give birth through cesarean section when they did not, and it also incorrectly predicted that 448 women would not have cesarean deliveries when they actually did. The support vector machine was found to have an accuracy of 63.0%, an AUC score of 63.0%, a precision of 63.0%, a recall score of 63.0%, and an F1 score of 63.0% when it came to accurately predicting cesarean deliveries.

Table 8:Evaluation metrics for SVM

Evaluation metrics for support vector machine	Confusion matrix		Precision	F1 score	Recall score	Accuracy	AUC Score
	Predicted		0.63	0.63	0.63	0.63	0.63
Observed	Cesarean	No cesarean					
Cesarean	1001	448					
No cesarean	627	827					

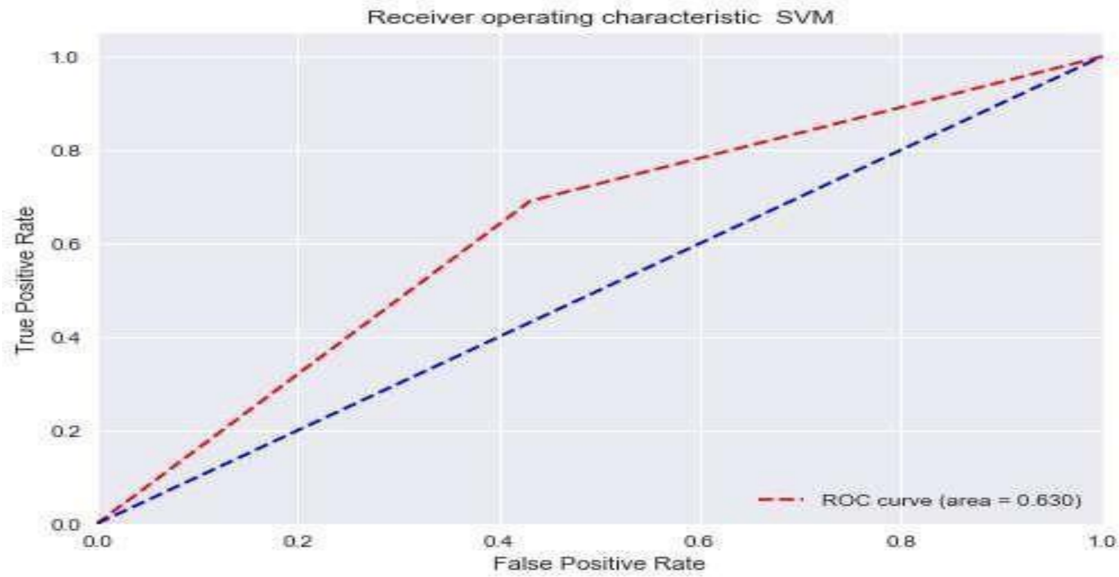

Figure 8:Receiver operating characteristic of SVM

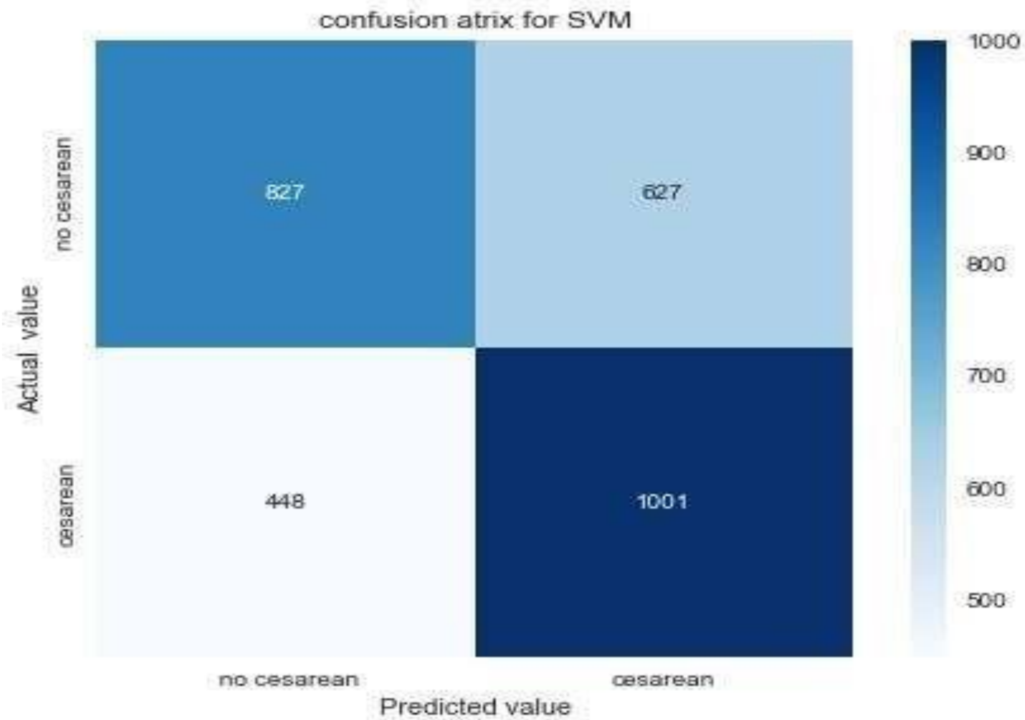

Figure 9:Confusion matrix of SVM

4.8.4.Decision tree classifier

The decision tree model correctly predicted that 1196 women would have a cesarean section, while 1091 women would not. It was also found that the decision tree model wrongly predicted 363 women would have a cesarean section when 253 women would not experience cesarean delivery at all. The decision tree model achieved a correct prediction rate of 79% accuracy, a 79.0% F1 score, a 79.0% precision score, an 80.7% AUC score, and a 79.0% recall score.

Table 9:Evaluation metrics for decision tree classifier

Evaluation metrics (decision tree classifier)	Confusion matrix		Precision	F1 score	Recall score	Accuracy	AUC Score
	Predicted		0.79	0.79	0.79	0.79	0.807
Observed	Cesarean	No cesarean					
Cesarean	1196	253					
No cesarean	363	1091					

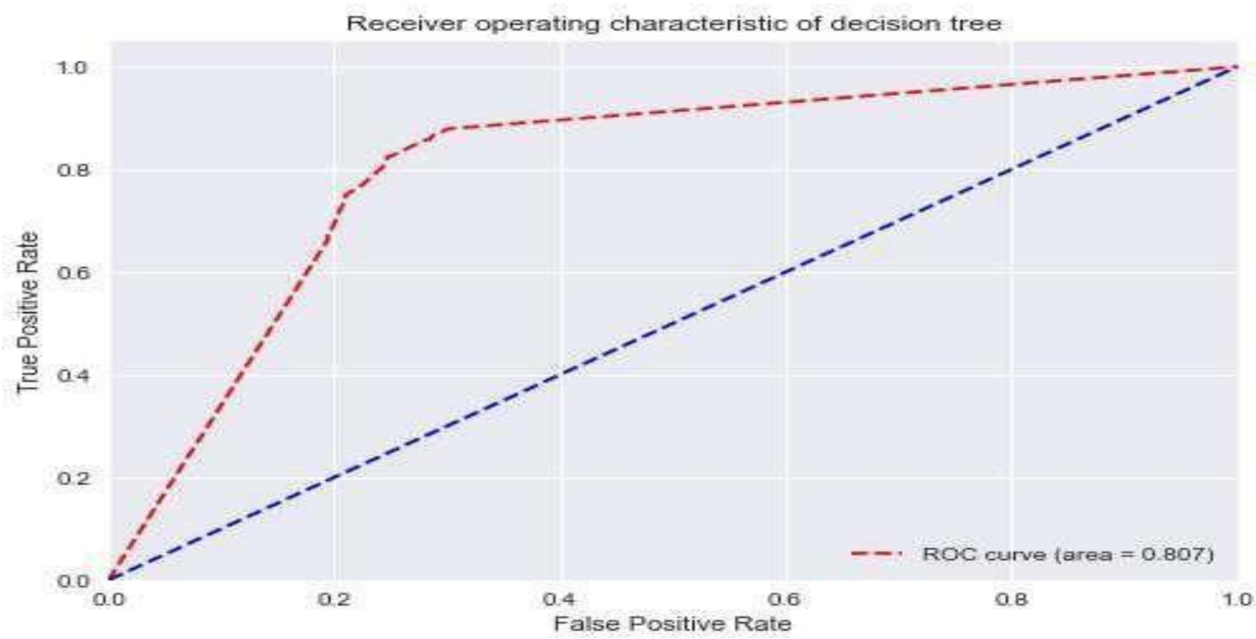

Figure 10:Receiver operating characteristic of decision tree model

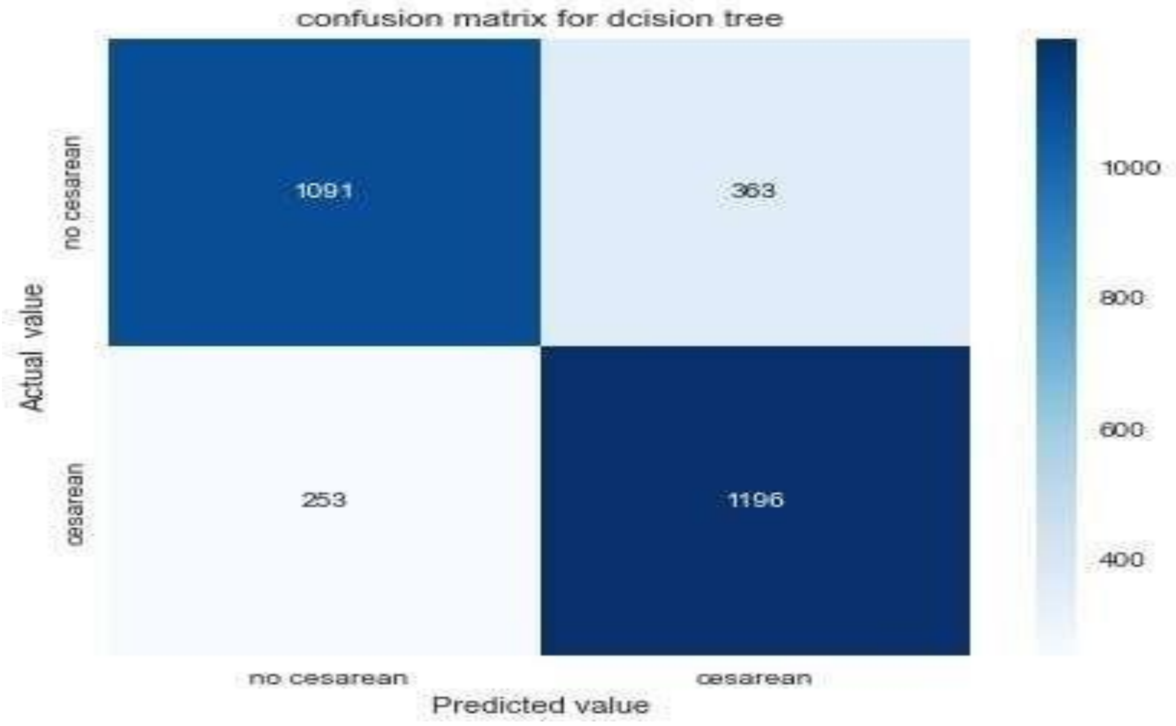

Figure 11:Confusion matrix of decision tree model

4.8.5.K nearest neighbor classifier

In this study, the k-nearest neighbor classifier correctly predicted that 1,256 women had experienced cesarean delivery, while 914 women had not. However, it made incorrect predictions for 540 women, falsely indicating that they had given birth through cesarean section when they had not. The k-nearest neighbor classifier achieved an accuracy of 75.0%, an AUC score of 81.4%, a precision rate of 76.0%, an F1 score of 74.0%, and a recall score of 75.0%.

Table 10:Evaluation metrics for KNN classifier

evaluation metrics (KNN)	Confusion matrix		Precision	F1 score	Recall score	Accuracy	AUC Score
	Predicted		0.76	0.74	0.75	0.75	0.814
Observed	Cesarean	No cesarean					
Cesarean	1256	193					
No cesarean	540	914					

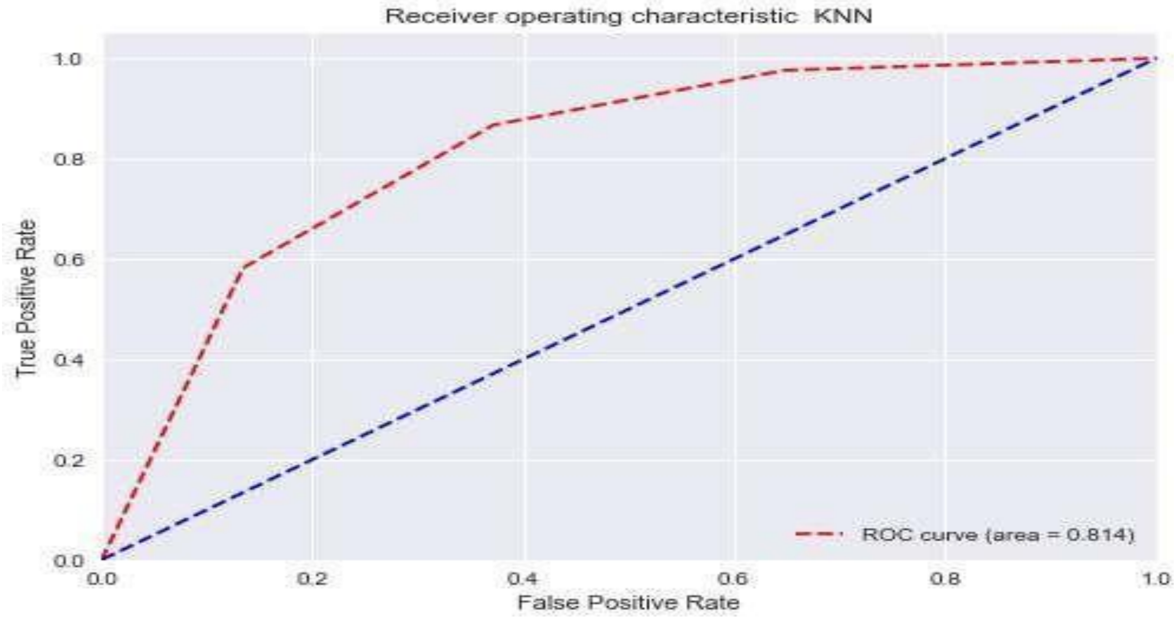

Figure 12:Receiver operating characteristic of KNN

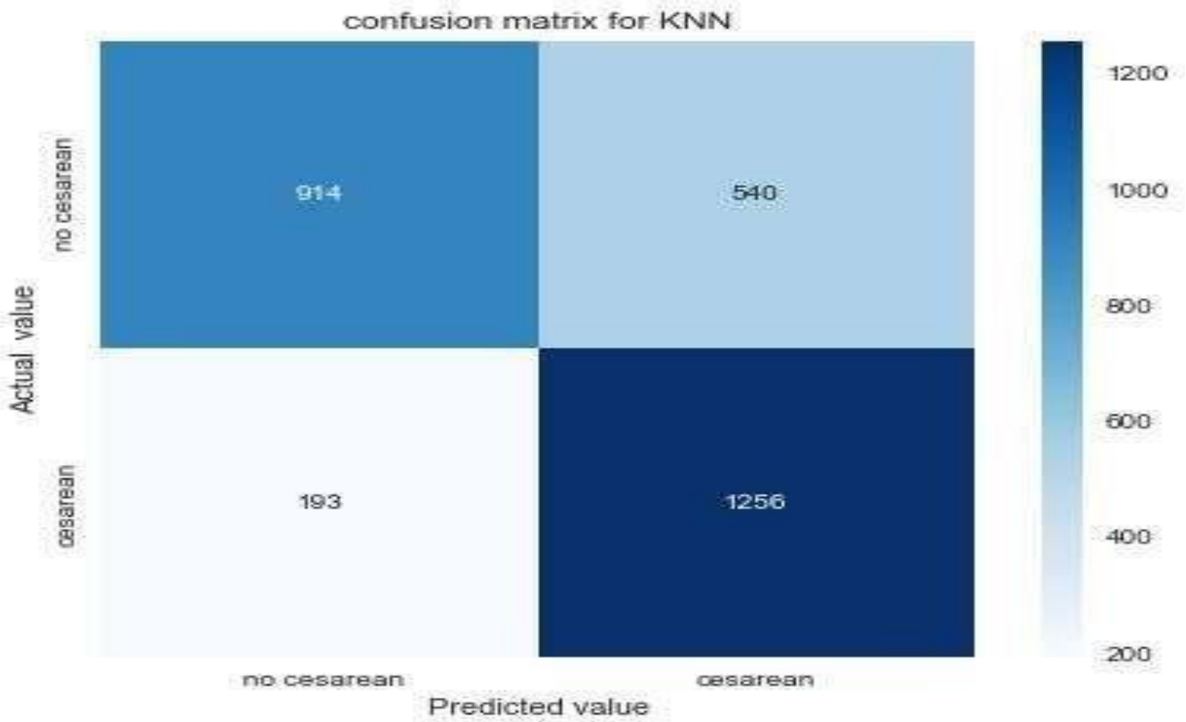

Figure 13:Confusion matrix of KNN

4.9.Comparison of all cesarean prediction models

In comparison to other predictive models employed in this study, the random forest model exhibited superior accuracy at 83.0% and achieved an area under the curve score of 90.20%, establishing itself as the most effective predictor of the factors driving the increase in cesarean deliveries among women. Our findings align with previous research, demonstrating that machine learning outperforms traditional models when it comes to predicting the factors contributing to cesarean deliveries, as opposed to mere estimation. These results are consistent with earlier studies that recognized the suitability of ML algorithms for predicting factors leading to the rise in cesarean deliveries among women [51].

Table 11:Summary of evaluation metrics of all models

Model	Precision score	F1 score	Recall score	Accuracy	AUC score
Logistics regression	**61.0%**	**61.0%**	**61.0%**	**61.0%**	**66.2%**
Decision tree	**79.0%**	**79.0%**	**79.0%**	**79.0%**	**80.7%**
Random forest	**83%**	**83%**	**83%**	**83.0%**	**90.2%**
K nearest neighbor	**76.0%**	**74.0%**	**75.0%**	**75%**	**81.4%**
Support vector machine	**63%**	**63%**	**63%**	**63%**	**63.0%**

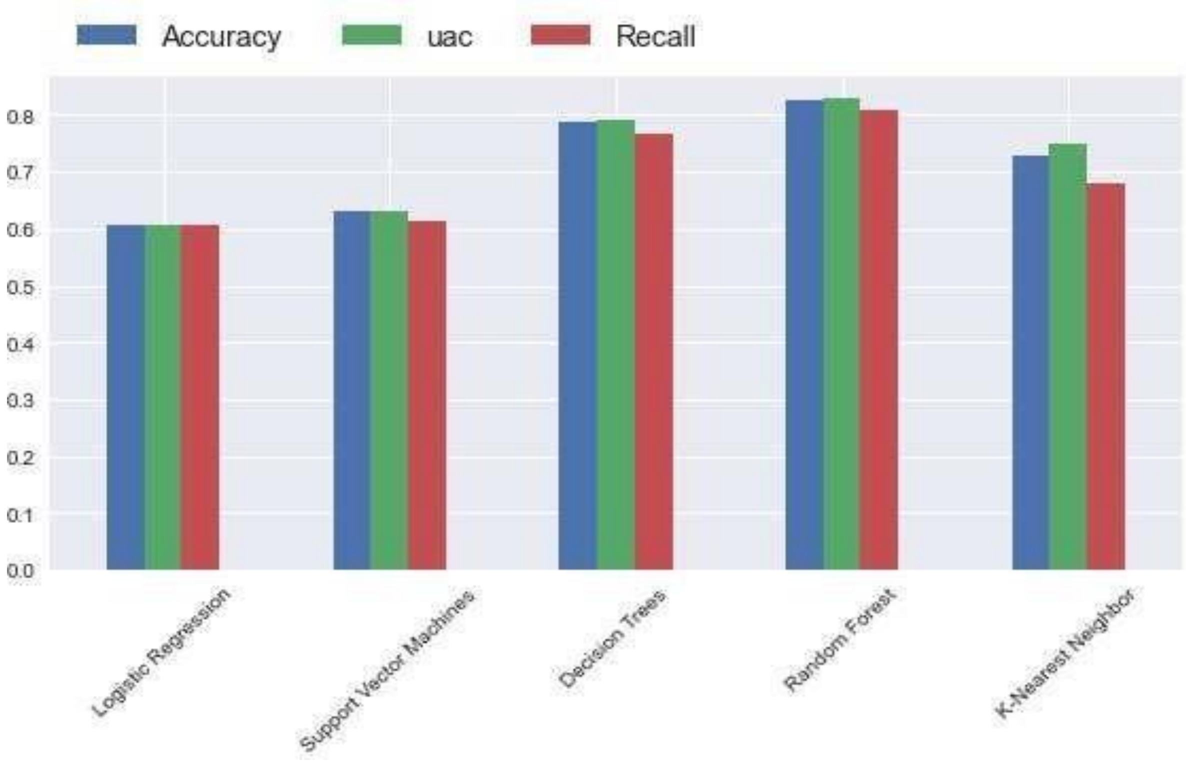

Figure 14:Graph of evaluation metric of all models

CHAPTER FIVE: DISSCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Discussion of findings

The study's primary objective was to utilize machine learning algorithms to predict cesarean sections among women. The secondary goal of the study was to identify the risk variables influencing cesarean deliveries in Rwanda. This was accomplished by employing multivariable binary logistic regression to pinpoint the risk factors associated with the increasing rate of cesarean sections in Rwanda.

The results of the multivariable logistic regression analysis in the current study suggest that eleven variables, namely birth order number, number of antenatal care visits, birth weight, mother's education level, mother's short-rapid breaths during labor, mother's wealth index, place of residence, mother's current age, the mother's age at the first birth, the number of pregnancies that were terminated, and whether the child is a twin, influence the occurrence of cesarean births in Rwanda. These findings align with previous research [52],[36] , indicating their reliability.

The first objective of the study was to develop predictive models for cesarean deliveries using five machine learning algorithms, including decision trees, random forests, logistic regression models, support vector machines, and k-nearest neighbors. The predictive results suggest that the random forest model has the best performance in predicting cesarean births. The findings of the current investigation were in line with those of a previous study that used machine learning algorithms to predict cesarean births and vaginal births, as highlighted in references [53],[54]. The random forest model correctly predicted that 1236 women would undergo cesarean sections, while 1163 women would not. It incorrectly predicted that 291 women would give birth through cesarean section, while 947 would not. The accuracy of the predictions made by the random forest model was also 83.0%, with a precision of 83.0%, an AUC score of 83.0%, an F1 score of 90.2%, and a recall score of 83.0%.

The logistic regression model correctly predicted that 885 women would have cesarean deliveries, while it predicted that 887 women would not. However, it incorrectly predicted that 557 women would undergo cesarean sections, when in fact, 564 did not. Our research findings indicate that the logistic regression model consistently predicted the cesarean section rate with a precision, recall, F1 score, and accuracy of 61.0%. These results align with earlier investigations [55],[51].

The support vector machine predicted that 1001 women would undergo a cesarean section based on the results of the prediction model. However, it also predicted that 827 women would not have a cesarean section. It incorrectly predicted that 448 women would give birth naturally, and it also incorrectly predicted that 627 women would have a cesarean delivery. Support vector machines were found to accurately predict cesarean sections with an accuracy of 63.0%, an AUC score of 63.0%, a precision of 63.0%, a recall score of 63.0%, and an F1 score of 63.0%. These results were consistent with the preceding study [46]. Additionally, the study results were similar to past studies that demonstrated SVM algorithms could classify cesarean delivery with perfect accuracy (100%) [56].

Predictive model results showed that KNN was correctly predicted that 1256 of women would have cesarean section and 914 women would not have it. It was wrongly predicted 540 women to give birth via cesarean section while 193 women did not. It was also found that k nearest neighbor correctly predicted 75.0% of accuracy, 81.4.0% of AUC score, 76.0%of precision, 74.0% of F1 score and 75% of recall score. These results were analogous with the previous studies [57].

The last objective was to rank the factors contributing to the rise in cesarean deliveries. Using the random forest feature selection technique, research findings revealed that the following factors were important contributors to the increase in cesarean sections among women: place of delivery, mother's short-rapid breaths, mother's current age, birth weight, number ofantenatal visits, coverage by health insurance, place of residence, body mass index, birth ordernumber, current marital status, mother's highest education level, and wealth index. These results were consistent with those of a previous study[58].

5.2. Conclusion

The overall objective of the project is to predict cesarean delivery among women using machine learning algorithms. To achieve the research goal, the RDHS 2029-20 individual records (IR) dataset were employed. Researchers utilized five machine learning (ML) methods,including logistic regression, decision trees, random forest classifier, K-nearest neighbors, andsupport vector machines, to create predictive models for cesarean births among Rwandan women.

In the current study, a multivariable binary logistic regression model was employed to identify the risk factors influencing cesarean sections in Rwanda. The identified risk factors included birth order, the number of antenatal care visits, birth weight, the mother's wealth index, place of residence, the mother's current age, the age of the mother at the first delivery, the number of pregnancies that were terminated, and whether or not the child was a twin. The results of our study demonstrated that machine learning algorithms excelled in developing precise predictive models for cesarean births [58].

The random forest model was determined to be the best classifier used for building a predictive model of cesarean delivery among Rwandan women. It achieved an accuracy of 83.0% and an area under the curve (AUC) score of 90.2%, outperforming the other models used in this research. This conclusion was reached based on various evaluation metrics for predictive models of cesarean section, including the confusion matrix, precision, accuracy, AUC, and F1 score.

This research concludes by encouraging additional institutions and future researchers to build on the work done in the current study. This can be achieved by employing various machine learning algorithms to predict the occurrence of cesarean sections among women usingdifferent RDHS datasets, in order to observe trends of cesarean section use and identify factorslinked to the rise in cesarean section rate among Rwandan women.

5.3. General recommendation

Based on the findings of the literature review, it is recommended that the RDHS collects all information related to women's deliveries, including factors such as fetal distress, a scarred uterus, breech presentation, and cephalopelvic disproportion. This comprehensive data collection is essential as cesarean deliveries may be planned if any of these conditions are present.

Based on the study's findings, the Ministry of Health and other health policy makers should develop guidelines and implementation plans to reduce the risk of pregnancy complications, such as shortness of breath and experiencing too many pregnancies, which may contribute to an increased rate of cesarean deliveries in Rwanda.

According to research findings, a significant number of women have undergone cesarean deliveries. Therefore, policymakers should encourage informed decision-making in this regard. Additionally, they can utilize a random forest model to predict the frequency of cesarean deliveries in the upcoming year, which would assist health planners in making informed decisions.

5.4. Further work

In the current study, we utilized the RDHS 2019–2020 dataset and employed five machine learning methods, including the decision tree classifier, random forest classifier, support vector machine, logistic regression, and k-nearest neighbor. To enhance the comparison of predictive models for cesarean birth across various RDHS datasets, we recommend that other researchers apply machine learning algorithms to predict cesarean deliveries among women using different RDHS datasets.

REFERENCES

[1] “Fetal Distress_ Causes, Symptoms, Diagnosis & Treatment.”

[2] A. Reid, “Breech Presentation - Risk Factors - Management - TeachMeObGyn.” 2018. [Online]. Available: http://teachmeobgyn.com/pregnancy/fetal-abnormality/breech-presentation/

[3] G. No, “Antepartum Haemorrhage,” no. 63, 2011.

[4] “Placental Abruption - StatPearls - NCBI Bookshelf.” [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK482335/

[5] “What happens when my baby is in transverse lie_ - BabyCentre UK.”

[6] D. Maharaj, “Assessing cephalopelvic disproportion: Back to the basics,” *Obstetrical and Gynecological Survey*, vol. 65, no. 6. pp. 387–395, 2010. doi: 10.1097/OGX.0b013e3181ecdf0c.

[7] “C-Section (Cesarean Section): Procedure, Risks & Recovery.” [Online]. Available: https://my.clevelandclinic.org/health/treatments/7246-cesarean-birth-c-section

[8] A. P. Betran, J. Ye, A. B. Moller, J. P. Souza, and J. Zhang, “Trends and projections of caesarean section rates: Global and regional estimates,” *BMJ Global Health*, vol. 6, no. 6. 2021. doi: 10.1136/bmjgh-2021-005671.

[9] A. P. Betrán, J. Ye, A. B. Moller, J. Zhang, A. M. Gülmezoglu, and M. R. Torloni, “The increasing trend in caesarean section rates: Global, regional and national estimates: 1990-2014,” *PLoS ONE*, vol. 11, no. 2. 2016. doi: 10.1371/journal.pone.0148343.

[10] H. Survey, *Rwanda*. 2019.

[11] National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], and ICF, *Rwanda Demographic and Health Survey 2019-20 Final Report*. 2021.

[12] P. M. Kibe and D. Shikuku, “Prevalence and Factors Associated With Caesarean Section in Rwanda : a Trend Analysis of Rwanda Demographic and Health Survey 2000 to 2019-20,” pp. 1–26, 2021.

[13] A. P. Betran *et al.*, “What is the optimal rate of caesarean section at population level? A systematic review of ecologic studies,” *Reprod. Health*, vol. 12, no. 1, 2015, doi: 10.1186/s12978-015-0043-6.

[14] L. Gibbons, J. M. Belizán, J. a Lauer, A. P. Betrán, M. Merialdi, and F. Althabe, “The Global Numbers and Costs of Additionally Needed and Unnecessary Caesarean Sections Performed per Year: Overuse as a Barrier to Universal Coverage,” *World Health Report (2010) Background Papers*. pp. 1–31, 2010. doi: 10.1017/CBO9781107415324.004.

[15] A. Shah *et al.*, “Cesarean delivery outcomes from the WHO global survey on maternal and perinatal health in Africa,” *International Journal of Gynecology and Obstetrics*, vol. 107, no. 3. pp. 191–197, 2009. doi: 10.1016/j.ijgo.2009.08.013.

[16] C. Burke and R. Allen, “COMPLICATIONS OF CESAREAN BIRTH CLINICAL

RECOMMENDATIONS FOR PREVENTION AND MANAGEMENT," vol. 45, no. 2, 2020.

[17] N. I. Khan, T. Mahmud, M. N. Islam, and S. N. Mustafina, "Prediction of Cesarean Childbirth using Ensemble Machine Learning Methods," *ACM Int. Conf. Proceeding Ser.*, no. November, pp. 331–339, 2020, doi: 10.1145/3428757.3429138.

[18] J. H. Wie *et al.*, "Prediction of Emergency Cesarean Section Using Machine Learning Methods : Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea," 2022.

[19] M. Z. Amin and A. Ali, "Performance Evaluation of Supervised Machine Learning Classifiers for Predicting Healthcare Operational Decisions Performance Evaluation of Supervised Machine Learning Classifiers for Predicting Healthcare Operational Decisions," no. January 2017, 2020, doi: 10.13140/RG.2.2.26371.25127.

[20] S. Dhakal-Rai, E. van Teijlingen, P. R. Regmi, J. Wood, G. Dangal, and K. B. Dhakal, "Factors contributing to rising cesarean section rates in South Asian countries: A systematic review," *Asian J. Med. Sci.*, vol. 13, no. 2, pp. 143–174, 2022, doi: 10.3126/ajms.v13i2.40904.

[21] M. Dikete, Y. Coppieters, P. Trigaux, Y. Englert, and P. Simon, "An analysis of the practices of caesarean section in sub-Saharan Africa: A summary of the literature," *Arch. Community Med. Public Heal.*, vol. 5, no. 2, pp. 077–086, 2019, doi: 10.17352/2455-5479.000058.

[22] C. Comparetto and F. Borruto, "Indications for cesarean sections," *Cesarean Sect. Proced. Complicat. Recover.*, pp. 39–65, 2012.

[23] S. Mumtaz, J. Bahk, and Y. H. Khang, "Rising trends and inequalities in cesarean section rates in Pakistan: Evidence from Pakistan Demographic and Health Surveys, 1990-2013," *PLoS One*, vol. 12, no. 10, pp. 1–14, 2017, doi: 10.1371/journal.pone.0186563.

[24] M. S. Id, T. Mehmood, and M. Aslam, "Identifying the factors associated with cesarean section modeled with categorical correlation coefficients in partial least squares," pp. 1–17, 2019.

[25] O. Kizito, "Determinants of caesarean section rates in private-not-for-profit healthcare facilities: St. Joseph's Hospital_ Kitovu," *Cogent Med.*, vol. 8, no. 1, 2021, doi:10.1080/2331205x.2021.1928939.

[26] F. E. Abebe, A. W. Gebeyehu, A. N. Kidane, and G. A. Eyassu, "Factors leading to cesarean section delivery at Felegehiwot referral hospital, Northwest Ethiopia: A retrospective record review," *Reproductive Health*, vol. 13, no. 1. 2016. doi: 10.1186/s12978-015-0114-8.

[27] E. Dankwah, S. Kirychuk, W. Zeng, C. Feng, and M. Farag, "Socioeconomic inequalities in the use of caesarean section delivery in Ghana: A cross-sectional study using nationally representative data," *International Journal for Equity in Health*, vol. 18, no. 1. 2019. doi:

10.1186/s12939-019-1063-6.

[28] B. D. Murray, “Macrosomia : Being Pregnant With a Large Baby.” pp. 1–21, 2020.

[29] G. J. Hofmeyr, J. F. Barrett, and C. A. Crowther, “Planned caesarean section for women with a twin pregnancy,” *Cochrane Database Syst. Rev.*, vol. 2015, no. 12, 2015, doi: 10.1002/14651858.CD006553.pub3.

[30] Y. S. Khader, A. Batieha, S. Khader, A. M. Jalam, and I. A. Amra, “Rate , causes , and risk factors of caesarean section in north of Jordan and its associated maternal and perinatal risks,” *Jordan Univ. Sci. Technol.*, 2015.

[31] M. Rouse, “What is machine learning? - Definition from WhatIs.com,” *WhatIs.com*. 2011. [Online]. Available: http://whatis.techtarget.com/definition/machine-learning

[32] Java Point, “Logistic Regression in Machine Learning - Javatpoint.” pp. 1–22, 2018. [Online]. Available: https://www.javatpoint.com/logistic-regression-in-machine-learning

[33] and I. 2021. National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], U. Rwanda Demographic and Health Survey 2019-20 Final Report. Kigali, Rwanda, and Rockville, Maryland, and N. and ICF., *Rwanda Demographic Health Survey, 2019-20*. 2021.

[34] NISR and MOH, *Rwanda Demographic and Health Survey 2019-20 Final report*. 2021.

[35] S. R. Manual, “Demographic and Health Surveys Methodology”.

[36] N. Burke *et al.*, “Prediction of cesarean delivery in the term nulliparous woman: results from the prospective, multicenter Genesis study,” *Am. J. Obstet. Gynecol.*, vol. 216, no. 6, pp. 598.e1-598.e11, 2017, doi: 10.1016/j.ajog.2017.02.017.

[37] A. Abubakari, G. Kynast-Wolf, and A. Jahn, “Prevalence of abnormal birth weight and related factors in Northern region, Ghana,” *BMC Pregnancy and Childbirth*, vol. 15, no. 1. 2015. doi: 10.1186/s12884-015-0790-y.

[38] K. Goyal, “Data Preprocessing in Machine Learning: 7 Easy Steps To Follow,” *upGrad*. 2021.

[39] “Python - Replace Missing Values with Mean, Median & Mode - Data Analytics.” [Online]. Available: https://vitalflux.com/pandas-impute-missing-values-mean-median-mode/

[40] J. Brownlee, “Random Oversampling and Undersampling for Imbalanced Classification,” *Machine Learning Mastery*. pp. 1–26, 2020. [Online]. Available: https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification/#:~:text=Random oversampling involves randomly selecting,them from the training dataset.

[41] “Chi-Squared Test of Independence - AI ML Analytics.” [Online]. Available: https://ai-ml-analytics.com/chi-squared-test-of-independence/

[42] G. C. Souza, “What is Logistic Regression? A Beginner’s Guide [2022],” *CareerFoundry*.

2021. [Online]. Available: https://careerfoundry.com/en/blog/data-analytics/what-is-logistic-regression/

[43] Nagesh Singh, "Decision Tree Algorithm, Explained - KDnuggets." 2018. [Online]. Available: https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html

[44] CFI, "Decision Tree - Overview, Decision Types, Applications," *Decision Tree*. 2020. [Online]. Available: https://corporatefinanceinstitute.com/resources/knowledge/other/decision-tree/

[45] Javapoint, "Machine Learning Random Forest Algorithm - Javatpoint." 2021. [Online]. Available: https://www.javatpoint.com/machine-learning-random-forest-algorithm

[46] Javatpoint, "Support Vector Machine (SVM) Algorithm - Javatpoint," *JavaTpoint*. 2018. [Online]. Available: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm

[47] P. Cunningham and S. J. Delany, "K-Nearest Neighbour Classifiers-A Tutorial," *ACM Comput. Surv.*, vol. 54, no. 6, 2021, doi: 10.1145/3459665.

[48] Javapoint-ConfusionMatrix, "Confusion Matrix in Machine Learning - Javatpoint." 2020. [Online]. Available: https://www.javatpoint.com/confusion-matrix-in-machine-learning

[49] G. Y. Lee, L. Alzamil, B. Doskenov, and A. Termehchy, "A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance," pp. 1–6, 2021, [Online]. Available: http://arxiv.org/abs/2109.07127

[50] P. Bhandari, "How to Find Outliers | 4 Ways with Examples & Explanation," *2021*. 2022. [Online]. Available: https://www.scribbr.com/statistics/outliers/

[51] J. H. Wie *et al.*, "Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea," *Life*, vol. 12, no. 4, 2022, doi: 10.3390/life12040604.

[52] H. C. C. De Souza *et al.*, "Development of caesarean section prediction models: Secondary analysis of a prospective cohort study in two sub-Saharan African countries," *Reprod. Health*, vol. 16, no. 1, pp. 1–11, 2019, doi: 10.1186/s12978-019-0832-4.

[53] A. De Ramón Fernández, D. Ruiz Fernández, and M. T. Prieto Sánchez, "Prediction of the mode of delivery using artificial intelligence algorithms," *Comput. Methods Programs Biomed.*, vol. 219, 2022, doi: 10.1016/j.cmpb.2022.106740.

[54] S. Rahman, M. I. Khan, M. S. Satu, and M. Z. Abedin, "Risk prediction with machine learning in cesarean section: Optimizing healthcare operational decisions," *Intell. Syst. Ref. Libr.*, vol. 192, no. October, pp. 293–314, 2021, doi: 10.1007/978-3-030-54932-9_13.

[55] L. Baoling and Z. Quanfu, "Development of a model for predicting success of trial of labor after cesarean delivery in China," *International Journal of Gynecology and Obstetrics*, vol. 143. pp. 655–656, 2018. [Online]. Available:

https://www.embase.com/search/results?subaction=viewrecord&id=L624605754&from=export%0Ahttp://dx.doi.org/10.1002/ijgo.12583

[56] A. A. Abdillah, Azwardi, S. Permana, I. Susanto, F. Zainuri, and S. Arifin, “Performance Evaluation Of Linear Discriminant Analysis And Support Vector Machines To Classify Cesarean Section,” *Eastern-European J. Enterp. Technol.*, vol. 5, no. 2–113, pp. 37–43, 2021, doi: 10.15587/1729-4061.2021.242798.

[57] S. Sodsee, “Predicting caesarean section by applying nearest neighbor analysis,” *Procedia Comput. Sci.*, vol. 31, pp. 5–14, 2014, doi: 10.1016/j.procs.2014.05.239.

[58] R. Meyer *et al.*, “Utilizing machine learning to predict unplanned cesarean delivery,” *Int. J. Gynecol. Obstet.*, no. April, pp. 1–9, 2022, doi: 10.1002/ijgo.14433.

APENDIX

Turnitin report

MASTERS DISSERTATION THESIS FOR AYINKAMIYE ROSE FINAL

ORIGINALITY REPORT

19%	15%	9%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dr.ur.ac.rw Internet Source	1%
2	www.researchgate.net Internet Source	1%
3	www.researchsquare.com Internet Source	1%
4	Submitted to University of Rwanda Student Paper	1%
5	etd.aau.edu.et Internet Source	<1%
6	erepository.uonbi.ac.ke Internet Source	<1%
7	Flavia Gladys Nakinobe, Charles Lwanga, Stephen Ojiambo Wandera, Ishmael Kalule-Sabiti, Kudzaishe Mangombe. "Caesarean delivery in Uganda: Do non-clinical factors explain the trend?", Journal of Biosocial Science, 2022 Publication	<1%

8	Submitted to University of Greenwich Student Paper	<1 %
9	worldwidescience.org Internet Source	<1 %
10	link.springer.com Internet Source	<1 %
11	ogma.newcastle.edu.au Internet Source	<1 %
12	repository.ju.edu.et Internet Source	<1 %
13	researchspace.ukzn.ac.za Internet Source	<1 %
14	www.panafrican-med-journal.com Internet Source	<1 %
15	Submitted to Bournemouth University Student Paper	<1 %
16	María Graciela Freyermuth, José Alberto Muños, María del Pilar Ochoa. "From therapeutic to elective cesarean deliveries: factors associated with the increase in cesarean deliveries in Chiapas", International Journal for Equity in Health, 2017 Publication	<1 %

18	www.mdpi.com Internet Source	<1 %
19	Luong Hoang Huong, Nguyen Hoang Khang, Le Nhat Quynh, Le Huu Thang, Dang Minh Canh, Ha Phuoc Sang. "A Proposed Approach for Monkeypox Classification", International Journal of Advanced Computer Science and Applications, 2023 Publication	<1 %
20	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
21	www.internationaljournalssrg.org Internet Source	<1 %
22	Submitted to International Health Sciences	<1 %
23	Ardeshir Sepehri. "Institutional setting and wealth gradients in cesarean delivery rates: Evidence from six developing countries", Birth, 2018 Publication	<1 %
24	Submitted to University of KwaZulu-Natal Student Paper	<1 %
25	bmcpregnancychildbirth.biomedcentral.com Internet Source	<1 %
26	eprints.bournemouth.ac.uk	

27	Nader Zad, Lisa A. Tell, Remya Ampadi Ramachandran, Xuan Xu et al. "Development of machine learning algorithms to estimate maximum residue limits for veterinary medicines", Food and Chemical Toxicology, 2023 Publication	<1 %
28	assets.researchsquare.com Internet Source	<1 %
29	scholar.mzumbe.ac.tz Internet Source	<1 %
30	www.medrxiv.org Internet Source	<1 %
31	hdl.handle.net Internet Source	<1 %
32	"Poster Abstracts", International Journal of Gynecology & Obstetrics, 2018 Publication	<1 %
33	Submitted to Liverpool John Moores University Student Paper	<1 %
34	www.ncbi.nlm.nih.gov Internet Source	<1 %

35	afribary.com Internet Source	<1 %
36	vital.seals.ac.za Internet Source	<1 %
37	www.thieme-connect.com Internet Source	<1 %
38	clinmedjournals.org Internet Source	<1 %
39	wiredspace.wits.ac.za Internet Source	<1 %
40	Laavanya Lokeesan, Elizabeth Martin, Yvette Miller. "Predictors of early initiation of breastfeeding in women who birth in government hospitals in Sri Lanka: an assessment across socio-demographic and clinical measures", Research Square Platform LLC, 2022 Publication	<1 %
41	Lilian Nuwabaine, Joseph Kawuki, Livingstone Kamoga, John Baptist Asiimwe et al. "Factors associated with anaemia among pregnant women in Rwanda: An analysis of the Rwanda Demographic and Health Survey of 2020", Research Square Platform LLC, 2022 Publication	<1 %
42	github.com	

43	pastebin.com Internet Source	<1 %
44	Abdulai Abubakari, Gisela Kynast-Wolf, Albrecht Jahn. "Prevalence of abnormal birth weight and related factors in Northern region, Ghana", BMC Pregnancy and Childbirth, 2015 Publication	<1 %
45	Submitted to Brunel University Student Paper	<1 %
46	Submitted to University of Melbourne Student Paper	<1 %
47	Submitted to University of Sydney Student Paper	<1 %
48	microdata.worldbank.org Internet Source	<1 %
49	pdf.usaid.gov Internet Source	<1 %
50	scholarspace.manoa.hawaii.edu Internet Source	<1 %
51	www.hindawi.com Internet Source	<1 %

52	Nazar P. Shabila. "Rates and trends in cesarean sections between 2008 and 2012 in Iraq", BMC Pregnancy and Childbirth, 2017 Publication	<1 %
53	web.iihe.ac.be Internet Source	<1 %
54	Vicky O'Dwyer, Jennifer L. Hogan, Nadine Farah, Mairead M. Kennelly, Christopher Fitzpatrick, Michael J. Turner. "Maternal mortality and the rising cesarean rate", International Journal of Gynecology & Obstetrics, 2012 Publication	<1 %
55	erepository.uonbi.ac.ke:8080 Internet Source	<1 %
56	hackmd.io Internet Source	<1 %
57	Mon H. Tun, Radha Chari, Padma Kaul, Fabiana V. Mamede et al. "Prediction of odds for emergency cesarean section: A secondary analysis of the CHILD term birth cohort study", PLOS ONE, 2022 Publication	<1 %
58	Zhenxing Li, Diming Zhang, Yuanjiang Li, Luohan Wang. "Visual Garment Collar Style Identification using Deep Neural Networks",	<1 %

61 Md. Akhtarul Islam, Mst. Tanmin Nahar, Md. Ashfikur Rahman, Sutapa Dey Barna, S.M. Farhad Ibn Anik. "Factors associated with cesarean delivery in Bangladesh: A multilevel modeling", Sexual & Reproductive Healthcare, 2022
Publication
<1 %

62 Menen Tsegaw, Bezawit Mulat, Kegnie Shitu. "Intimate partner violence and associated factors among reproductive age women in Liberia: a cross-sectional study using a recent Liberian demographic and health survey", BMC Women's Health, 2022
Publication
<1 %

63 Niamh C. Murphy, Fionnuala M. Breathnach, Naomi Burke. "Prediction of Caesarean Delivery", IntechOpen, 2019
Publication
<1 %

64 Submitted to University of Liverpool
Student Paper
<1 %

66	Florin Leon, Silvia Curteanu, Cătălin Lisa, Nicolae Hurduc. "Machine Learning Methods Used to Predict the Liquid-Crystalline Behavior of Some Copolyethers", Molecular Crystals and Liquid Crystals, 2007 Publication	<1%
67	Submitted to Higher Education Commission Pakistan Student Paper	<1%
68	Submitted to Intercollege Student Paper	<1%
69	Wei Zhang, Wei Li, Lei Han. "A Three-Dimensional Convolutional-Recurrent Network for Convective Storm Nowcasting", 2019 IEEE International Conference on Big Knowledge (ICBK), 2019	<1%
70	Submitted to Addis Ababa University Student Paper	<1%
71	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	<1%
72	Bismeen Jadoon, Ramez Mahaini, Karima Gholbzouri. "Determinants of over and underuse of caesarean births in the Eastern Mediterranean Region: an updated review", Eastern Mediterranean Health Journal, 2019 Publication	<1%

135	Joyce K. Edmonds. "Clinical Indications Associated With Primary Cesarean Birth", Nursing for Women's Health, 2014 Publication	<1 %
136	Munira Nargis Ananna, Md Belal Hossain. "Caesarean Delivery- a Pressured Choice for Women: A Meta-Analysis of 23 Cross-Sectional Studies", Dhaka University Journal of Science, 2023 Publication	<1 %
137	Sulochana Dhakal-Rai, Edwin van Teijlingen, Pramod Raj Regmi, Juliet Wood, Ganesh Dangal, Keshar Bahadur Dhakal. "Factors contributing to rising cesarean section rates in South Asian countries: A systematic review", Asian Journal of Medical Sciences, 2022 Publication	<1 %
138	Avijit Roy, Pintu Paul, Jay Saha, Margubur Rahaman, Nanigopal Kapasia, Bikash Barman, Pradip Chouhan. "Assessing Socio-Demographic Factors of Cesarean Delivery in India: Insights From National Family Health Survey-4", Research Square, 2020 Publication	<1 %

139	François Niragire, Celestin Ndikumana, Marie Gaudence Nyirahabimana, Cyprien Mugemangango. "Child stunting and associated risk factors in selected food-insecure areas in Rwanda: an analytical cross-sectional study", Pan African Medical Journal, 2022 Publication	<1%
140	Kizito Omona, Dorothy Babirye. "Maternal near misses (MNM) and their determinants among women who sought obstetric care from fort portal regional referral hospital, Western Uganda", Cogent Public Health, 2022 Publication	<1%
143	Sana Al Azwari. "Predicting Myocardial Rupture after Acute Myocardial Infarction in Hospitalized Patients using Machine Learning", 2021 National Computing Colleges Conference (NCCC), 2021 Publication	<1%
144	Yasir B. Nisar, Michael J. Dibley. "Iron/folic acid supplementation during pregnancy prevents neonatal and under-five mortality in Pakistan: propensity score matched sample from two Pakistan Demographic and Health Surveys", Global Health Action, 2016 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		