AFRICAN CENTRE OF EXCELLENCE IN DATA SCIENCE

COLLEGE OF BUSINESS & ECONOMICS

MACHINE LEARNING PREDICTION OF LOW BIRTH WEIGHT IN KENYA USING MATERNAL RISK FACTORS

By

SHARON JEPKORIR SAWE

Registration Number: 220000140

A dissertation submitted in partial fulfilment of the requirements for the degree of Master of Data Science in Biostatistics

University of Rwanda, College of Business and Economics

Supervisor: Dr Dieudonné Muhoza

September 2022

Declaration

I declare that this dissertation entitled **Machine Learning Prediction of Low Birth Weight in Kenya using Maternal Risk Factors** 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.

Name: Sharon Jepkorir Sawe

Registration Number: 220000140

Signature: 

Date: 21st September 2022

Approval sheet

This dissertation entitled **Machine Learning Prediction of Low Birth Weight in Kenya using Maternal Risk Factors** written and submitted by **SHARON JEPKORIR SAWE** in partial fulfilment of the requirements for the degree of Master of Science in Data Science majoring in **Biostatistics** is hereby accepted and approved. The rate of plagiarism tested using Turnitin is **19 %** which is less than 20% accepted by the African Centre of Excellence in Data Science (ACE-DS).

_____

Supervisor: Dr. Dieudonné Muhoza

_____

Head of Training: Dr. Ignace Kabano

Dedication

I dedicate this work to my family and to every person who constantly helped me during conceptualization and write up.

Acknowledgements

I wish to express my sincere gratitude to my supervisor Dr. Dieudonné N. Muhoza for his guidance, positive criticism and a lot of patience during his supervision of this thesis.

My sincere gratitude also goes to my dear parents Mr. Daniel Sawe and Mrs Teresa Sawe, for their love, support and prayers which made everything worthwhile.

I must express lots of appreciation to the entire staff of Africa Centre of Excellence including Dr Charles Ruranga, Dr Ignace Kabano and other members of staff at large for their support. Finally, I wish to thank my friends and classmates for their constant assistance during write up of this thesis.

Abstract

A new born's health is a primary factor that determines the overall health of a human being and its life expectancy. Therefore, its health should be monitored not only after birth but also when the baby is still growing in the womb. Birth weight is one of the crucial aspects to be observed. Low birth weight is among the main problems that new borns face. Low birth weight (LBW) is the weight at birth less than 2500g as defined by the World Health Organization. A global estimate of 15 to 20 percent of total live births are low birth weight representing over twenty million births every year. In Kenya, the rate of children born with low weight is 8 percent. Several methods have been used to measure and approximate birth weight in clinical practice including obstetric ultrasound, symphysio-fundal height measurements and abdominal palpation. However, these methods are associated with reliability and accuracy challenges therefore, calling for more robust methods. This research aimed at creating a machine learning model for predicting low birth weight using the maternal risk factors that have been found to be associated with low birth weight. Secondary data from the 2014 Kenya Demographic Health Survey was utilized where the variables were extracted from the births recode file. The study population included mothers between the age of 15 to 49 years. The machine learning algorithms employed were logistic regression, decision trees, random forest, support vector machines, gradient boosting and xtreme gradient boosting. Using performance evaluation metrics namely; accuracy, precision, recall, F1 score, and ROC-AUC, the random forest model was found out to be the most robust with 0.956679 accuracy, 0.956831 precision, 0.956679 recall an F1 score of 0.95666 and an AUC of 0.988. In addition, variable importance was performed using the random forest approach to ascertain the maternal risk factors that are the most important to predict low birth weight. It was found out that mother's weight was the most important variable for predicting low birth weight. The other important variables found were; mothers height, mother's age and the number of antenatal visits attended by the mother during pregnancy. Machine learning techniques are increasingly being used to provide information to guide health policy. This research merits further modelling, research and more consultation.

Keywords

Machine learning, birth weight, low birth weight, maternal risk factors, prediction, algorithm

Table of contents

List of tables

List of figures

List of symbols and acronyms

LBW-Low Birth Weight

WHO-World Health Organization

UNICEF-United Nations Children's Fund

WHA-World Health Assembly

UHC-Universal Health Coverage

SMOTE-Synthetic Minority Oversampling Technique

XGB-Xtreme Gradient Boosting

SVM-Support Vector Machine

EDA-Exploratory Data Analysis

ROC-Receiver Operating Characteristic Curve

AUC-Area Under the Curve

MRE-Mean Relative Error

IUGR-Intrauterine Growth Restriction

BMI-Body Mass Index

SGA-Small for Gestational Age

PAR-Population Attriburable Risk

IQ-Intelligence Quotient

KDHS-Kenya Demographic and Health Surveys

HIV-Human Immunodeficency Virus

CHAPTER ONE: INTRODUCTION

1.1 Background

The health of a new born is a crucial element in the general health of a nation and even global health. It is the primary factor that determines the overall health of a human being and the life expectancy. Therefore, a baby's health and well-being should be monitored not only after birth but also when the baby is still growing in the womb. One of the aspects that should be observed before the baby is born is its weight. Birth weight is the new born baby's first weight measured immediately after being born within the first hour before occurrence of significant loss of weight due to postnatal effects [1]. A new-born's weight signifies a lot about the future health and survival of the baby. Therefore, it is advisable to know whether the baby is going to have normal weight or low weight during birth in order to make early interventions before birth. Low birth weight (LBW) is the weight below 2500 grams that is measured at birth as the World Health Organization defines [2].

The major causes of low birth weight are preterm birth and growth faltering in the womb / Intrauterine growth restriction (IUGR) [3]. Preterm births occur in a period below 259 days since the start of the last menstruation of a woman preceding conception or before completing a gestation period of 37 weeks as WHO defines [4]. On the other hand, Intrauterine growth restriction is the below normal rate of foetal growth with respect to the growth potential of the infant in terms of its gender and race. An infant's normal birth weight ranges between the 10^{th} and 90^{th} percentile with exclusion of malnutrition and growth retardation features [5]. Low birth weight is highly associated with neonatal and foetal morbidity and mortality, slow cognitive development and growth, and there after in life they may develop chronic diseases [1].

Low birth weight prevalence is estimated regionally to be 9 percent in Latin America, 28 percent in South Asia and 13 percent in Sub-Saharan Africa. However, these rates could probably be an underestimate because, not all

women get access to giving birth in hospitals therefore these deliveries are not recorded since they deliver at home. Moreover, deliveries that occur in small clinics may go unreported by public official figures [2].

Globally, a prevalence reduction of low birth weight by 30 percent in 2025 has been targeted by the World Health Assembly [2]. In Kenya, Universal Health Coverage (UHC) is one of the big 4 agenda of which new born health is among its important indicators. The rate of children born in Kenya with low birth weight is 8 percent according to a report by [6]. This rate is still alarming and therefore appropriate solutions towards this problem should be sought.

Several methods have been used to measure and approximate birth weight in clinical practice. The methods include; obstetric ultrasound, symphysio-fundal height measurements and abdominal palpation. Obstetric ultrasound stands out to be the most reliable method of examining the growth of the foetus. However, ultrasound is not easily accessed in low-resource areas and poor communities. Therefore, the other two methods are applied which are not very reliable in terms of accuracy [7]. Moreover, training for ultrasound is very crucial. Unskilled ultrasound sonographers might lead to inaccurate foetal weight measurements. Therefore, good training is paramount [8].

Due to these challenges, another route towards tackling LBW estimation should be taken. Robust methods to estimate birth weight and predict low birth weight should be taken into high consideration. This is because early detection allows for proper and effective obstetric interventions. Recently, data mining methods particularly machine learning have been discovered to be of great help in predicting low birth weight.

Prediction of low birth weight can be done by using machine learning techniques particularly supervised learning approaches. Machine learning is a computer science subfield that involves pattern recognition and computational learning. It inspects the construction and the study of algorithms which can make predictions by learning from data [9]. Supervised learning also known as classification is a paradigm of machine learning that is used to acquire the system's information based on a set of labelled input-output samples. The goal is to predict output given new inputs [10]. In this research, the data will be labelled in such a way that the output variable low birth weight is binary in nature. Therefore, the machine learning task will follow a supervised learning approach.

In Kenya particularly, several studies have been conducted to identify the risk factors associated with low birth weight. However, limited research has been geared towards predicting babies at risk of being born with low weight. Therefore, the identified maternal risk factors from previous researches can be consolidated and used to develop a machine learning model to predict low birth weight.

This study aims to use the 2014 Kenya demographic health survey data to perform predictive modelling of low birth weight in unborn babies using mainly the maternal risk factors of low birth weight.

1.2 Problem statement

Low birth weight is still a major challenge globally and nationally. Several interventions have been put in place but it remains a public health problem. It is a global war whereby it led to the 2012 global nutrition targets. A reduction by 30 percent in low birth weight between 2012 and 2025 worldwide was adopted by the member states during the 65th World Health Assembly (WHA). However, up to date the globe is still far from accomplishing this objective. The 2000-2015 report on global trend in low birth weight prevalence by WHO and UNICEF reports that in 2010 to 2015, the reduction in low birth weight prevalence was slow in comparison to 2000 to 2009. It further reports that if the current annual average rate of reduction of 1.00 percent yearly continues, the low birth weight prevalence that was projected to be 10.5 percent, would be 13.2 percent by 2025 [3].

Worldwide, approximately 15 percent to 20 percent of total live births are of low weight. This represents over twenty million births every year [2]. Moreover, 91 percent of the low birth weight livebirths are from countries of middle and low income majorly South Asia with 48 percent and sub-Saharan Africa with 24 percent [11]. A recent study carried out in West Bengal India revealed that 21.49 percent of born babies had low weight. The risk was highly likely for women with less than 20 years of age and a BMI of less than 18.5 kg/m^2. The odds were higher for women having a weight below 45 kg and a height below 150 cm, those who never attended antenatal care visits and those who never took iron folic acid tablets and an extra diet while pregnant. Moreover, the situation was higher in illiterate women, those who lived in rural areas and those who came from low social economic families [12].

A study done in Benin Nigeria at a traditional birth home depicted that, the prevalence of low birth weight was 6.3 percent. This figure was affected significantly by gestational age, maternal age, maternal height, time of registration and maternal anaemia [13]. In rural Cameroon, the LBW prevalence was approximated to be 6.1 percent. The study indicated that babies born with low weight possess a higher probability of being still born or asphyxiated at the 5th minute as compared to heavier babies [14]. Another study conducted in Africa on prevalence and its association to maternal body weight showed the prevalence of LBW in Uganda, Malawi, Senegal, Ghana and Burkina Faso was 10 percent, 12.1 percent, 15.7 percent, 10.2 percent and 13.4 percent respectively where underweight mothers had a bigger probability of giving birth to babies with a lower weight than women of normal weight except in Ghana [15]. The pooled prevalence of LBW in Ethiopia was found to be 14.1 percent. Factors that were highly associated with low birth weight were; female babies, prematurity, not attending prenatal care, pregnancy-induced hypertension and mothers from rural areas [16]. In Tanzania, a recent study found out that the incidence of LBW was 7.1 percent whereas the recurrent prevalence was estimated to be 28.1 percent. The important low birth weight recurrence predictors were; preterm birth, less than 4 antenatal care visits, HIV positive status and pre-eclampsia during pregnancy [17]. The recurrence of Low birth weight is described as repetition of a low birth weight delivery in a subsequent pregnancy [18].

Kenya being the centre of this research, low birth weight is still a major concern in the nation's health. Using the 2009 Kenya demographic and health survey, WHO and UNICEF reported that the low birth weight estimates were 11 percent and 6 percent. Moreover, in central province Kenya alone, the prevalence was estimated as 5.5 percent [19]. Another study carried out in Olkalau district hospital central Kenya depicted that the prevalence of low birth weight was 12.3 percent. LBW was found to be associated with delivery in a previous birth, premature rapturing of membranes, premature births and female infants [20]. In Pumwani maternity hospital, the situation was even worse. The prevalence of LBW was found out to be 32.8 percent. It was associated with number of meals taken per day while pregnant, vaginal bleeding, maternal anaemia, hypertension, pelvic pressure, abdominal pain and lower back ache [21]. The prevalence was also high (29 percent) in a study carried out at Coast General Hospital Mombasa County. It was discovered that low birth weight was significantly associated with caesarean section birth, a previous low birth weight delivery, twin birth and less than four antenatal care visits.

Moreover, women with college education level and a normal concentration of haemoglobin had a lower risk of giving birth to low weight babies [22].

Low birth weight is greatly associated with a lot of difficulties and challenges in the life of a child which might extend to adulthood. Yearly, 1.1 million babies succumb from preterm birth which is a major cause of low birth weight [2]. A particular study found out that babies who are born small for gestational age (SGA) are associated with lower performance in school at 12 and 18 years [23]. The baby's overall nutrition status is also affected by low birth weight. Most of these babies end up being malnourished. A certain research revealed that LBW was associated with higher odds of underweight, stunting and wasting. For overall SGA, the population attributable risk (PAR) for childhood wasting and stunting was found out to be 30 percent and 20 percent respectively [24]. Low birth weight consequences may proceed to adulthood and result to onset of chronic diseases such as diabetes and obesity. It was discovered that women born with low weight had higher levels of insulin, diabetes, metabolic syndrome, fasting and plasma glucose. A similar trend which was insignificant was found out in male adults [25]. Several studies have also reported that low birth weight is associated with low IQ than children born with normal weight. Gradient relationship has been demonstrated for different levels of LBW and IQs. A discrepancy of 10 to 11 points for IQ was revealed between low birth weight and normal birth weight children [26]. Low birth weight is also found to be associated with adulthood mortality. For all-cause mortality, a 6 percent lower risk of mortality was observed per one extra higher birthweight kilogram for both men and women [27]. Another research on effects of low birth weight on number of nephrons and long-term renal health discovered that Intrauterine growth restriction and preterm birth which are the major causes of low birth weight contribute to reduced number of nephrons in the body therefore posing a high risk of long-term renal disease [28].

All these challenges and consequences that come with low birth weight are due to the fact that the birth weight of new born babies are not predicted early enough in order to make necessary interventions to mitigate the problem in case a low birth weight instance is foreseen. Therefore, this research aims at building a model that accurately predicts low birth weight using machine learning techniques.

1.3 Research objectives

1.3.1 General Objective

To predict low birth weight in Kenya using machine learning techniques

1.3.2 Specific Objectives

1. To evaluate the performance of the machine learning techniques on low birth weight prediction using performance metrics
2. To determine the most robust machine learning technique for predicting low birth weight
3. To identify the most important variables for predicting low birth weight

1.4 Research questions

1. What is the performance of the machine learning techniques on low birth weight prediction using performance metrics?
2. What is the most robust machine learning technique for predicting low birth weight?
3. Which are the most important variables for predicting low birth weight?

1.5 Scope and significance of the study

This research was based on a Kenyan setting using the recent 2014 Kenya demographic health survey data. The KDHS is a national survey that reflects the national situation of the subject under study. The study variables were drawn from the births recode KDHS file to be used for predicting the risk of low birth weight using machine learning techniques.

This study will bring a lot of impact in the Kenyan health sector particularly obstetrics health. The findings will help the obstetrics and gynaecology departments in the health sector to be in a better position to make appropriate and better interventions concerning babies at risk of low birth weight. In addition, it will contribute to the general body of health related research to the academic world. It will create a platform to other researchers to explore more on the same field and discover more and new findings.

1.6 Conceptual framework

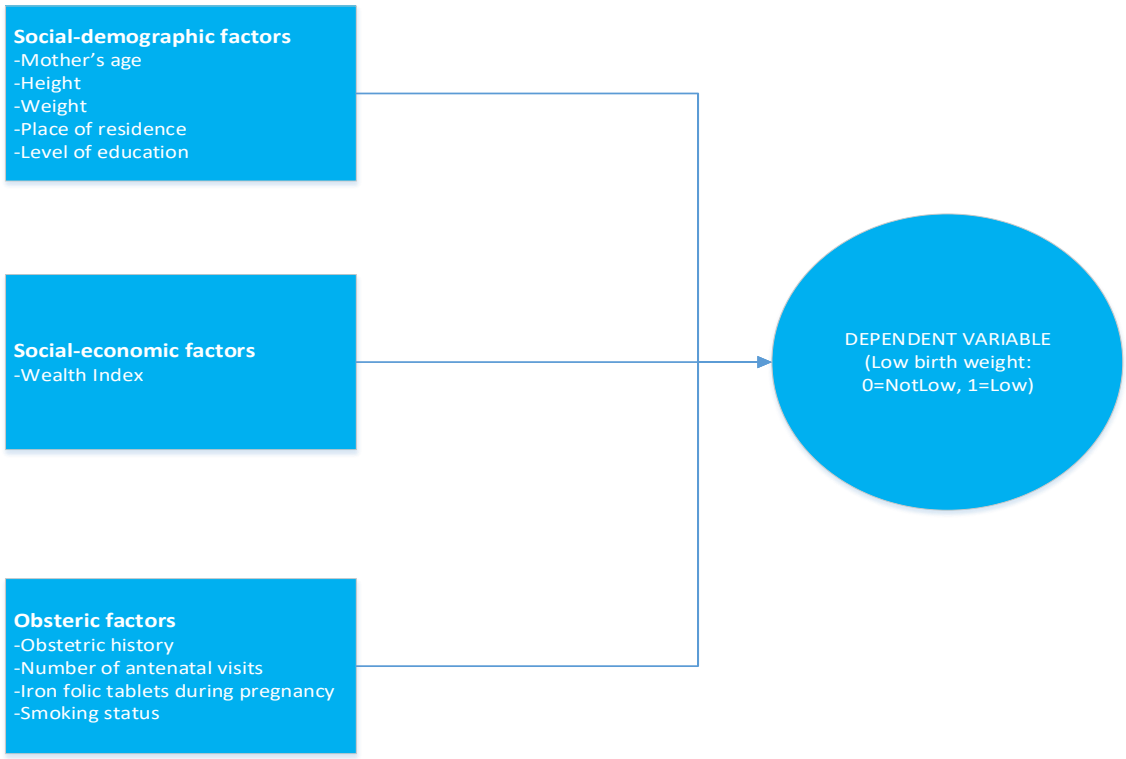

Figure 1.1: Conceptual framework

The figure 1.1 above displays the conceptual framework that was adapted in this research. It is essentially a diagram showing the independent variables and the dependent variable. The independent variables which are maternal risk factors for low birth weight are divided into three categories; the socio-demographic factors, the socio-economic factors and the obstetric factors. The socio-demographic category contains five variables. These are variables which incorporate the social and demographic aspects of the subjects under study. The variables in this category are; mother's age, weight, height, place of residence and level of education. The socio-economic category indicates variables that measure the economic status of the individuals. In this category, one variable was used which is the wealth index of the mother. The last category of independent variables is the obstetric factors category which describes aspects of pre-partum, pregnancy and childbirth. Four variables were included in this category. These are; the obstetric history which describes scenarios such as whether the respondent ever had a pregnancy that terminated in a miscarriage, abortion or still birth, the other variable in this category is number of antenatal visits, whether the mother took iron folic tablets while

pregnant and the smoking status of the mother. On the other hand, the dependent variable is low birth weight labelled as LBW which is binary in nature with two categories coded as 1 for the category with low birth weight and 0 for the category not having low birth weight.

1.7 Limitations of the study

There emerged some limitations during the research including the following:

1.7.1 Limited access to data

The DHS data did not provide all the information and variables needed. This led to limited findings in the study.

1.7.2 Time Constrains

The time required to complete the study was somehow short given a number of previous studies both empirical and theoretical had to be undertaken before embarking on the analysis and drawing conclusion on the same field scope. Again, more time was required on analysis in order to obtain detailed findings.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction to the literature review

Three sections in relation with low birth weight were developed here: Techniques for predicting fetal weight, factors which are associated with low birth weight across the world and utilization of machine leaning in low birth weight prediction.

2.2 Techniques for predicting fetal weight

Birth weight in obstetrics has been used as a primary indicator of foetal health status and foetal growth [29]. It helps to check whether the foetal growth is abnormal thereby preventing the risk of neonatal mortality and still birth. Therefore, there is a crucial need to accurately predict the birth weight of a foetus before birth.

Several studies have been conducted to build models for predicting the weight of babies and also using the clinical methods such as ultrasound and fundal height measurements. Prediction models for fetal weight where ultrasound facilities were absent was applied in Indonesia to reduce low birth weight delivery risks at given ages of gestation. The models efficacy was evaluated using multi-prediction measures of accuracy. The models that were proposed showed more accuracy compared to the existing ones in fetal weight prediction between 35 to 41 weeks of gestation [30].

One case control study carried out in a hospital at Arak, Iran used a decision curve analysis (DCA) to estimate the probability of having a newborn child with LBW. Out of the 15 factors discovered to be associated with low birth weight, with DCA the model used for prediction had a 0.3110 net benefit [31]. This value was considered to be substantial.

A different study in China proposed a model for predicting fetal weight using a genetic algorithm which optimizes a neural network on back propagation. The accuracy if this method was 76.3 percent which turned out to be 14.6 percent better than the traditional methods [32].

In Africa, ultrasound and clinical fetal weight measurement methods have also been used to estimate fetal weight. A study in South Nigeria compared the accuracy of sonographic and clinical methods in predicting term fetal weights. It was discovered that the accuracy of Dare's formula in estimating fetal weight is comparable to estimates from the ultrasound method [33].

In Kenya, a cross sectional based study conducted at the Kenyatta National hospital compared the ultrasound method and the clinical weight method based on actual birth weight after birth. The clinical weight estimation method appeared to be better than the ultrasound method in all the categories of weight [34].

2.3 Factors associated with low birth weight across the world

Low birth weight has been associated with several factors. Several researchers worldwide have focused on researching on these factors.

Many studies have been conducted to identify the risk factors associated with low birth weight in the American countries. In Rio Grande do Sul state in Brazil, an ecological study carried out established that mothers who have had less than seven prenatal check-ups have 3.8 times the risk of LBW. Maternal age also proved to be a predisposing factor since the mothers whosa age was above 35 years and those whose age was below 20 years had a higher LBW outcome [35]. Another study conducted a research on disparities of race on the risk of low birth weight. It was discovered that socio-economic status was a LBW risk factor in black women. Race was also discovered to be a risk factor since the chance of women of black race to give birth to babies with low weight was 2.6 times more than white women. Moreover, mothers with an underweight had a higher risk of giving birth to LBW infants than obese mothers. Health status was also studied and discovered that mothers who were healthy were less likely to give birth to LBW babies than those with poorer health [36]. Furthermore, smoking habits during pregnancy and other modifiable lifestyles have been highlighted as behavioural risks which trigger a several complications leading to LBW [37].

A different study in Afghanistan by Das Gupta showed that children of female gender, lower education for mothers, wealth index categorized as poor and urban settings were important factors associated with LBW [38]. In Malaysia, one study done in a tertiary hospital suggests a number of interplaying factors which lead to getting LBW children. The factors that were most significant were antenatal care, age of the mother, level of education and economic status [39].

Africa too has not been left behind in this research. In a study conducted in Tshwane, South Africa, to analyze the factors associated with giving birth to a baby with low weight suggested that LBW was associated with inadequate prenatal care, infant sex, older maternal age, premature rapture of membranes, maternal HIV, preterm birth, preeclampsia, and syphilis infections [40].

In hospitals of North Wello zone, Ethiopia, a research on low birth weight risk factors showed that Maternal weight during pregnancy, previous obstetric complication, place of antenatal follow-up and paternal education were associated with low birth weight [41].

A secondary data analysis study on risk factors for low birth weight in Zimbabwean Women reported that antepartum haemorrhage, prenatal care, preterm labor, infant sex, premature rapture of membranes and anaemia were associated with LBW. In addition, pregnancy induced hypertension, history of abortion or still-birth, eclampsia, history of LBW and malaria were associated with LBW [42].

Another study done at the teaching hospital of Butare in Rwanda on risk factors of preterm delivery of low birth weight in an African population reported that the major contributing factors to low birth weight in Rwanda were maternal weight, maternal height, preterm delivery and women with a poor nutritional status [43].

In Kenya, a certain study further illustrated that, the main factors associated with low birth weight include frequency of antenatal visits, maternal nutritional status, type of birth, region of residence, birth order and ethnicity. Other factors include maternal height and sex of the child [20]. In Pumwani Maternity hospital in Nairobi, a study conducted revealed that low birth weight was associated with number of meals taken per day while pregnant, vaginal bleeding, maternal anaemia, hypertension, pelvic pressure, abdominal pain and lower back ache [21]. At Coast General Hospital Mombasa County, it was found out that low birth weight was significantly associated with caesarean section birth, a previous low birth weight delivery, twin birth and less than four antenatal care visits. Moreover, women with college education level and a normal concentration of haemoglobin had a lower risk of giving birth to low weight babies [22].

2.4 Utilization of machine learning techniques in prediction of low birth weight

Machine learning techniques have proved to be a valuable tool in all spheres of research including health related research. Several studies have appreciated the utilization of machine learning techniques in predicting low birth weight.

The study done which used 2006 data from North Carolina State Centre for Health Statistics utilized machine learning techniques to build a data mining model to predict low birth weight with a high Area Under the Curve (AUC). Several procedures were followed to extract meaningful patterns from the data. These include; selecting the data, handling missing values,

dealing with imbalanced data, building the model, feature selection and evaluating the model. The machine learning models that were used for classification were; REPTree model, J48 and Random tree. The experimentation was done in two phases whereby the first one was done using an imbalanced dataset whereas the second one was performed using a balanced dataset using Synthetic Minority Oversampling Technique (SMOTE) that generates synthetic samples. Different results were obtained for the two experiments. The evaluation metrics that were employed are; AUC, sensitivity and specificity. After handling data imbalance, The J48 algorithm outperformed the other two models with a sensitivity of 66.3 percent, a specificity of 95.4 percent and AUC of 90.3 percent [44].

Ensemble machine learning techniques were utilized in a study conducted in China to estimate foetal weight at varying gestational age. Using data from electronic health records of pregnant women from a big hospital there in China, the ensemble model that was employed comprised of three models; that is random forest, XGBoost and Light GBM. A multi-parameter parallel optimization was used to perform foetal weight prediction in comparison with a multi-parameter formula applied in examination of ultrasound. The performance metrics that were used are accuracy and mean relative error (MRE). The MRE is used to measure the credibility of the prediction. Another evaluation metric applied was the Intersection over Union (IoU) to prove algorithm effectiveness. This index was formerly used for image processing. The experimental results were encouraging whereby an IoU index of 0.64 was achieved. Compared with the ultrasonic examination method, a 12 percent improvement in accuracy and a 3 percent reduction in MRE was achieved [7].

In India, data mining techniques were applied in predicting infants at risk of low birth weight and its factors. Compared to other classification methods, classification tree produced the best results; AUC of 93.80 percent, prediction accuracy of 89.95 percent, specificity of 72.88 percent, F-value of 93.04 percent and precision of 88.81 percent [45].

Indonesia has not been left behind too. Two studies in Indonesia have utilized the Indonesia Demographic and Health Survey data to predict and classify low birth weight using machine learning techniques. The first study compared binary logistic regression and random forest in prediction and classification. Random forest proved to be the best model in both tasks [46]. Using the same data with same variables, another study used Support Vector Machines (SVMs) for classification of LBW. The results revealed that SVMs with four kernel functions (hyperbolic

tangent, polynomial, linear and radial) were fit for binary classification of LBW. Moreover, their average predictive performance was satisfactory since the predictive error was below 10 percent [47].

2.5 Gap in the past studies

This study builds on the previous studies. Prediction of low birth weight based on the risk factors that have been identified can be of very big importance in identification of pregnant mothers who are at risk of giving birth to low birth weight infants. In Kenya particularly, several studies have been done concerning identification of risk factors for low birth weight. However, to my knowledge, no study has been done concerning predicting low birth weight using machine learning techniques based on Kenyan data. Therefore, this study is aimed at filling this gap.

CHAPTER THREE: RESEARCH METHODOLOGY

3.1 Introduction

This section describes the fundamental procedures and techniques that were applied to meet the objectives of the research. This consists of the data that was used, pre-processing techniques, handling data imbalance, machine learning models training, hyperparameter tuning and evaluation of performance metrics, variable importance and the software tools for the whole process. The different steps are presented in the diagram below:

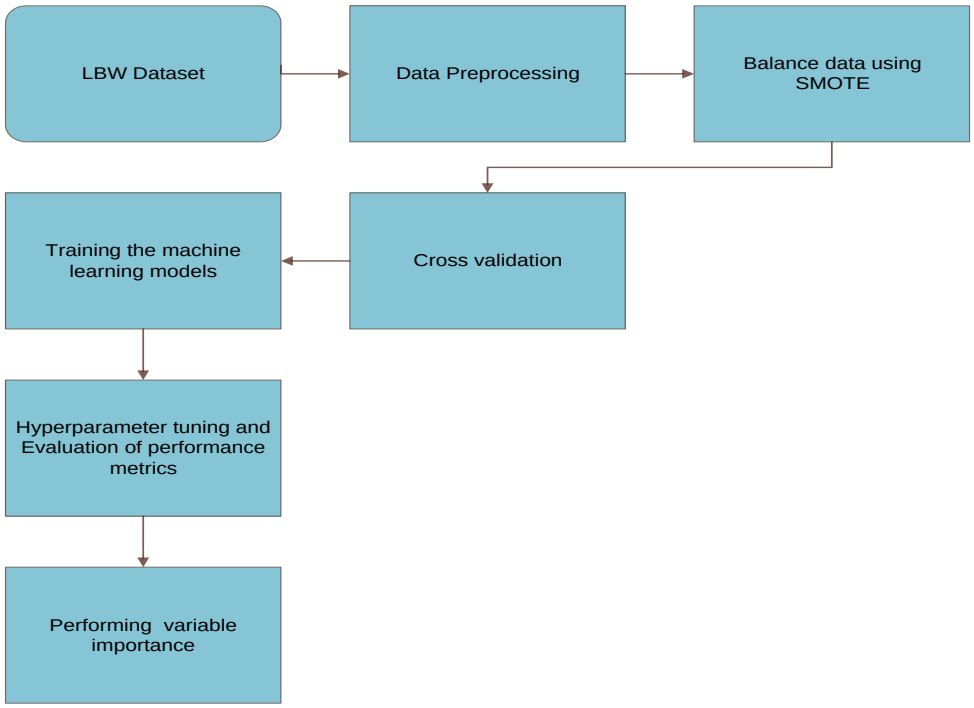

Figure 3.1: Illustration of procedure followed in predicting low birth weight.

Figure 3.1 above shows the procedural steps followed by the research towards constructing the model. The steps began with the dataset which contains the variables that were used in the research for prediction. The data was then subjected to a pre-processing process where exploratory analysis and cleaning of the data to prepare it for analysis is performed. After pre-processing, the data was balanced using the Synthetic Minority Oversampling Technique (SMOTE) in order to avoid the models from skewing results towards the majority class. Before the actual modelling, the data was subjected to cross validation in order to estimate how the models would perform before applying hyper-parameter tuning. Model building was then performed using the classification machine learning algorithms because the problem under study entails a target variable which is categorical in nature. After building the models, they were subjected to performance evaluation using evaluation metrics including precision, recall, F1 score, accuracy and ROC-AUC. A comparative analysis of the models was then done based on the evaluation metrics to get the most robust model. Variable importance was then performed to identify the independent variables that contributed most to the performance of the most robust model. These steps are discussed in depth below:

3.2 Data and variables

This study utilized secondary data from the 2014 Kenya Demographic and Health Survey (KDHS) which is the latest complete KDHS. The data focused mainly on the maternal risk factors that have been found out in previous researches to contribute to the risk of low birth weight in order to perform prediction of low birth weight. The study population included interviewed mothers between the age of 15 to 49 years.

The study variables which were extracted from the 2014 KDHS report data include both the dependent and independent variables. The dependent variable is low birth weight which is a binary variable consisting of two categories i.e. low birth weight and those without low birth weight. The independent variables, which are the maternal risk factors for low birth weight, are place of residence, mother's education level, age of the mother, mother's weight, mother's height, smoking habits, mother's obstetric history, wealth index, number of antenatal visits during pregnancy and whether the mother took iron folic tablets while pregnant. Below is their list and types:

Table 3.1: Variables used in the study and their categories

Variable	Type	Symbol	No of categories
Weight	Continous	MomWeight	-
Height	Continous	MomHeight	-
Age	Continous	MomAge	-
Number of antenatal visits	Continous	ANCVisits	-
Wealth index	Categorical	WealthIndex	5
Level of education	Categorical	HighestEduc	4
Taken iron folic tablets	Categorical	IronFolic	2
Place of residence	Categorical	Residence	2
Obstetric history	Categorical	ObstHist	2
Smoking status	Categorical	MomSmoke	2

3.3 Exploratory data analysis (EDA)

Data exploration is a very crucial step before any model building procedure can be undertaken. Exploratory data analysis consists of several tasks such as getting the summary statistics of the variables, visualizing the data to identify skewness and outliers in the dataset. Checking missing values is also another important aspect of exploration.

The first step was to get the summary statistics particularly descriptive statistics of all the variables. Descriptive statistics will be very important because they will allow presentation of data in a simpler way that is easier to interpret. After getting insight from the summary statistics, visualizations were considered more interactive. Data visualization is the presentation of data in pictorial or graphical form in a way that can be easily understood by the human brain more than just numbers. Several visualization techniques were employed to accomplish tasks such as checking the distribution of the variables, skewness and outliers in the data. Histograms are a popular method of data visualization to check the distribution and skewness of the variables. Boxplots and scatterplots were used to check for any outliers in the

dataset. This is because if outliers are not properly handled, they will skew the results of analysis. Bar graphs were employed to check the distribution of categorical variables and more specifically the outcome variable low birth weight which is binary in nature.

Missing values in the dataset were checked and replaced using the KNN (K-Nearest Neighbors) method.

Checking the correlation between the variables is also an important aspect that should not be ignored. A correlation matrix in form of a heat map sufficed for this task. This was done in order to identify the variables that have a high impact on the dependent variables. Correlation between independent variables was also keenly done to ascertain whether variable selection is important.

3.4 Data pre-processing

After data visualization has been fully explored, the next aspect was to pre-process the data since the faults were already identified at the EDA step. This ensured that the data to be used for building the model is of high quality. Several aspects of data pre-processing were taken into consideration. They include; completeness, accuracy, uniqueness, timeliness and consistency. Data completeness was accomplished by making sure that missing values have been handled. All the missing values were filled using the KNN machine learning method.

Data accuracy is also important to ascertain reliability of the information drawn from it. To ensure accuracy is taken care of, outliers in the data set were handled. The appropriate method to handle them was considered. These outliers were removed.

The next aspect that was accounted for during data cleaning was the uniqueness of the data. This was accomplished by removing duplicates in the data since duplicates contain the same information which bring about data redundancy.

Timeliness of the data is paramount too. Since this research utilized the latest KDHS in Kenya, this means that the data is timely and up to date. In addition, the integrity of the data is also high since it is based on DHS survey which is globally used by researchers and conducted by several governments supported by the USAID. Therefore, this rendered the research to be very trustworthy.

3.5 Handling data imbalance

In real life settings, most of the datasets are usually imbalanced whereby one of the classes has a higher percentage than the other. This results to skewness of the results. As evidenced in fig 3.2 below, data was highly imbalanced with the majority class (those without LBW) being 92.6 percent whereas the minority class (those with LBW) being 7.4 percent. This means that if the data was used this way without handling the imbalance issue, modelling would produce skewed results.

Synthetic Minority Oversampling Technique (SMOTE) was employed to balance the data. This resampling technique involves creating synthetic samples. It has a close similarity to oversampling. This technique is very handy in cases where data is limited. SMOTE technique applies the KNN algorithm in such a way that it randomly selects a minority class instance at random and identifies its k nearest neighbours. The synthetic instance is found by selecting one of the k nearest neighbours randomly and connecting it to the chosen minority class instance forming a line segment in the feature space [48].

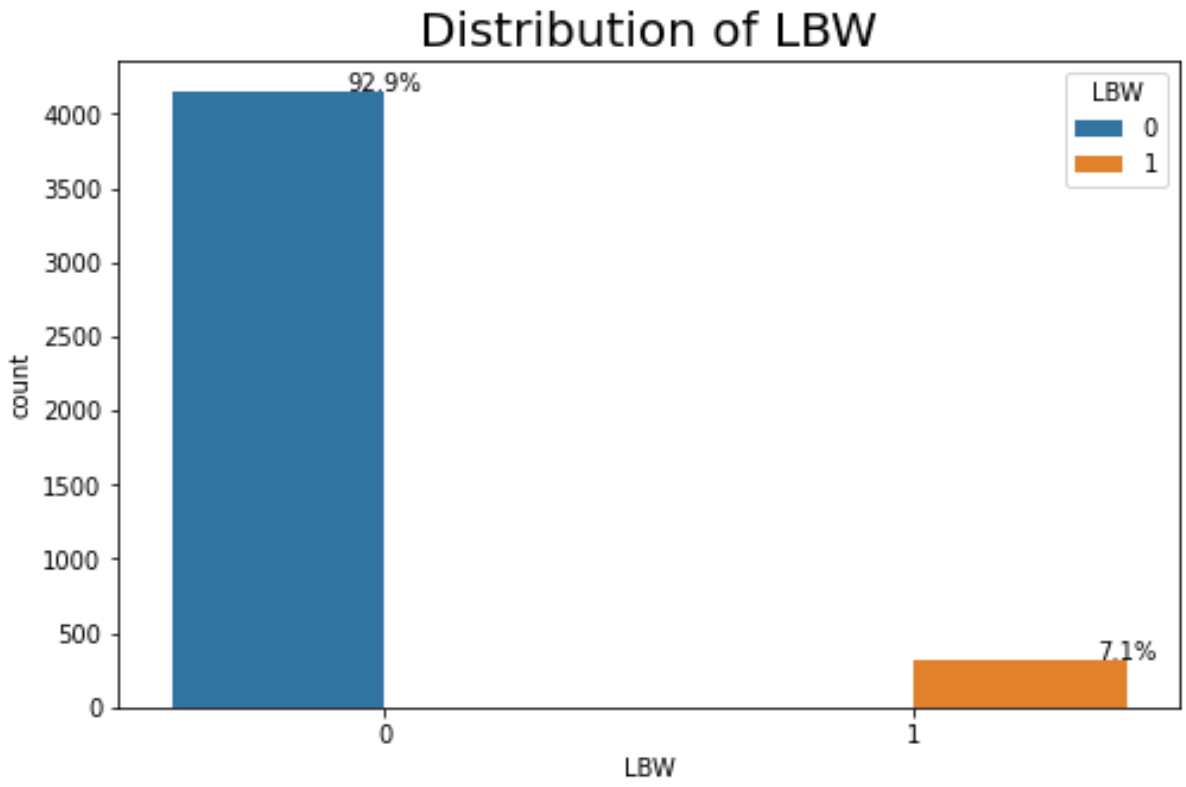

Figure 3.2: Bar graph of low birth weight showing imbalanced data

3.6 Cross validation

Cross-validation is a resampling method which aids in estimating the true prediction of the models [49]. It gives an insight of the performance of the machine learning models therefore providing guidance on the model tuning procedures including the models that are worth to be tuned in order to give better performance. K-fold cross validation was applied in this research specifically 5-fold whereby the data was split into five equal subsets. Each time, four subsets were used as a training set and the remaining set was used as a test set. The procedure was repeated until every subset was used as a training set and a test set at a particular time.

3.7 Hyper-parameter tuning

Hyper-parameter tuning involves finding the model parameters, which are used in the learning process of the model building to give optimum results. Grid search CV and randomized searchCV were used to tune the models. However, due to the computationally expensive nature of the grid searchCV, randomized searchCV was preferred. The goal of this procedure was to optimize the models to give the best results.

3.8 Model building

Model building accounts for the biggest focus in this research. This involved building a prediction model using various machine learning techniques in order to find out the one that performs the best. Since this is a classification problem whereby the output variable is binary in nature, specific machine learning models were employed for this task. Before fitting the machine learning models, the data was split into two, that is the training set and the test set. It was split in the ratio of 80:20 so that 80 percent of the data was used for training the models whereas the remaining 20 percent was used to test the models. The models that were used were; logistic regression, decision trees, random forest, Support Vector Machine (SVM), Gradient boosting and Extreme gradient boosting method popularly known as XGBoost. These six models were implemented in this research based on the history of their performance in predicting low birth weight. Logistic regression is popurlarly known for predictive analysis of classification problems. It was implemented by [45] and [46] in predicting low birth weight and produced good performance. The next algorithm implemented for this research was the decision tree. This model has previously also been used for predicting low birth weight by [45] and [44] among other researchers. Moreover , random forest which is an ensemble method combining several decision trees was applied. Random forest has previously been used by

researchers in predicting low birth weight including research by [46] and [45]. Support Vector Machines was also worth to be tested in this research. An indepth research on using SVM for predicting low birth weight based on different kernels was performed by [47]. Gradient boosting methods is another class of machine learning algorithms which have been used and produced exemplary performance as applied by [7].

Below are the models discussed:

3.8.1 Logistic Regression

Logistic regression is the simplest type of classification algorithm in machine learning. This research utilized the binary logistic regression which has two outcome categories. Its main goal is to find the relationship between the dependent variable Low Birth Weight and the independent variables X. [50] defines binary logistic regression in the following form:

$$\pi(x) = \frac{\exp(\beta 0 + \beta 1 + \cdots \beta p)}{1 + (\beta 0 + \beta 1 + \cdots \beta p)}$$

Where $\pi(x)$ is the outcome probability, $\beta 0 \ldots \beta p$ are unknown parameters and x's are the independent variables.

3.8.2 Decision Trees

A decision tree is a flow-chart like machine learning algorithm which comprises of leaves, branches and roots. Every internal node represents a test on an attribute, each branch denotes a test outcome and every leaf node represents a class label. For any tuple, $\boldsymbol{X}$, with an unknown class label, the tuple attribute values are tested against the decision tree. A path is tracked down from the root node up to the leaf node, which carrys the tuple's class prediction [51].

3.8.3 Random Forest

Random forest is defined as a group of combined trees where every tree depends on the values of a random vector independently sampled with equal distribution for all trees in the forest [52]. The algorithm of random forest follows the following steps [53]:

1. From the original data, create n trees bootstrap samples.
2. Grow a classification tree which is unpruned.
3. Predict new data.

3.8.4 SVM (Support vector machine)

A support vector machine is an algorithm which transforms training data into a higher dimension. It then finds a data separater called a hyperplane which separates that data by class by use of vital training tuples known as support vectors [51].

3.8.5 Gradient boosting

Gradient boosting is a class of machine learning techniques that is highly used. It is a boosting method which is gradient based. This algorithm follows a principle of constructing up to date base-learners that are correlated at maximum with the loss function's negative gradient which is associated with the ensemble as a whole [54].

3.8.6 Xtreme Gradient Boosting (XGBoost)

It is a gradient boosting technique whereby every training tuple is assigned weights. A sequence of k classifiers is learned repetitively. After a classifier termed Mi, has been learned, the updation of the weights is done to enable the next classifier, Mi+1, to "pay more attention" to the training tuples which Mi misclassified. Then M∗ the last classifier that is boosted, combines every individual classifier's votes, whereby each classifier's vote has a weight that is a function of its accuracy [51].

3.9 Model evaluation

After building the model, several evaluation metrics were used to evaluate the performance of the model. For this classification problem, the metrics used are; Accuracy, ROC-AUC, precision, F1 score and recall. The calculations of these metrics emanate from the confusion matrix.

3.9.1 Confusion Matrix

The confusion matrix is a contingency table which compares actual class to the model predictions. It is divided into true positive, false positive, true negative and false negative values:

True positive (TP): this is a case where actual positive values are predicted as positive. For instance, the number of cases correctly classified as low birth weight.

False positive (FP): this is a case where actual negative values are predicted as positive. For instance, the number of cases falsely classified as low birth weight.

True negative (TN): this is a case where actual negative values are predicted as negative. For instance, the number of cases correctly classified as not low birth weight.

False negative (FN): this is a case where actual positive values are predicted as negative. For instance, the number of cases falsely classified as not low birth weight.

3.9.2 Accuracy

Accuracy is described as a ratio between number of correctly classified points to the total number of points.

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

3.9.3 Precision

Precision is the fraction of correctly classified instances from the total classified instances.

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

3.9.4 Recall

Recall is the fraction of correctly classified instances from the total classified instances.

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

3.9.5 F1 Score

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

3.9.6 ROC-AUC

ROC (Receiver Operating Characteristic curve) is constructed by plotting true positives against false positives using various thresholds. To evaluate the performance of the model, the Area under curve (AUC) was used. The higher the AUC value, the better the classification ability of the model, meaning the model is able to clearly distinguish between the low birth weight class and the class without low birth weight.

3.10 Feature importance

Feature importance involves the techniques that are used to rank the predictor variables based on how important they are at predicting the dependent variable. There are several machine learning methods that can be employed to check feature importance. In this research, Random Forest technique was employed to check for the important input features for predicting Low birth weight since the algorithm Random Forest proved to be the best algorithm in predicting low birth weight. In addition, XGBoost was used to do the variable importance inorder to compare it with the random forest and check whether there is consistency in the results. This section helped in identifying the variables that highly contributed to the performance of the robustness of the model.

3.11 Software tools

The software that was used for this research during data analysis is Python Jupiter notebook which has several powerful libraries useful for all the analysis tasks. However, some of the exploratory data analysis tasks were performed using R.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction

This chapter displays the results and discussions of the research. It covers the preliminary exploration of the data displayed by the descriptive statistics table and a heatmap which shows the correlation of the continuous independent variables and the dependent variable. The results of the model performance including cross-validation and the main modelling are displayed and discussed. Finally, variable importance is illustrated and discussed.

Table 4.1. Descriptive statistics

	LBW No (N=4153)	**Yes (N=316)**	**Overall (N=4469)**
Obstetric history			
Good	3736 (93.03%)	280 (6.97%)	4016
Bad	417 (92.05%)	36.0 (7.95%)	453
Age			
Mean (SD)	28.5 (6.46)	28.8 (7.25)	28.5 (6.52)
Median [Min, Max]	28.0 [15.0, 49.0]	28.0 [15.0, 48.0]	28.0 [15.0, 49.0]
Weight			
Mean (SD)	61.5 (13.4)	60.4 (15.6)	61.4 (13.6)
Median [Min, Max]	59.5 [0, 168]	56.5 [36.9, 166]	59.3 [0, 168]
Height			
Mean (SD)	159 (13.0)	159 (6.39)	159 (12.7)
Median [Min, Max]	160 [0, 197]	158 [138, 187]	160 [0, 197]
Iron Folic			

	LBW No (N=4153)	Yes (N=316)	Overall (N=4469)
No	1043 (92.22%)	88.0 (7.78%)	1131
Yes	3110 (93.17%)	228 (6.83%)	3338
Smoking Status			
Non-smoker	4139 (92.93%)	315 (7.07%)	4454
Smoker	14.0 (93.33%)	1.00 (6.67%)	15.0
Residence			
Rural	2276 (92.80%)	167 (7.20%)	2443
Urban	1877 (92.65%)	149 (7.35%)	2026
Education level			
No education	329 (90.38%)	35.0 (9.62%)	364
Primary	2169 (92.14%)	185 (7.86%)	2354
Secondary	1201 (94.57%)	69.0 (5.43%)	1270
Higher	454 (94.39%)	27.0 (5.61%)	481
Wealth index			
Poorest	640 (90.65%)	66.0 (9.35%)	706
Poorer	833 (94.44%)	49.0 (5.56%)	882
Middle	801 (92.92%)	61.0 (7.08%)	862
Richer	933 (92.28%)	78.0 (7.71%)	1011
Richest	946 (93.85%)	62.0 (6.15%)	1008
ANCvisits			
Mean (SD)	4.23 (1.62)	4.28 (2.12)	4.23 (1.66)

	LBW No (N=4153)	**Yes (N=316)**	**Overall (N=4469)**
Median [Min, Max]	4.00 [1.00, 15.0]	4.00 [1.00, 20.0]	4.00 [1.00, 20.0]

Table 4.1 describes the descriptive statistics of the ten independent variables used for predicting low birth weight. A total of 4469 respondents were finally included in the modelling procedure after dropping the missing values and outliers. 4153 of the respondents translating to 92.93% gave birth to babies without low birth weight whereas only 316 translating to 7.07% gave birth to low birth weight babies. This proportion signifies an imbalanced class data and therefore imbalance data handling methods were considered before building the machine learning models. The mean age of mothers who gave birth to low birth weight babies was 28.8 years whereas those who gave birth to babies without low birth weight was 28.5 years. The mean weight for those who gave birth to low birth weight babies was 60.4 kg whereas those who gave birth to babies without low birth weight was 61.5 years. In terms of height, the mean height was 159 for both mothers with low birth weight and those without low birth weight. Antenatal care visits was another continuous variable that was considered. Approximately, a mean number of 4.23 visits was found for mothers who did not give birth to low birth weight babies whereas those who gave birth to low birth weight babies were 4.28. Categorical variables were also important to be looked at. The percentage of low birth weight was highest among mothers with a bad obstetric history with 7.95 % and lowest for mothers with a good obstetric history which was 6.97%. The percentage of low birth weight among mothers who did not take iron folic tablets is 7.78% whereas for those who took iron folic tablets while pregnant, the percentage of low birth weight is 6.83%. It can be observed that out of the 19 respondents who smoke, only 1 mother gave birth to a low birth weight baby translating to 6.67% whereas the percentage of low birth weight among mothers who do not smoke is 7.07%. The percentage of low birth weight was highest for mothers from the urban residence with 7.35% whereas it was least for mothers from the rural residence with 7.2%. The percentage of low birth weight was highest among mothers with the poorest wealth index with 9.35%. It was followed by those from the richer category with 7.71%, followed by those from the middle

category with 7.08% then those from the richest category with 6.15%. The percentage of low birth weight was least among mothers from the poorer category with 5.56%. The percentage of low birth weight was highest among mothers with no education with 9.62 %. It was followed by those with primary level of education with 7.86%, then those with higher education with 5.61% whereas the percentage of low birth weight was least among mothers with secondary level of education with 5.43%.

4.2. Correlation between the dependent variable and independent variables

Below is the summary of correlation between continuous independent variables and the dependent variable.

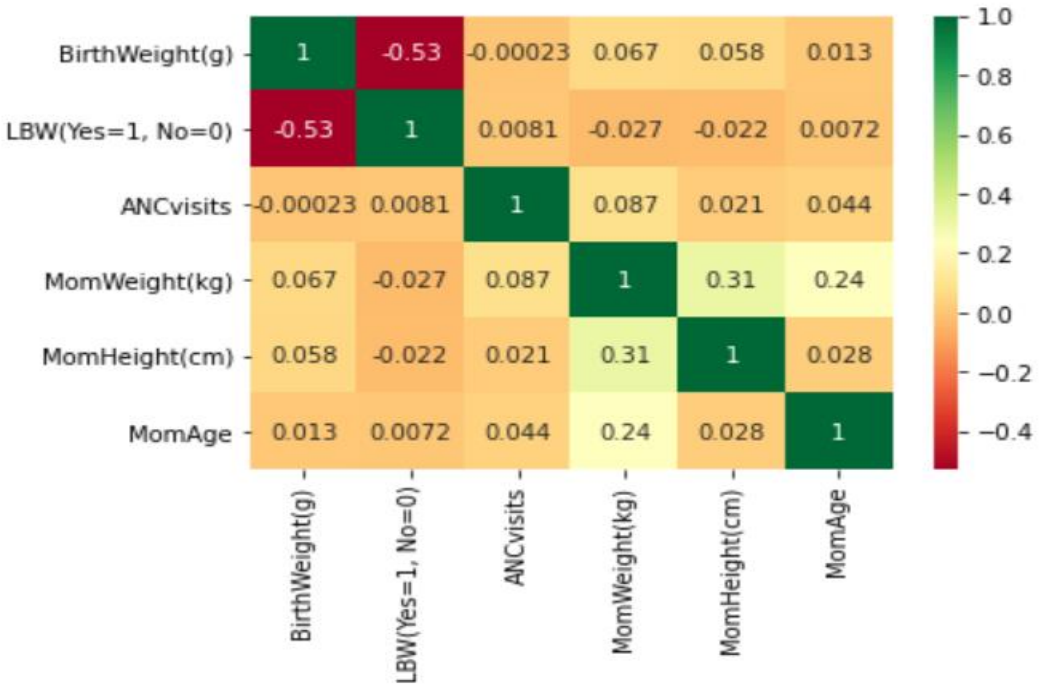

Figure 4.1: Correlation of dependent variable and continuous independent variables

Figure 4.1 above shows a heat map diagram of correlation matrix which displays the correlation between the dependent variable and the independent continuous variables. The variable BirthWeight has a positive correlation with all the independent variables except number of antenatal care visits. This means that as mother's weight, age and height increases, the birth weight of the baby also increases. Moreover, there exists no multicollinearity among the independent variables.

4.3 Model results

4.3.1: Predictive Performance

After the data was balanced and cross validation was performed, the next aspect was to build a machine learning model to predict low birth weight. Several classification algorithms were tested to predict Low birth weight. They are: logistic regression, decision tree classifier, random forest, SVM, gradient boosting and extreme gradient boosting. Hyperparameter tuning was performed on the models to optimize the results. However for some algorithms, the default parameters gave the best performance metric scores. After hyper-parameter tuning, the following results were found.

Table 4.2 Performance metrics of the machine learning algorithms

Model	Precision score	Recall score	F1 score	Accuracy
Logistic Regression	0.673333	0.666667	0.664621	0.666667
Random Forest	0.956831	0.956679	0.956666	0.956679
Decision Tree	0.883662	0.883273	0.883283	0.883273
Gradient Boosting	0.88827	0.871841	0.870153	0.871841
XGBoost	0.941236	0.939832	0.939746	0.939832
SVM	0.770404	0.761733	0.760404	0.761733

To evaluate the predictive ability of the classification algorithms, accuracy metric was computed. Out of the six models that were built to predict low birth weight, it is evident that the random forest model emerged to be the best with an accuracy of 0.956679. It was followed by XGBoost which had an accuracy of 0.939832. The third best was the decision tree with an accuracy of 0.883273, followed by the gradient boosting with an accuracy of 0.871841, then the Support Vector Machine (SVM) with 0.761733 and finally logistic regression with 0.666667 accuracy.

After checking the predictive ability of the models using accuracy metric, the effectiveness of the models was checked using F1 score, precision score and recall score. Starting with the precision score, random forest turned up to be the best by scoring 0.956831 followed by XGBoost with 0.941236. Gradient boosting succeeded with precision of 0.888270, then decision tree with 0.883662, followed by SVM with 0.770404 and finally logistic regression

with 0.673333. Moreover, the models were evaluated based on the recall score. Using this measure, random forest again emerged to be the best with a recall score of 0.956679. It was followed by XGBoost with 0.939832 which was succeeded by the decision tree with 0.883273. The gradient boosting technique then followed with 0.871841. SVM came after with a recall score of 0.761733 and finally the logistic regression with 0.666667. In addition, F1 score was utilized as another performance evaluation metric to compare the models. Random forest emerged the best with an F1 score of 0.956666. XGBoost succeeded it with 0.939746 followed by the decision tree with 0.883283, then the gradient boosting technique with 0.870153. SVM scored 0.760404 and finally the logistic regression managed to score 0.664621.

4.3.2: Discrimination analysis

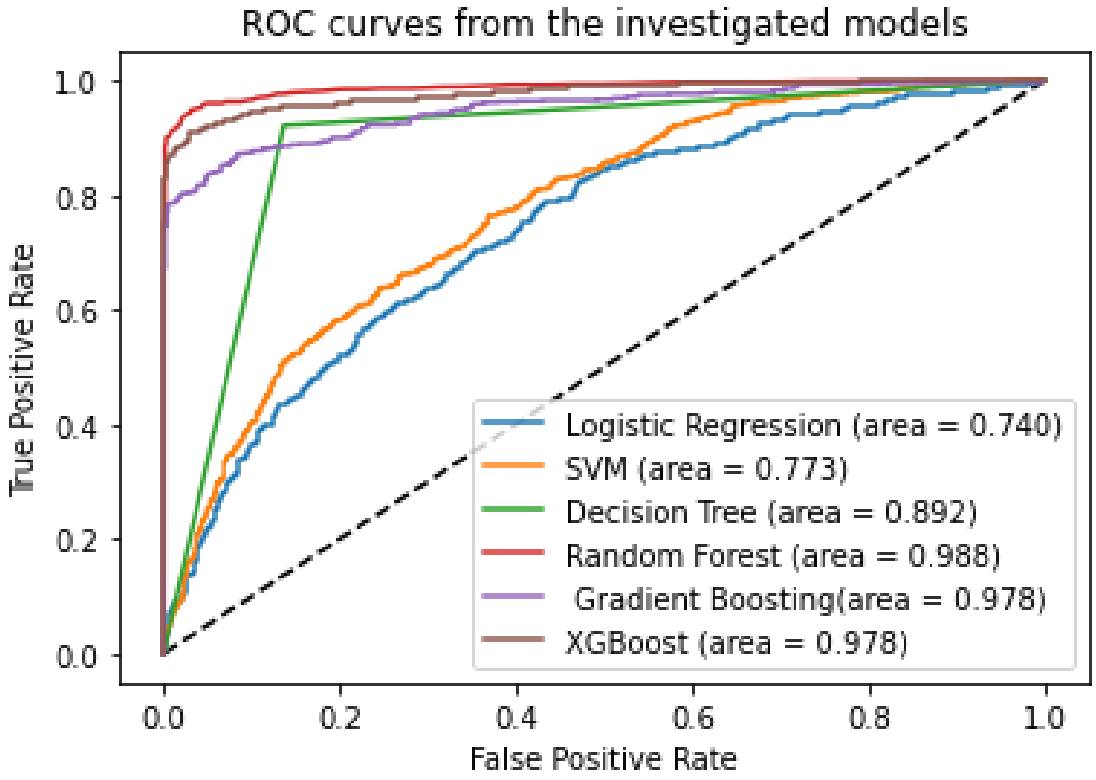

Figure 4.2: Receiver Operating Characteristic Curve (ROC)

Figure 4.2, exhibits the ROC curves of the machine learning models that were experimented in this research to predict low birth weight. The ROC curve shows the ability of the machine learning models to discriminate birth weight as either low birth weight or not low birth weight. It plots the trade-off between the true positive rate and the false positive rate at various thresholds. Out of the six machine learning algorithms tested in this research, random forest produced the highest area under the curve (AUC) of 0.988. The second best AUC was taken by the gradient boosting technique with a value of 0.978 which tied with the XGBoost with the same value. The decision tree then followed with an AUC of 0.892 followed by SVM with 0.773 and lastly the logistic regression which scored an AUC of 0.740.

4.4 Variable contribution to the robustness of the models

The contribution of the variables to the robustness of the machine learning models that were investigated in this research was good enough to be looked into. According to the performance of the models based on the predictive performance evaluation metrics that is the accuracy, F1 score, precision and recall score as well as the ROC-AUC, random forest emerged to be the best followed by the XGBoost. Therefore, the order of the importance of the variables was evaluated based on these two machine learning models as shown in the figures below.

4.4.1 Variable importance based on random forest algorithm

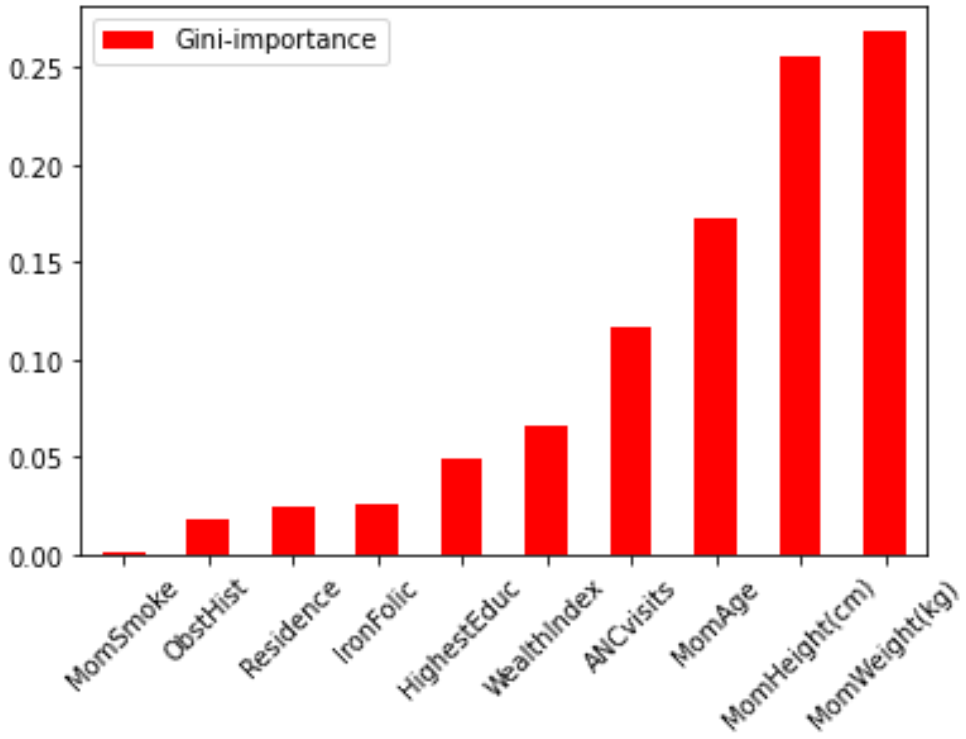

Figure 4.3: Plot of variable importance using the random forest algorithm

Figure 4.3 displays the order of importance of the independent variables based on the gini importance of the random forest algorithm that outperformed the other algorithms. Gini importance calculates the importance of every predictor as an addition of splits across all trees including predictor proportionality to the number of samples it splits. It was observed that the four most important variables for predicting low birth weight, which had a gini-importance value of 0.1 and above, were mother's weight, mother's height, mother's age and number of antenatal visits attended during pregnancy.

4.4.2 Variable importance based on XGBoost algorithm

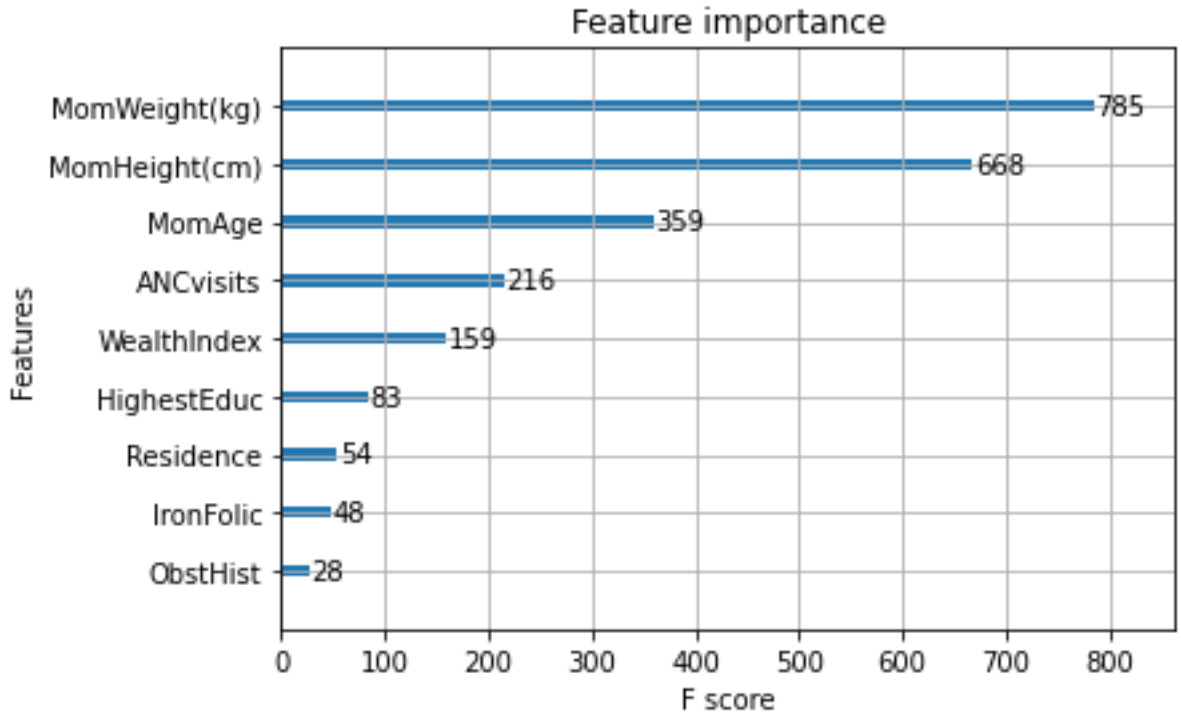

Figure 4.4: Plot of variable importance using the XGBoost algorithm

XGBoost emerged to be the second best model in this research. Variable importance was therefore checked based on it in order to ascertain whether the order is the same as the order according to the random forest model. Figure 4.4 shows the order of the importance of the variables. The order is similar to that of the random forest model. The importance was measure based on the F score. Those with an F score value of 200 and above were taken to be the most important variables. The most important was mother's weight followed by mother's height, then mother's age and lastly the number of antenatal visits during pregnancy.

4.5 Limitations

This research also exhibited some limitations. It was a challenge to get clinical data from an obstetric clinical records. This data could have given more information as compared to DHS data which is based on interviewing individuals which might be affected by recall bias and also a lot of missing values as was experienced in the DHS data used for this research.

CHAPTER FIVE: SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter describes the summary of the results, the conclusion and the recommendations suggested based on the results of the research.

5.2 Summary of the results

The main objective of this research was to build a machine learning model for predicting low birth weight. Three specific objectives were studied. To achieve these objectives, the 2014 KDHS data was used. The data was subjected to pre-processing in order to clean it and prepare for the modelling tasks. Six machine learning algorithms were trained and tested namely; the random forest, decision tree, gradient boosting, XGBoost, SVM and logistic regression. After training and testing them, they were evaluated based on performance metrics specifically used for classification problems. The metrics used were accuracy, precision score, recall score and F1 score. Furthermore, the ROC-AUC was used to test the classification ability of the models to differentiate between the low birth weight cases and the cases without low birth weight. In terms of accuracy, the best machine learning model was the random forest with an accuracy of 0.956679. The other five models produced an accuracy that ranged between 0.666667 to 0.939832. Based on the precision score, random forest again emerged to be the best with a value of 0.956831. The rest of the models had a precision score that ranged from 0.673333 to 0.941236. Recall score was also evaluated and random forest had the best recall score value of 0.956679 whereas the other models managed a recall score ranging from 0.666667 to 0.939832. Moreover, the F1 score was also examined. Random forest model had the best F1 score value of 0.956666. The rest of the models F1 score ranged from 0.664621 to 0.939746. Furthermore, the ROC curves of all the tested models were plotted and the area under the curved evaluated. The random forest had the highest area under the curve of 0.988. The AUC of the other models ranged from 0.740 to 0.978. Therefore, from these results based on the performance metrics and ROC-AUC, random forest emerged to be the most robust model.

In addition, variable importance was examined. This specific objective was geared to ascertain the variables which are the most important to be considered when predicting low birth weight. The random forest algorithm was used to perform the variable importance since it proved to be

the most robust model for predicting low birth weight. In addition, XGBoost which was the second best robust model was used to compare with the random forest model to comprehend whether the variable importance would produce the same results as that of the random forest. Fortunately, the two models produced the same results. The metrics used to measure the variable importance were gini importance for the random forest model and F measure for the XGBoost model. It was found out that out of the 10 independent variables used the most important ones for predicting low birth weight were; mother's weight, mother's height, mother's age and the number of antenatal care visits during pregnancy.

5.3 Conclusion

To conclude, machine learning techniques have lately been a very useful tool for performing prediction tasks in the field of science. Low birth weight is still a major challenge that should be looked into in order to reduce child mortality. Therefore, machine learning models were employed to perform prediction. It was found out that random forest is the best model for predicting low birth weight since it had the highest accuracy and effectiveness in terms of recall and precision. Moreover, random forest yielded the best AUC therefore the best classification model. It was also important to identify the variables that contributed most to the robustness of the model. This technique known as feature importance was performed using the random forest technique. It was ascertained that mother's weight, height, age and number of antenatal visits attended during pregnancy are the most important variables that contributed most to the models accuracy.

5.4 Recommendations

After completing this study, some recommendations and suggestions transpired for further research. First, the variables that were found to be the most important in predicting low birth weight, that is; mother's weight, height, age and number of antenatal visits should be the key factors looked at. The mother should be given appropriate advice to mitigate low birth weight based on these variables. Second, clinical data should be considered to be used for further studies in order to get more relevant variables to be included in the prediction since clinical data would give more insight. Moreover, apart from the six machine learning algorithms used in this research, more algorithms should be employed in order to get more improvements and discover more robust models for predicting low birth weight. In addition, other variables apart from maternal risk factors should be experimented.

REFERENCES

1. UNICEF and WHO. Low birthweight. *East Afr Med J*. 2004;63(2):89-90. doi:10.1787/9789264183902-17-en

2. WHO. Low Birth WeightPolicy Brief. *WHA Glob Nutr Targets 2025*. 2014;28(3):1-66.

3. UNICEF and WHO. Low birthweight estimates. *Lancet Glob Heal*. 2019.

4. WHO. *Born Too Soon:The Global Action Report on Preterm Birth*. Vol 164. Geneva; 2012.

5. Sharma D, Shastri S, Sharma P. Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. *Clin Med Insights Pediatr*. 2016;10:CMPed.S40070. doi:10.4137/cmped.s40070

6. EVERY PREEMIER SCALE. *PROFILE OF PRETERM AND LOW BIRTH WEIGHT PREVENTION AND CARE*.; 2017. www.EveryPreemie.org. Accessed January 8, 2021.

7. Lu Y, Zhang X, Fu X, Chen F, Wong KKL. Ensemble machine learning for estimating fetal weight at varying gestational age. *33rd AAAI Conf Artif Intell AAAI 2019, 31st Innov Appl Artif Intell Conf IAAI 2019 9th AAAI Symp Educ Adv Artif Intell EAAI 2019*. 2019:9522-9527. doi:10.1609/aaai.v33i01.33019522

8. Jan-Simon Lanowski A, Lanowski G, Schippert C, et al. Ultrasound versus Clinical Examination to Estimate Fetal Weight at Term Vergleich von Ultraschall und klinischer Untersuchung zur fetalen Gewichtsschätzung am Geburtstermin. *GebFra Sci*. 2017. doi:10.1055/s-0043-102406

9. Dönmez P. Introduction to Machine Learning, 2nd ed., by Ethem Alpaydın. Cambridge, MA: The MIT Press2010. ISBN: 978-0-262-01243-0. $54/£ 39.95 + 584 pages. *Nat Lang Eng*. 2013;19(2):285-288. doi:10.1017/s1351324912000290

10. Liu Q, Wu Y. Encyclopedia of the Sciences of Learning. *Encycl Sci Learn*. 2012;(April). doi:10.1007/978-1-4419-1428-6

11. Blencowe H, Krasevec J, de Onis M, et al. National, regional, and worldwide estimates

of low birthweight in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Heal*. 2019;7(7):e849-e860. doi:10.1016/S2214-109X(18)30565-5

12. Pal A, Manna S, Das B, Dhara PC. The risk of low birth weight and associated factors in West Bengal, India: a community based cross-sectional study. *Egypt Pediatr Assoc Gaz*. 2020;68(1). doi:10.1186/s43054-020-00040-0

13. Bankole Oladeinde H, Oladeinde OB, Omoregie R, Onifade AA. Prevalence and determinants of low birth weight: the situation in a traditional birth home in Benin City, Nigeria. *Afri Heal Sci*. 2015;15(4):1123-1132. doi:10.4314/ahs.v15i4.10

14. Agbor VN, Ditah C, Tochie JN, Njim T. Low birthweight in rural Cameroon: An analysis of a cut-off value. *BMC Pregnancy Childbirth*. 2018;18(1):1-5. doi:10.1186/s12884-018-1663-y

15. He Z, Bishwajit G, Yaya S, Cheng Z, Zou D, Zhou Y. Prevalence of low birth weight and its association with maternal body weight status in selected countries in Africa: A cross-sectional study. *BMJ Open*. 2018;8(8):1-8. doi:10.1136/bmjopen-2017-020410

16. Endalamaw A, Engeda EH, Ekubagewargies DT, Belay GM, Tefera MA. Low birth weight and its associated factors in Ethiopia: A systematic review and meta-analysis. *Ital J Pediatr*. 2018;44(1):1-12. doi:10.1186/s13052-018-0586-6

17. Mvunta MH, Mboya IB, Msuya SE, John B, Obure J, Mahande MJ. Incidence and recurrence risk of low birth weight in Northern Tanzania: A registry based study. *PLoS One*. 2019;14(4):1-10. doi:10.1371/journal.pone.0215768

18. Mazaki-Tovi S, Romero R, Kusanovic JP, et al. Recurrent Preterm Birth. *Semin Perinatol*. 2007;31(3):142-158. doi:10.1053/j.semperi.2007.04.001

19. National Bureau of Statistics K, Dhs M, Macro I. *Kenya Demographic and Health Survey*.; 2008.

20. Muchemi OM, Echoka E, Makokha A. Factors associated with low birth weight among neonates born at Olkalou district hospital, central region, Kenya. *Pan Afr Med J*. 2015;20(2):1-11. doi:10.11604/pamj.2015.20.108.4831

21. Mogire GK. Factors Associated With Low Birth Weight Deliveries in Pumwani

Maternity Hospital , Master of Science Jomo Kenyatta University of. *FactorsAssociated with Low Birth Weight Deliv Pumwani Matern Hosp Nairobi-Kenya*. 2013;3(2):1-99. http://ir.jkuat.ac.ke/bitstream/handle/123456789/1304/MOGIRE%2C GRACE -Msc Epidemiology- 2013.pdf?sequence=1&isAllowed=y.

22. Jumbale CM, Karanja S, Udu R. Factors Influencing Low Birth Weight (Lbw) Among Mother-Neonate Pairs and Associated Health Outcomes At Coast General Hospital Mombasa County Kenya. *Glob J Heal Sci*. 2018;3(3):41-53.

23. Larroque B, Bertrais S, Czernichow P, Léger J. School difficulties in 20-year-olds who were born small for gestational age at term in a regional cohort study. *Pediatrics*. 2001;108(1):111-115. doi:10.1542/peds.108.1.111

24. Christian P, Lee SE, Angel MD, et al. Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *Int J Epidemiol*. 2013;42(5):1340-1355. doi:10.1093/ije/dyt109

25. Jornayvaz FR, Vollenweider P, Bochud M, Mooser V, Waeber G, Marques-Vidal P. Low birth weight leads to obesity, diabetes and increased leptin levels in adults: The CoLaus study. *Cardiovasc Diabetol*. 2016;15(1):1-10. doi:10.1186/s12933-016-0389-2

26. Gu H, Wang L, Lingfei L, et al. OPEN A gradient relationship between low birth weight and IQ : A meta- analysis. *Sci Rep*. 2017;(December):1-14. doi:10.1038/s41598-017-18234-9

27. Risnes KR, Vatten LJ, Baker JL, Jameson K, Sovio U. Birthweight and mortality in adulthood : a systematic review and meta-analysis. *Int J Epidemiol*. 2011;(February). doi:10.1093/ije/dyq267

28. Zohdi V, Sutherland MR, Lim K, Gubhaju L, Zimanyi MA, Black MJ. Low birth weight due to intrauterine growth restriction and/or preterm birth: Effects on nephron number and long-term renal health. *Int J Nephrol*. 2012;2012. doi:10.1155/2012/136942

29. Camerota M, Bollen KA. Birth weight, birth length, and gestational age as indicators of favorable fetal growth conditions in a us sample. *PLoS One*. 2016;11(4):1-15.

doi:10.1371/journal.pone.0153800

30. Anggraini D, Abdollahian M, Marion K. Foetal weight prediction models at a given gestational age in the absence of ultrasound facilities: Application in Indonesia. *BMC Pregnancy Childbirth*. 2018;18(1):1-12. doi:10.1186/s12884-018-2047-z

31. Rejali M, Mansourian M, Babaei Z, Eshrati B. Prediction of Low Birth Weight Delivery by Maternal Status and Its Validation: Decision Curve Analysis. *Int J Prev Med*. 2017;8. doi:10.4103/IJPVM.IJPVM_146_16

32. Gao H, Wu C, Huang D, Zha D, Zhou C. Prediction of fetal weight based on back propagation neural network optimized by genetic algorithm. 2021;18(March):4402-4410. doi:10.3934/mbe.2021222

33. Njoku C, Emechebe C, Odusolu P, Abeshi S, Chukwu C, Ekabua J. Determination of Accuracy of Fetal Weight Using Ultrasound and Clinical Fetal Weight Estimations in Calabar South, South Nigeria. *Int Sch Res Not*. 2014;2014:1-6. doi:10.1155/2014/970973

34. Wanjaria DK. a Correlation of Ultrasound and Clinical Fetal Weight. *J Med Ultrasound*. 2016;24(2004):144-148.

35. da Silva TRSR. Nonbiological maternal risk factor for low birth weight on Latin America: a systematic review of literature with meta-analysis. *Einstein (Sao Paulo)*. 2012;10(3):380-385. doi:10.1590/S1679-45082012000300023

36. Clay SL, Andrade FCD. Racial disparities in low birthweight risk: An examination of stress predictors. *J Racial Ethn Heal Disparities*. 2015;3(2):200-209. doi:10.1007/s40615-015-0128-5

37. Veloso HJF, da Silva AAM, Bettiol H, et al. Low birth weight in São Luís, northeastern Brazil: Trends and associated factors. *BMC Pregnancy Childbirth*. 2014;14(1):1-12. doi:10.1186/1471-2393-14-155

38. Gupta R Das, Swasey K, Burrowes V, Hashan MR, Al Kibria GM. Factors associated with low birth weight in Afghanistan: A cross-sectional analysis of the demographic and health survey 2015. *BMJ Open*. 2019;9(5). doi:10.1136/bmjopen-2018-025715

39. Yadav H, Lee N. Maternal factors in predicting low birth weight babies. *Med J Malaysia*. 2013;68(1):44-47.

40. Tshotetsi L, Dzikiti L, Hajison P, Feresu S. Maternal factors contributing to low birth weight deliveries in Tshwane District, South Africa. *PLoS One*. 2019;14(3). doi:10.1371/journal.pone.0213058

41. Wachamo TM, Yimer NB, Bizuneh AD. Risk factors for low birth weight in hospitals of North Wello zone, Ethiopia: A case-control study. *PLoS One*. 2019;14(3):1-15. doi:10.1371/journal.pone.0213054

42. Feresu SA, Harlow SD, Woelk GB. Risk factors for low birthweight in Zimbabwean women: A secondary data analysis. *PLoS One*. 2015;10(6):1-17. doi:10.1371/journal.pone.0129705

43. Bayingana C, Muvunyi CM, Africa CWJ. Risk factors of preterm delivery of low birth weight (plbw) in an African population. *J Clin Med Res*. 2010;2(7):114-118. http://www.academicjournals.org/JCMR.

44. Hange U, Selvaraj R, Galani M, Letsholo K. A data-mining model for predicting low birth weight with a high AUC. *Stud Comput Intell*. 2018;719(June 2020):109-121. doi:10.1007/978-3-319-60170-0_8

45. Senthilkumar, D; Paulraj S. Prediction of Low Birth Weight Infants and Its Risk Factors Using Data Mining Techniques. *Proc 2015 Int Conf Ind Eng Oper Manag Dubai, United Arab Emirates*. 2015;3:186-194.

46. Faruk A, Cahyono ES, Eliyati N, Arifieni I. Prediction and classification of low birth weight data using machine learning techniques. *Indones J Sci Technol*. 2018;3(1):18-28. doi:10.17509/ijost.v3i1.10799

47. Eliyati N, Faruk A, Kresnawati ES, Arifieni I. Support vector machines for classification of low birth weight in Indonesia Support vector machines for classification of low birth weight in Indonesia. *IOP J Phys*. 2019. doi:10.1088/1742-6596/1282/1/012010

48. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority

over-sampling technique. *J Artif Intell Res*. 2002;16(February 2017):321-357. doi:10.1613/jair.953

49. Berrar D. Cross-validation. *Encycl Bioinforma Comput Biol ABC Bioinforma*. 2018;1-3(April):542-545. doi:10.1016/B978-0-12-809633-8.20349-X

50. Kleinbaum DG, Klein M. Ordinal Logistic Regression. *Springer*. 2010:463-488. doi:10.1007/978-1-4419-1742-3_13

51. Agarwal S. *Data Mining: Data Mining Concepts and Techniques*.; 2014. doi:10.1109/ICMIRA.2013.45

52. Pavlov YL. Random forests. *Random For*. 2019:1-122. doi:10.1201/9780429469275-8

53. Liaw A, Wiener M. Classification and Regression by randomForest. *R news*. 2002;2(December):18-22. http://cran.r-project.org/doc/Rnews/.

54. Natekin A, Knoll A. Gradient boosting machines , a tutorial. *Front Neurorobot*. 2013;7(December). doi:10.3389/fnbot.2013.00021

APPENDIX: Turnitin Report

9/13/21, 9:26 PM Turnitin

1647592067

Turnitin Originality Report

Word Count: 13327
Document Viewer

Processed on: 13-Sep-2021 20:48 EAT ID:

Similarity Index
Similarity by Source

Submitted: 1

DISSERTATION By shaz sawe

19% Internet Sources: 13%
Publications: 13%
Student Papers: 5%

include quoted include bibliography exclude small matches mode: quickview (classic) report Change mode print download

1% match (Internet from 14-Apr-2021)
http://erepository.uonbi.ac.ke

1% match (Internet from 17-Dec-2020)
https://towardsdatascience.com/performance-metrics-for-classification-machine-learning-problems-97e7e774a007? gi=a3db002f50d1

<1% match (Internet from 14-Apr-2021)
http://erepository.uonbi.ac.ke

<1% match (Internet from 03-May-2021)
http://erepository.uonbi.ac.ke

<1% match (Internet from 04-May-2021)
http://erepository.uonbi.ac.ke

<1% match (Internet from 04-May-2021)
http://erepository.uonbi.ac.ke

<1% match (Internet from 14-Apr-2021)
http://erepository.uonbi.ac.ke

<1% match (Internet from 09-Sep-2020)
https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0213054

<1% match (Internet from 22-May-2020)
https://journals.plos.org/plosone/article/file?id=10.1371%2Fjournal.pone.0221257&type=printable

<1% match (Internet from 20-Jun-2021)
https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0129705

<1% match (Internet from 18-Aug-2020)

https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0215768

<1% match (Internet from 20-Jun-2021)
https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0213058

<1% match (Internet from 03-Aug-2020)
https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0123962

<1% match (Internet from 08-Nov-2020)
https://journals.plos.org/plosone/article/file?id=10.1371%2Fjournal.pone.0235626&type=printable

<1% match (Internet from 01-Mar-2020)
https://worldwidescience.org/topicpages/b/birth+weight+rate.html

<1% match (Internet from 26-Sep-2020)
https://worldwidescience.org/topicpages/l/low-birth+weight+preterm.html

<1% match (Internet from 07-Mar-2020)
https://worldwidescience.org/topicpages/w/weight+lbw+preterm.html

<1% match (Internet from 03-Dec-2017)
https://link.springer.com/content/pdf/10.1007%2F978-3-319-60170-0_8.pdf

<1% match (Internet from 28-Jul-2020)
https://link.springer.com/chapter/10.1007%2F978-981-15-6318-8_7

<1% match (Internet from 26-Aug-2018)
https://link.springer.com/article/10.1007%2Fs00404-017-4457-y

<1% match (publications)
"Machine Learning, Image Processing, Network Security and Data Sciences", Springer Science and Business Media LLC, 2020

<1% match (publications)
Han, Jiawei, Micheline Kamber, and Jian Pei. "Classification", Data Mining, 2012.

<1% match (publications)
Han, Jiawei, Micheline Kamber, and Jian Pei. "Classification", Data Mining, 2012.

<1% match (Internet from 28-Jul-2021)
http://ugspace.ug.edu.gh

<1% match (Internet from 13-Aug-2021)
http://ugspace.ug.edu.gh

<1% match (publications)
Sofia Mawaddah, Sulis Tiyawati. "Passive Smokers Pregnant Women with Low Birth Weight", JURNAL INFO KESEHATAN, 2021

<1% match (Internet from 29-Oct-2020)
https://www.hindawi.com/journals/ijpedi/2019/4628301/

<1% match (Internet from 14-Dec-2020)
https://www.hindawi.com/journals/ijpedi/2020/8394578/

<1% match (Internet from 26-Jun-2020)
https://www.hindawi.com/journals/apm/2020/8459694/

<1% match (Internet from 07-Sep-2021)
https://www.hindawi.com/journals/isrn/2014/970973/

<1% match (Internet from 08-May-2019)
http://ir.jkuat.ac.ke

<1% match (Internet from 12-Apr-2021)
http://Erepository.uonbi.ac.ke

<1% match (Internet from 25-Dec-2019)
https://www.panafrican-med-journal.com/content/article/20/108/full/

<1% match (Internet from 10-Jun-2021)
https://ir
library.ku.ac.ke/bitstream/handle/123456789/20054/Determinants%20of%20The%20Rising%20Numbers%20of%20Traffic%20Road%e2%80%a6% isAllowed=y&sequence=1

<1% match (Internet from 07-Jun-2021)
https://ir-
library.ku.ac.ke/bitstream/handle/123456789/19048/Determination%20of%20Fecal%20Contamination%20Status.pdf? isAllowed=y&sequence=1

<1% match (Internet from 03-Apr-2021)
https://ir-
library.ku.ac.ke/bitstream/handle/123456789/14335/Travel%20agencies%20response%20to%20internet.....pdf? isAllowed=y&sequence=1

<1% match (publications)
Federico Perrotta, Tony Parry, Luis C. Neves. "Application of machine learning for fuel consumption modelling of trucks", 2017 IEEE International Conference on Big Data (Big Data), 2017

<1% match (student papers from 04-Jul-2018)
Submitted to Mount Kenya University on 2018-07-04

<1% match (student papers from 01-Aug-2018)
Submitted to Mount Kenya University on 2018-08-01

<1% match (publications)
Preetham Ganesh, Reza Etemadi Idgahi, Chinmaya Basavanahally Venkatesh, Ashwin Ramesh Babu, Maria Kyrarini. "Personalized system for human gym activity recognition using an RGB camera", Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, 2020

<1% match (Internet from 04-Sep-2021)
https://www.iprjb.org/journals/index.php/GJHS/article/download/743/887/

<1% match (publications)
Girum Gebremeskel Kanno, Adane Tesfaye Anbesse, Mohammed Feyisso Shaka, Miheret Tesfu Legesse, Sewitemariam Desalegn Andarge. "Investigating the effect of biomass fuel use and Kitchen location on Maternal Report of Birth size: A Cross-Sectional Analysis of 2016 Ethiopian Demographic Health Survey data", Cold Spring Harbor Laboratory, 2020

<1% match (student papers from 07-May-2021)
Submitted to University of Hertfordshire on 2021-05-07

<1% match (Internet from 10-May-2020)
http://www.datascienceconsultant.net

<1% match (publications)
M. H. Alderman. "Low Birth Weight: Race and Maternal Nativity--- Impact of Community Income", PEDIATRICS, 01/01/1999

<1% match (publications)
Omkar G. Kaskar, Elaine Wells-Gray, David Fleischman, Landon Grace. "Evaluating Machine Learning Classifiers for Glaucoma Referral Decision Support in Primary Care Settings", Research Square Platform LLC, 2021

<1% match (student papers from 31-Jan-2021)
Submitted to Universiti Teknologi MARA on 2021-01-31

<1% match (student papers from 19-Jul-2006)
Submitted to University of Ulster on 2006-07-19

<1% match (Internet from 15-Jul-2020)
https://pdfs.semanticscholar.org/2115/1cdcd3112dcb0ab0169b02c6fdc798db06f1.pdf

<1% match (publications)
Fangxiong Chen, Guoheng Huang, Huishi Wu, Ke Hu, Weiwen Zhang, Lianglun Cheng. "Foetal weight prediction based on improved PSO-GRNN model", International Journal of Data Mining and Bioinformatics, 2020

<1% match (Internet from 08-Sep-2021)
https://www.ijbio.com/articles/evaluation-of-logistic-regression-on-mode-of-the-delivery-of-expectant-mothers.pdf

<1% match (publications)
Sundararaghavan, P.S.. "Sequencing questions to ferret out terrorists: Models and heuristics", Omega, 201002/04

<1% match (Internet from 02-Jun-2021)
https://arxiv.org/abs/2001.09636

<1% match (Internet from 25-Oct-2018)
https://dhsprogram.com/pubs/pdf/WP131/WP131.pdf

<1% match (Internet from 10-Jan-2021)
https://dhsprogram.com/pubs/pdf/QRS22/QRS22.pdf

<1% match (publications)
Abdullah Zahirzada, Kittichai Lavangnananda. "Implementing Predictive Model for Low Birth Weight in Afghanistan", 2021 13th International Conference on Knowledge and Smart Technology (KST), 2021

<1% match (publications)
Mendes, Carolina Queiroz de Souza, Bruna Cristina de Almeida Cacella, Myriam

Aparecida Mandetta, and Maria Magda Ferreira Gomes Balieiro. "Baixo peso ao nascer em município da região sudeste do Brasil", Revista Brasileira de Enfermagem, 2015.

<1% match (publications)
Rajat Das Gupta, Krystal Swasey, Vanessa Burrowes, Mohammad Rashidul Hashan, Gulam Muhammed Al Kibria. "Factors associated with low birth weight in Afghanistan: a cross-sectional analysis of the demographic and health survey 2015", BMJ Open, 2019

<1% match (publications)
Yu Lu, Xianghua Fu, Fangxiong Chen, Kelvin K.L. Wong. "Prediction of fetal weight at varying gestational age in the absence of ultrasound examination using ensemble learning", Artificial Intelligence in Medicine, 2020

<1% match (Internet from 21-May-2021)
https://revistas.unal.edu.co/index.php/revfacmed/article/view/61577

<1% match (publications)
Alireza Kajabadi, Mohamad Hosein Saraee, Sedighe Asgari. "Data mining cardiovascular risk factors", 2009 International Conference on Application of Information and Communication Technologies, 2009

<1% match (publications)
Gauri Shrestha. "Factors Affecting Maternal Health Care Services Utilization in Nepal: Insight from the Nepal Demographic Health Survey 2006 and 2011", Nepalese Journal of Statistics, 2017

<1% match (publications)
Shekhar Chauhan , Ratna Patel. "Risk of low birth weight and exposure to type of cooking fuel in India", International Journal of Pregnancy & Child Birth, 2020

<1% match (Internet from 23-Aug-2020)
http://etd.aau.edu.et

<1% match (Internet from 28-Dec-2020)
http://export.arxiv.org

<1% match (Internet from 27-Aug-2020)
http://export.arxiv.org

<1% match (Internet from 10-Apr-2019)
https://gupea.ub.gu.se/bitstream/2077/53615/1/gupea_2077_53615_1.pdf

<1% match (Internet from 26-May-2021)
http://repository.kemu.ac.ke:8080

<1% match (Internet from 15-Feb-2015)
http://www.saathii.org

<1% match (Internet from 04-Jan-2018)
https://www.scribd.com/document/325455554/1-The-Effect-of-Nonperforming-Loans-on-Profitability

<1% match (student papers from 30-Jan-2021)
Submitted to Accra Business School on 2021-01-30

<1% match (student papers from 10-Aug-2021)
Submitted to Coventry University on 2021-08-10

<1% match (Internet from 26-May-2019)
https://es.scribd.com/document/377642469/3003034-Je-My-Mercy-Thomas

<1% match (Internet from 19-Jul-2020)
https://ruor.uottawa.ca/bitstream/10393/28695/1/MR73775.PDF

<1% match (Internet from 09-Jul-2020)
https://www.imedpub.com/articles/a-review-of-low-birth-weight-in-ethiopia-sociodemographic-and-obstetric-risk-factors.php? aid=22406

<1% match (student papers from 09-Nov-2017)
Submitted to International Health Sciences University on 2017-11-09

<1% match (Internet from 01-Apr-2021)
http://studentsrepo.um.edu.my

<1% match (publications)
Qing Cai, Mohamed Abdel-Aty, Scott Castro. "Explore effects of bicycle facilities and exposure on bicycle safety at intersections", International Journal of Sustainable Transportation, 2020

<1% match (student papers from 20-Jan-2021)
Submitted to RDI Distance Learning on 2021-01-20

<1% match (publications)
Rosnah Sutan, Mazlina Mohtar, Aimi Nazri Mahat, Azmi Mohd Tamil. "Determinant of Low Birth Weight Infants: A Matched Case Control Study", Open Journal of Preventive Medicine, 2014

<1% match (student papers from 28-Jul-2017)
Submitted to University of Nairobi on 2017-07-28

<1% match (student papers from 06-Nov-2017)
Submitted to University of Queensland on 2017-11-06

<1% match (Internet from 26-Jul-2020)
https://vm36.upi.edu/index.php/ijost/article/view/10799

<1% match (student papers from 28-Sep-2015)
Submitted to Chester College of Higher Education on 2015-09-28

<1% match (publications)
Graciane Radaelli, Eduardo Leal-Conceição, Felipe K. Neto, Melissa R. G. Taurisano et al. "Motor and cognitive outcomes of low birth weight neonates born in a limited resource country: a systematic review", Cold Spring Harbor Laboratory, 2020

<1% match (student papers from 12-Aug-2021)
Submitted to Liverpool John Moores University on 2021-08-12

<1% match (student papers from 13-Jan-2015)
Submitted to University of Lancaster on 2015-01-13

<1% match (student papers from 01-Apr-2015)
Submitted to University of Wolverhampton on 2015-04-01

<1% match (Internet from 25-Feb-2021)
http://hnmj.gums.ac.ir

<1% match (student papers from 20-Aug-2018)
Submitted to Imperial College of Science, Technology and Medicine on 2018-08-20

<1% match (publications)
Nuradin Abusha Katiso, Getachew Mullu Kassa, Gedefaw Abeje Fekadu, Abadi Kidanemariam Berhe, Achenef Asmamaw Muche. "Prevalence and Determinants of Low Birth Weight in Ethiopia: A Systematic Review and Meta-Analysis", Advances in Public Health, 2020

<1% match (publications)
Weizhe Ding, Li Zhang, Yang Nan, Juanshu Wu, Xiangxin Xin, Chenyang Han, Siyuan Li, Hongsheng Liu. "Combining Multi Dimensional Molecular Fingerprints to Predict hERG Cardiotoxicity of Compounds", Cold Spring Harbor Laboratory, 2021

<1% match (publications)
Kamalesh Kumar Patel, Jyoti Vijay, Abha Mangal, Daya Krishan Mangal, Shiv Dutt Gupta. "Burden of anaemia among children aged 6–59 months and its associated risk factors in India – Are there gender differences?", Children and Youth Services Review, 2021

<1% match (student papers from 26-Sep-2018)
Submitted to University of Rwanda on 2018-09-26

<1% match (Internet from 28-Jul-2021)
https://prism.ucalgary.ca/bitstream/handle/1880/113211/ucalgary_2021_lowton_dana.pdf

<1% match (Internet from 20-Jul-2021)
https://turkiyeklinikleri.com/article/tr-dogum-agirligini-etkileyen-faktorler-uzerine-bir-arastirma-51189.html

<1% match (Internet from 18-Apr-2021)
https://www.tripdatabase.com/search?criteria=water+birth&lang=en&page=2

<1% match (publications)
Feresu, Shingairai A., Siobán D. Harlow, and Godfrey B. Woelk. "Risk Factors for Low Birthweight in Zimbabwean Women: A Secondary Data Analysis", PLoS ONE, 2015.

<1% match (publications)
François R. Jornayvaz, Peter Vollenweider, Murielle Bochud, Vincent Mooser, Gérard Waeber, Pedro Marques-Vidal. "Low birth weight leads to obesity, diabetes and increased leptin levels in adults: the CoLaus study", Cardiovascular Diabetology, 2016

<1% match (publications)
Mesfin Wudu Kassaw, Ayele Mamo Abebe, Ayelign Mengesha Kassie, Biruk Beletew Abate, Seteamlak Adane Masresha. "Trends of proximate low birth weight and associations among children under-five years of age: Evidence from the 2016 Ethiopian demographic and health survey data", PLOS ONE, 2021

<1% match (Internet from 25-Jun-2021)
https://eprajournals.com/jpanel/upload/EPRA%20IJMR%20JANUARY-2021%20FULL%20JOURNAL.pdf#page=162

<1% match (Internet from 07-Apr-2021)
https://iopscience.iop.org/article/10.1088/1742-6596/1282/1/012010

<1% match (Internet from 13-Sep-2021)
https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00461-7

<1% match (Internet from 12-Jul-2021)
https://repository.cardiffmet.ac.uk/bitstream/handle/10369/8525/Bamokarah%2c%20Saud%20Salem.pdf

<1% match (publications)
"Poster Presentations", International Journal of Gynecology & Obstetrics, 2015.

<1% match (publications)
Emmanuel Biracyaza, Samuel Habimana, Donat Rusengamihigo, Heather Evans. "Regular antenatal care visits were associated with low risk of low birth weight among newborns in Rwanda: Evidence from the 2014/2015 Rwanda Demographic Health Survey (RDHS) Data", F1000Research, 2021

<1% match (publications)
M. J. Msolla, J. L. Kinabo. "Prevalence of anaemia in pregnant women during the last trimester", International Journal of Food Sciences and Nutrition, 2009

<1% match (publications)
Susan Thurstans, Natalie Sessions, Carmel Dolan, Kate Sadler et al. "The relationship between wasting and stunting in young children: A systematic review", Maternal & Child Nutrition, 2021

<1% match (student papers from 28-Aug-2020)
Submitted to University of Leeds on 2020-08-28

<1% match (publications)
Wei Sun, Zhiwei Xu. "A novel hourly PM2.5 concentration prediction model based on feature selection, training set screening, and mode decomposition-reorganization", Sustainable Cities and Society, 2021

<1% match (Internet from 17-Aug-2019)
https://nutritionj.biomedcentral.com/articles/10.1186/s12937-018-0409-z

<1% match (Internet from 01-Aug-2021)
https://www.jpmh.org/index.php/jpmh/issue/download/98/54

<1% match (Internet from 24-Aug-2021)
https://www.orfonline.org/research/covid-19-compounds-global-challenges-to-food-security/

<1% match (Internet from 16-Jun-2019)
http://www.renupublishers.com

<1% match (Internet from 17-May-2019)
https://www.science.gov/topicpages/b/birth+weight+prematurity.html

<1% match (publications)
Abdulbasit Musa, Catherine Chojenta, Deborah Loxton. "The association between intimate partner violence and low birth weight and preterm delivery in eastern Ethiopia:

Findings from a facility-based study", Midwifery, 2021

<1% match (publications)
Chhorvann Chhea, Por Ir, Heng Sopheab. "Low birth weight of institutional births in Cambodia: Analysis of the Demographic and Health Surveys 2010-2014", PLOS ONE, 2018

<1% match (publications)
Desalegn Abebaw Jember, Zeleke Argaw Menji, Yibeltal Asmamaw Yitayew. " Low Birth Weight and Associated Factors Among Newborn Babies in Health Institutions in Dessie, Amhara, Ethiopia ", Journal of Multidisciplinary Healthcare, 2020

<1% match (publications)
Ergaz, Z.. "Intrauterine growth restriction-etiology and consequences: What do we know about the human situation and experimental animal models?", Reproductive Toxicology, 200509/10

<1% match (publications)
Eugene Budu, Abdul-Aziz Seidu, Ebenezer Kwesi Armah-Ansah, Francis Sambah, Linus Baatiema, Bright Opoku Ahinkorah. "Women's autonomy in healthcare decision-making and healthcare seeking behaviour for childhood illness in Ghana: Analysis of data from the 2014 Ghana Demographic and Health Survey", PLOS ONE, 2020

<1% match (publications)
GOBOPAMANG LETAMO, ROLANG MAJELANTLE. "FACTORS INFLUENCING LOW BIRTH WEIGHT AND PREMATURITY IN BOTSWANA", Journal of Biosocial Science, 2001

<1% match (publications)
Onesmus Maina Muchemi, Elizabeth Echoka, Anselimo Makokha. "Factors associated with low birth weight among neonates born at Olkalou District Hospital, Central Region, Kenya", Pan African Medical Journal, 2015

<1% match (publications)

Rebecca Niemiec, Megan S. Jones, Andrew Mertens, Courtney Dillard. "The effectiveness of COVID-related message framing on public beliefs and behaviors related to plant-based diets", Appetite, 2021

<1% match (publications)
Yunqi Li. "A novel scoring function for discriminating hyperthermophilic and mesophilic proteins with application to predicting relative thermostability of protein mutants", BMC Bioinformatics, 2010

<1% match (publications)
Zakir Hussain, Malaya Dutta Borah. "Birth weight prediction of new born baby with application of machine learning techniques on features of mother", Journal of Statistics and Management Systems, 2020

<1% match (publications)
Ziaullah Momand, Pornchai Mongkolnam, Pichai Kositpanthavong, Jonathan H. Chan. "Data Mining Based Prediction of Malnutrition in Afghan Children", 2020 12th

International Conference on Knowledge and Smart Technology (KST), 2020

<1% match (Internet from 01-Apr-2021)
https://bmc.edu.np/wp-content/uploads/2021/03/BMC-Journal-of-Sc.-Research-Vol-3-Dec-2020.pdf#page=28

<1% match (Internet from 14-Nov-2020)
https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-09456-0

<1% match (Internet from 29-Aug-2017)
https://brage.bibsys.no/xmlui/bitstream/handle/11250/2414034/14506_FULLTEXT.pdf?isAllowed=y&sequence=1

<1% match (Internet from 19-Jul-2021)
https://bspace.buid.ac.ae/bitstream/handle/1234/1449/2016228205.pdf

<1% match (Internet from 19-Feb-2021)
Shttps://dokumen.pub/in-silico-drug-design-repurposing-techniques-and-methodologies-1nbsped-0128161256-9780128161258.html

<1% match (Internet from 06-Jul-2021)
http://eprints.poltekkesjogja.ac.id

<1% match (Internet from 19-Sep-2019)
https://escholarship.org/content/qt23d272xg/qt23d272xg.pdf?t=pxprzz

<1% match (Internet from 11-Aug-2021)
http://essay.utwente.nl

<1% match (Internet from 15-Jul-2020)
https://file.scirp.org/Html/6-1761221_74760.htm

<1% match (Internet from 24-Sep-2020)
https://journals.sagepub.com/doi/10.4137/CMPed.S40070

<1% match (Internet from 23-Aug-2021)
http://research.wsulibs.wsu.edu:8080

<1% match (Internet from 29-Nov-2017)
http://scholarbank.nus.edu.sg

<1% match (Internet from 25-Nov-2019)
https://serval.unil.ch/notice/serval:BIB_294E477D2C7F

<1% match (Internet from 19-Apr-2021)
https://www.e3s-conferences.org/articles/e3sconf/pdf/2020/74/e3sconf_ebldm2020_01023.pdf

<1% match (Internet from 13-Aug-2019)
https://www.frontiersin.org/articles/10.3389/fendo.2019.00055/full

<1% match (Internet from 09-Aug-2020)
https://www.nature.com/articles/s41598-019-39071-y?code=d7d89f99-5e6e-4c2e-bbe8-a326efdd1f9b&error=cookies_not_supported

<1% match (Internet from 07-Oct-2020)
https://www.omicsonline.org/open-access/stillbirth-in-a-university-maternity-of-portonovo-in-southern-benin-epidemiological-and etiological-aspects-2376-127X-1000353-94801.html

<1% match (Internet from 17-Mar-2020)
https://www.researchsquare.com/article/rs-4025/v1

<1% match (Internet from 09-May-2019)
https://www.tandfonline.com/doi/full/10.1080/02102412.2015.1118903

<1% match (publications)
Amare Alamirew Aynie, Tigabu Birhan Kassa, Dagninet Derebe Abie. "Prevalence of Low Birth Weight and Its Determinants in Bahir Dar City, Amhara Region, North West Ethiopia: Health Facility Based Cross-Sectional Study", Biomedical Statistics and Informatics, 2020

<1% match (publications)
Huaiting Gu, Lixia Wang, Lingfei Liu, Xiu Luo et al. "A gradient relationship between low birth weight and IQ: A meta-analysis", Scientific Reports, 2017

<1% match (publications)

Tesfahun Mulatu Wachamo, Nigus Bililign Yimer, Asmamaw Demis Bizuneh. "Risk factors for low birth weight in hospitals of North Wello zone, Ethiopia: A case-control study", PLOS ONE, 2019

<1% match (publications)
Vasudha Bhatnagar. "VisTree: Generic Decision Tree Inducer and Visualizer", Lecture Notes in Computer Science, 2010

<1% match ()
Nesara, Paul. "Determinants of Low Birth Weight in a Population-Based Sample of Zimbabwe", ScholarWorks, 2018

<1% match (publications)
Andualem Zenebe, Kaleab Tesfaye Tegegne, Berhanu Bifato, Abiyu Ayalew Assefa. "Association between iron and folic acid supplementation and birth weight in Ethiopia: systemic review and meta analysis", Bulletin of the National Research Centre, 2021

<1% match (publications)
Carmen Martínez-Gil, A. López-López. "Chapter 14 Answer Extraction for Definition Questions using Information Gain and Machine Learning", Springer Science and Business Media LLC, 2008

<1% match (publications)
Helma Jane Ferreira Veloso, Antônio Augusto Moura da Silva, Heloísa Bettiol, Marcelo Zubarán Goldani et al. "Low birth weight in São Luís, northeastern Brazil: trends and associated factors", BMC Pregnancy and Childbirth, 2014

<1% match (publications)
Ning Eliyati, Alfensi Faruk, Endang Sri Kresnawati, Ika Arifieni. "Support vector machines for classification of low birth weight in Indonesia", Journal of Physics: Conference Series, 2019

Machine Learning Prediction of Low Birth Weight in Kenya using Maternal Risk Factors. Masters Research Thesis By Sharon Jepkorir Sawe Registration Number:

220000140 A dissertation submitted in partial fullment of the requirements for the degree of MASTER OF SCIENCE IN DATA SCIENCE (OPTION: BIOSTATISTICS). In the college of Business and Economics Supervisor: Dr Dieudonné Muhoza September 2021 i DECLARATION I declare that this dissertation entitled Machine Learning Prediction of Low Birth Weight in Kenya using Maternal Risk Factors 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. Sharon Jepkorir Sawe Reg No: 220000140 Signature: Date: __9/13/2021_____ ii APPROVAL SHEET This dissertation entitled Machine Learning Prediction of Low Birth Weight in Kenya using Maternal Risk Factors written and submitted by SHARON JEPKORIR SAWE in partial fulfilment of the requirements for the degree of Master of Science in Data Science majoring in Biostatistics is hereby accepted and approved. The rate of plagiarism tested using Turnitin is 18 % which is less than 20% accepted by ACE-DS. _____ Supervisor _____ Head of Training iii ACKNOWLEDGEMENTS I wish to express my sincere gratitude to my supervisor Dr. Dieudonné N. Muhoza for his guidance, positive criticism and a lot of patience during his supervision of this thesis. My sincere gratitude also goes to my dear parents Mr. Daniel Sawe and Mrs Teresa Sawe, for their love, support and prayers which made everything worthwhile. I must express lots of appreciation to the entire staff of Africa Centre of Excellence including Dr Charles Ruranga, Dr Ignace Kabano and other members of staff at large for their support. Finally I wish to thank my friends and classmates for their constant assistance during write up of this thesis. iv ABSTRACT A new born's health is a primary factor that determines the overall health of a human being and its life expectancy. Therefore, its health should be monitored not only after birth but also when the baby is still growing in the womb. Birth weight is one of the crucial aspects to be observed. Low birth weight is among the main problems that new borns face. Low birth weight (LBW) is the weight at birth less than 2500g as defined by the World Health Organization. A global estimate of15 to 20 percent of total live births are low birth weight representing over twenty million births every year . In Kenya, the rate of children born with low weight is 8 percent. Several methods have been used to measure and approximate birth weight in clinical practice including obstetric ultrasound, symphysio-fundal height measurements and abdominal palpation. However, these methods are associated with reliability and accuracy challenges therefore, calling for more robust methods. This research aimed at creating a machine learning model for predicting low birth weight using the maternal risk factors that have been found to be associated with low birth weight. Secondary data from the 2014 Kenya Demographic Health Survey was utilized where the variables were extracted from the births recode file. The study population included mothers between the age of 15 to 49 years. The machine learning algorithms employed were logistic regression, decision trees, random forest, support vector machines, gradient boosting and xtreme gradient boosting. Using performance evaluation metrics namely; accuracy, precision, recall, F1 score, and ROC- AUC, the random forest model was found out to be the most robust with 0.956679 accuracy, 0.956831 precision, 0.956679 recall an F1 score of 0.95666 and an AUC of 0.988. In addition, variable importance was performed using the random forest approach to ascertain the maternal risk factors that are the most important to predict low birth weight. It was found out that mother's weight was the most important variable for predicting low birth weight. The other important variables found were; mothers height, mother's age and the number of antenatal visits attended by the mother during pregnancy. Machine learning techniques are increasingly being used to provide

information to guide health policy. This research merits further modelling, research and more consultation. v KEYWORDS Machine learning, birth weight, low birth weight, maternal risk factors, prediction, algorithm vi LIST OF SYMBOLS AND ACRONYMS LBW-Low Birth Weight WHO-World Health Organization UNICEF-United Nations Children's Fund WHA-World Health Assembly UHC-Universal Health Coverage SMOTE-Synthetic Minority Oversampling Technique XGB Xtreme Gradient Boosting SVM-Support Vector Machine EDA-Exploratory Data Analysis ROC-Receiver Operating Characteristic Curve AUC-Area Under the Curve MRE-Mean Relative Error IUGR-Intrauterine Growth Restriction BMI-Body Mass Index SGA Small for Gestational Age PAR-Population Attriburable Risk IQ-Intelligence Quotient KDHS-Kenya Demographic and Health Surveys HIV-Human Immunodeficency Virus vii Table of Contents

LIST OF TABLES

LIST OF FIGURES

 1HASH(0x7f233782da38) crucial element in the general health of a nation and even global health. It is the primary factor that determines the overall health of a human being and the life expectancy. Therefore, a baby's health and well-being should be monitored not only after birth but also when the baby is still growing in the womb. One of the aspects that should be observed before the baby is born is its weight. Birth weight is the new born baby's first weight measured immediately after being born within the first hour before occurrence of significant loss of weight due to postnatal effects (UNICEF and WHO, 2004). A new-born's weight signifies a lot about the future health and survival of the baby. Therefore, it is advisable to know whether the baby is going to have normal weight or low weight during birth in order to make early interventions before birth. HASH(0x7f233782db40) that is measured at birth as HASH(0x7f233782e960) major HASH(0x7f233782df18) womb / HASH(0x7f233782eb40) births occur in a period below 259 days since the start of the last menstruation of a woman preceding conception or before completing a gestation period of 37 weeks as WHO defines (WHO, 2012). On the other hand, Intrauterine growth restriction is the below normal rate of foetal growth with respect to the growth potential of the infant in terms of its gender and race. An infant's normal HASH(0x7f233782f080) with exclusion of malnutrition and growth retardation features (Sharma et al., 2016). HASH(0x7f233782f128) there after in life they may develop chronic diseases (HASH(0x7f233782f788) estimated regionally to be 9 percent in Latin America, 28 percent HASH(0x7f2337832120). However, these rates could probably be an underestimate because, not all 2 women get access to giving birth in hospitals therefore these deliveries are not recorded since they deliver at home. Moreover, deliveries that occur in small clinics may go unreported by public official figures (WHO, 2014). Globally, a prevalence reduction of low birth weight by 30 percent in 2025 has been targeted by the World Health Assembly (WHO, 2014). In Kenya, Universal Health Coverage (UHC) is one of the big 4 agenda of which new born health is among its important indicators. The rate of HASH(0x7f233782f740) a report by (EVERY PREEMIER SCALE, 2017). This rate is still alarming and therefore appropriate solutions towards this problem should be sought. Several methods have been used to measure and approximate birth weight in clinical practice. The methods include; obstetric ultrasound, symphysio-fundal height measurements and abdominal palpation. Obstetric ultrasound stands out to be the most reliable method of examining the growth of the foetus. However, ultrasound is not easily accessed in low-resource areas and poor communities. Therefore, the other two methods are applied which are not very reliable in terms of accuracy (Lu et al., 2019). Moreover, training for ultrasound is very crucial. Unskilled ultrasound sonographers might lead to inaccurate foetal weight measurements. Therefore, good training is paramount (Jan-Simon Lanowski et al., 2017). Due to these challenges, another route towards tackling LBW estimation should be taken. Robust methods to estimate HASH(0x7f23378324f8) taken into high consideration. This is because early detection allows for proper and effective obstetric interventions. Recently, data mining methods particularly machine learning have been discovered to be of great help in predicting HASH(0x7f2337832a20) done by using machine learning techniques particularly supervised learning HASH(0x7f2337832e88) involves pattern recognition and

computational learning. It inspects the construction and the study of algorithms which can make predictions by learning from data (Dönmez, 2013). Supervised HASH(0x7f23378339f0) that is used to acquire the system's information based on a set of labelled input-output samples. The goal is to predict output given new inputs (Liu & Wu, 2012). In this research, the data will be labelled in such a way that the output variable low birth weight is binary in nature. Therefore, the machine learning task will follow a supervised learning approach. 3 In Kenya particularly, several studies have been conducted HASH(0x7f2337833dc8). However, limited research has been geared towards predicting babies at risk of being born with low weight. Therefore, the identified maternal risk factors from previous researches can be consolidated and HASH(0x7f2337833e70)HASH(0x7f2337833ff0) use the 2014 Kenya demographic health survey data to perform predictive modelling of HASH(0x7f2337833a50)HASH(0x7f23378343e0). HASH(0x7f2337837258) major challenge globally and nationally. Several interventions have been put in place but it remains a public health problem. It is a global war whereby it led to the 2012 global nutrition targets. HASH(0x7f23378376c0) 2025 worldwide was adopted by the member states HASH(0x7f23378379f0), up to date the globe is still far from accomplishing this objective. The 2000- 2015 report on global trend in low birth weight prevalence by WHO and UNICEF reports that in 2010 to 2015, the reduction in low birth weight prevalence was slow in comparison to 2000 to 2009. It further reports that if the current annual average rate of reduction of 1.00 percent yearly continues, the low birth weight prevalence that was projected to be 10.5 percent, would be 13.2 percent by 2025 (UNICEF and WHO, 2019). Worldwide, approximately 15 percent to 20 percent of total live births are of low weight. This represents over twenty million births every year (WHO, 2014). Moreover, 91 percent of the low birth weight livebirths are from countries of middle and low income majorly South Asia with 48 percent and sub-Saharan Africa with 24 percent