A Performance Study of Prediction Model for Preterm Birth and Mode of Delivery Based on Machine Learning Tools.

Prakash S. Chougule [1], Tejaswi S.Kurane[2] , Rahual. H. Waliv[3] ,Ramesh D.Shinde[4], Miss. Ankita V. Chougale[5], Miss. Sakshi R. Chaugule[6], Miss. Rutuja R. Chavan[7]

[1] Associate Professor, Rajarshi Chhatrapati Shahu College, Kolhapur(MS), India
[2] Assistant Professor, Rajarshi Chhatrapati Shahu College, Kolhapur (MS), India
[3] Associate Professor, Kisan Veer Mahavidyalaya, Wai Dist.Satara (MS), India
[4] Assistant Professor, Jaysingpur College, Jaysingpur (MS), India
[5,6,7] Research Students, Rajarshi Chhatrapati Shahu College, Kolhapur (MS), India

Abstract: Preterm births affect around 15 million children a year worldwide. Current development in science and technology in medical field the efforts focus on mitigating the effects of the pre-maturity, not on preventing it. In this study, we studied different maternal health factors like Age, Height, Weight, BMI, BMI Category that affect on preterm birth and Similarly, prediction of mode of delivery based on Age, Height, Weight, Blood Pressure, TSH Level, HB . The objective of this study was to develop and compare machine learning predictive models for preterm birth based on logistic regression, random forest and stratified k-fold cross validation machine learning algorithms for detecting the delivery method is Normal or C- section. our study shows that model based on logistic regression show high accuracy than random forest for predicting Preterm births whereas model based on random forest shows excellent performance with high accuracy than logistic regression and stratified k-fold cross validation for prediction of mode of delivery. Therefore we use random forest model is used for real data prediction.

Keywords: Machine Learning, Logistic Regression, Stratified K-fold Cross Validation, Random Forest, ROC -AUC Curve.

Introduction: Preterm births affect around 15 million children a year worldwide. Current development in science and technology in medical field the efforts focus on mitigating the effects of the pre-maturity, not on preventing it.Although caesarean(C-Section) deliveries hold life-saving potential, their increasing rate poses a substantial global health challenge. By 2030, it is projected that around 28.5% of all global births will involve caesarean sections(C-Section), which equates to approximately 38 million women annually [1]. In India, CD rates have steadily increased. Data from the National Family Health Surveys (NFHS) demonstrate an increase in the CD rates 8.5% in 2005–2006, 17.2% in 2015–2016, and 21.5% in 2019–2021[2].Preterm birth refers to the delivery of a baby completing 37 weeks of gestation it is a global public health issue and one of the leading causes of Neonatal Mortality and Morbidity. Nearly 15 million infants are born prematurely worldwide, and more than 1 million die from preterm birth and its complications before the age of 5 [3]. Understanding the cause risk factors and prevention strategies is essential for improving maternal and child health outcomes. Recent data indicates that India recorded the highest number of preterm births old wide in 2020 with approximately 3.02 million cases, according for 20% of all preterm births globally.PTB not only causes death and diseases in the new born, but also causes anxiety and depression in postpartum women [4]. The choice of delivery mode has a significant impact on the health of both mothers and infants. With the continuous advancement of medical technology, the global utilization rate of caesarean section as a crucial delivery method is increasing [5]. Therefore, it is essential to deeply understand the factors influencing the mode of delivery for predicting and preventing caesarean section. The rate of C-section deliveries in India has risen significantly over the years, increasing from 8.5% in 2005-06 to 21.5% in 2019-21. This increasing rate also raises questions about whether all C-section are medically necessary. Maternal health place a crucial role in determining the mode of delivery, which can be either vaginal or caesarean section (C-section). Understanding and the predicting model of delivery based on maternal health parameters have becomes essential for optimizing maternal health outcomes. Advancement and technology, particularly in machine learning (ML) and artificial intelligence (AI) have, significantly enhanced the ability to predict mode of delivery by analysing various maternal health parameters. Our research focuses on exploring factors affecting both C-section and normal deliveries. In this study examine how Age, Weight, Height, TSH levels, Blood Pressure, HB contribute to delivery outcomes. The objective of this study was to develop and compare machine learning predictive models for preterm birth based on logistic regression, random forest and stratified k-fold cross validation machine learning algorithms for detecting the delivery method is Normal or C- section.our predictive model assist health care providers in making informed decisions there by improving maternal health outcomes.

Objectives:

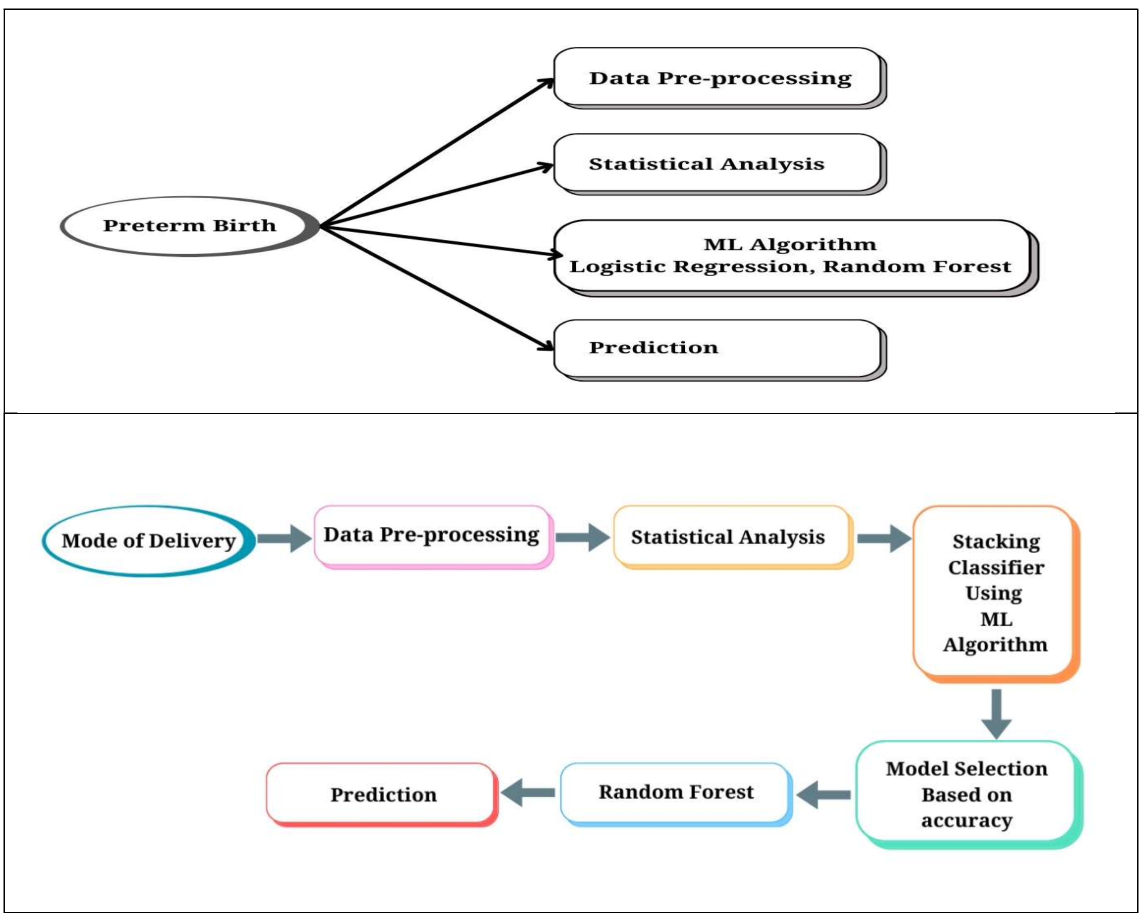
- Statistical Analysis of factors affecting Preterm birth.
- To build a predictive model to detect preterm birth based on maternal health factors.
- To analyze whether Age and BMI Category differ significantly in both group (preterm risk and no preterm risk).
- To analyze age-wise distribution of mode of delivery.
- To analyse the relationship between the mode of Normal vs. caesarean delivery and maternal age.
- To evaluate the prevalence of anemia among women following different delivery methods
- To build a predictive model based on maternal health factors.

Literature Review: Preterm birth is becoming a vital public health concern, given its correlation with neonatal mortality and morbidity. The causation of preterm birth is complex and multifactorial issue therefore remains the subject of considerable research and investigation. A study conducted by Richard P Dickey, Xu Xiong examined the effect of Height, Weight, Body mass index ion the risk of preterm birth, this study showed that obesity and short stature significantly increase the risk of preterm birth. Another study conducted by Sarah, zhen Han showed that risk of induced preterm birth was increased in over-weight and obese one. Therefore it is also essential to evaluate effect of weight. In maternal care, a quick decision has to be made about most suitable delivery type for the current patient. The increasing rate are observed rate in India ha increased from 17.2% in 2016 to 21.15% in 2021.However challenges remain in identifying who is at greater risk and ensuring that intervention are delivered early when they have greatest potential to benefit. A study conducted by Ibewesi, MaryJane Chinyer employed machine learning methods like decision tree algorithm, SVM, logistic regression to predict outcome of a pregnant women based on her vital signs and concluded that decision tree achieved highest accuracy (86.99%). Another study conducted by Tamala Gondwe, Kalpana Betha showed that more women to delivered by C-Section where older than 25 at their first pregnancy and had been diagnosed with gestational diabetes, hypertension and anemia. More women who delivered by C-Section where overweight in BMI Category before pregnancy, more women who delivered by cesarean had one or more labor complications, Cephalic Disproportion (CPD), non reassuring fetal heart rate pattern, and the fetus in breech or transverse position. So it is essential to study the role of Age, Height, Weight, Blood Pressure, TSH Level in determining the delivery methods In our study conducted on 64 patients, 16% of overweight women have the risk of preterm birth. Overweight BMI category of women have high risk of preterm birth.

Statistical Tools: MS-Excel ,Ms-Word ,

Statistical software : Python

Methodology:



Data Collection:

a) Data is collected from different maternal hospital of Kolhapur city
b) Delivery method (Normal = 0, C-Section = 1) is the dependent variable.

c) The sample size is determined by using Yamen's formula:

$$n = \frac{N}{1+Ne^2}$$; where n=sample size,N=population size and e=margin of error

2) Data Preprocessing:

a) Handle missing values using central tendency.

3) Graphical and Statistical Analysis:

a) Use bar charts, Pie chart, and stacked bar chart, stacked horizontal bar chart for data visualization.

b) Conduct Shapiro-Wilk test for Normality testing and Mann-Whitney U Test for comparing means of variables across mode of delivery.

4) Model Building:

a) Model building for Predicting risk of preterm birth and Mode of Delivery using Machine Learning model.

5) Validation and Testing:

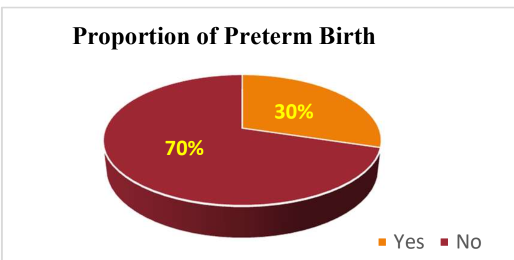
a) Split data into training and testing sets.

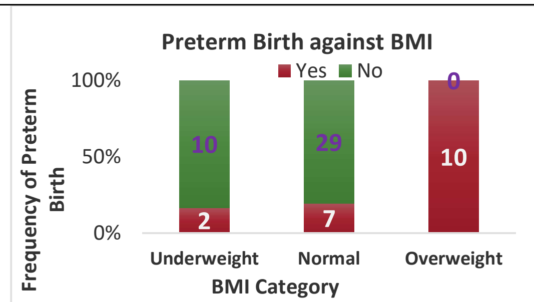
b) Evaluate model performance using accuracy, precision, recall, and F1 Score.

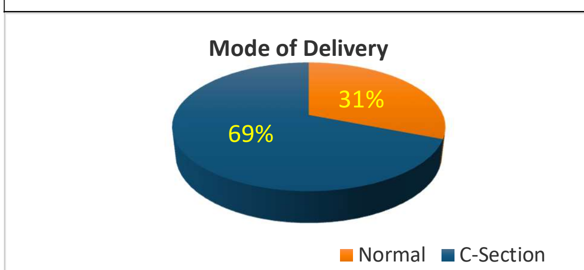
6) Interpretation and Reporting:

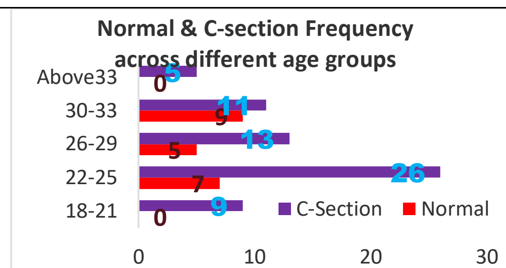
a) Visualize results through graphs (ROC-AUC Curves) and tables.

Graphical Representation:

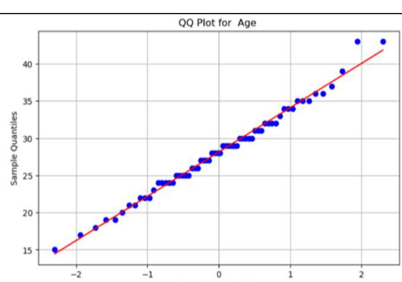
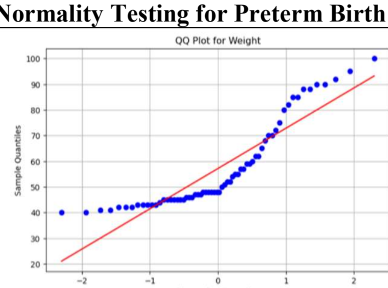
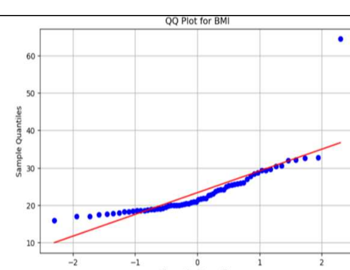


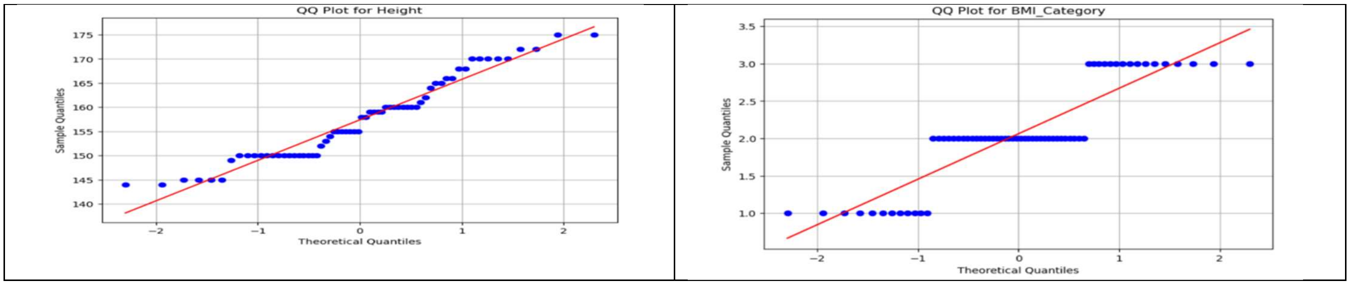




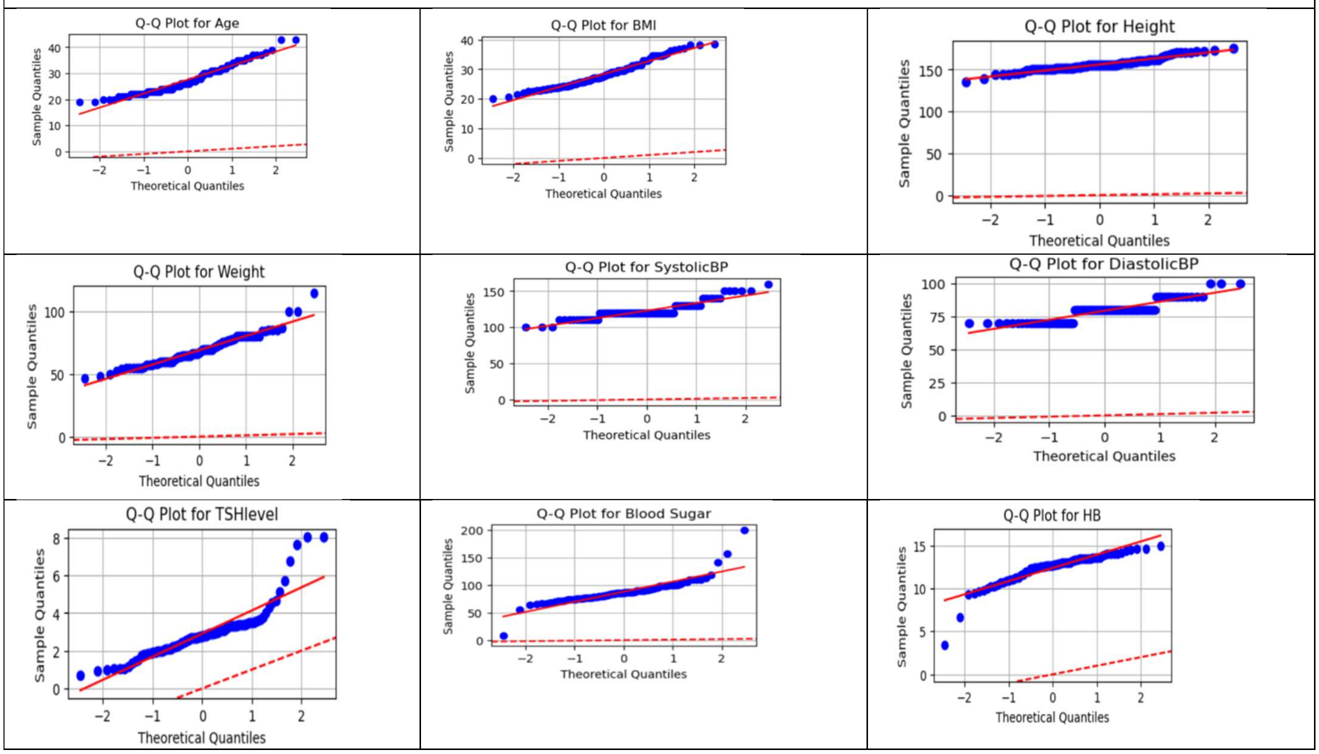


Statistical Analysis:

Normality Testing for Preterm Birth		
		



Normality Testing for Mode of Delivery



Testing Part:

a) **Chi-square test:**

Chi-square test to check independence between BMI-Category and Preterm Birth.
$\mathbf{H_0}$**:** There is no association between BMI Category and Preterm Birth against
$\mathbf{H_1}$**:** There is an association between BMI Category and Preterm Birth.
Under $\mathbf{H_0}$**,** χ^2-statistic: 11.0367 and p-value: 0.004012

a) **Mann Whitney U test:**

Mann Whitney U test to to check whether There is significant difference in the mean age between women who had preterm births and those who did not.
$\mathbf{H_0}$**:** There is no significant difference in the mean age between women who had preterm births and those who did not .against
$\mathbf{H_1}$**:** There is significant difference in the mean age between women who had preterm births and those who did not.
Under $\mathbf{H_0}$**,** Z-statistic: -1.8061 and p-value: 0.0709

c) **Mann Whitney U test:**

Mann Whitney U test to check whether there is a significant difference in the mean age between the two modes of normal and c-section delivery .
$\mathbf{H_0}$**:** There is no significant difference in the mean age between the two modes of normal and c-section delivery.against
$\mathbf{H_1}$**:** There is a significant difference in the mean age between the two modes of normal and c-section

delivery.

Under H_0, Z-statistic: 1451.0 and p-value: 0.00049

Model Building:

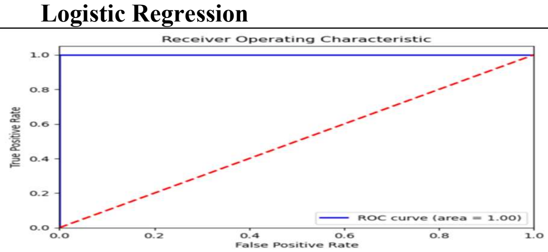
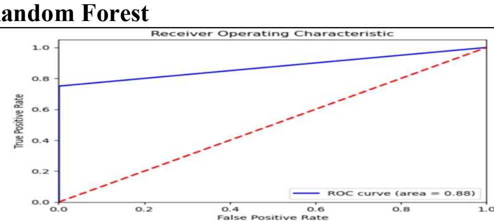
Logistic Regression: It is a classification algorithm for supervised data and also similar to linear regression, the difference between them is logistic regression uses Logistic function/ sigmoid function which is an S shaped curve. Logistic regression, unlike linear regression, does not assume normality of errors or predictor variables.

Random Forest: It is a machine learning algorithm that uses multiple decision trees to make predictions. It's a type of ensemble learning method that's used for classification and regression to improve accuracy and control overfitting**.** Random Forest algorithms, unlike some statistical methods, do not assume normality of the data or any specific distribution. They are robust to non-normal data and can handle complex, non-linear relationships.

Stratified K-fold Cross-Validation: It is particularly useful when dealing with biases and when data sets contain imbalances or when there's a class imbalance in the target variable. Stratified K-fold cross-validation not assuming normality is particularly useful for datasets with imbalanced classes, ensuring each fold maintains the same class distribution as the original dataset, which is crucial for fair model evaluation and training.

a) **Prediction model based on Logistic Regression and_Random Forest for Preterm birth:** The dependent Variable is taken as Preterm birth history and Independent Variables are Age, Height, Weight, BMI_Category

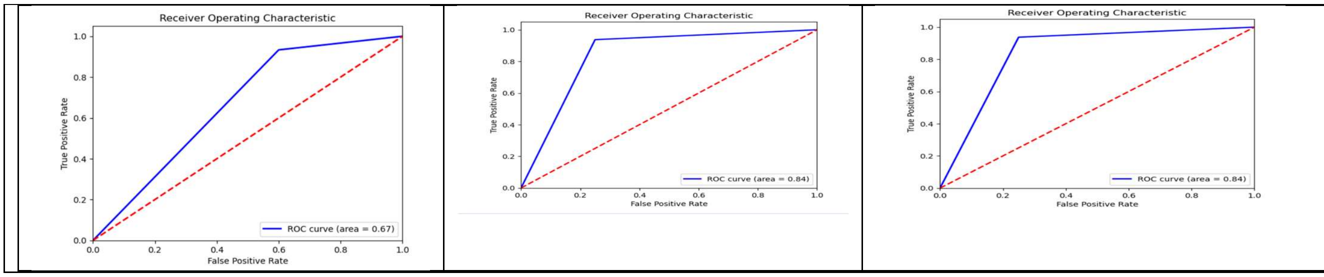
Classification Report for Preterm birth							
		Precision	recall	F1-score	support	Accuracy	ROC-AUC Score
Logistic Regression	1	1.00	1.00	1.00	12	1.0	1.0
	0	1.00	1.00	1.00	4		
Random Forest	0	0.90	1.00	0.95	9	0.9230	0.88
	1	1.00	0.75	0.86	4		

Logistic Regression	**Random Forest**
	

b) **Prediction model based on Logistic Regression and_Random Forest and Stratified k-fold cross validation for mode of delivery:** Dependent Variable is taken as Mode of delivery and Independent Variables are Age, Height, Weight, TSH level, Systolic BP, Diastolic BP, HB

Classification Report for mode of delivery							
		Precision	recall	F1-score	support	Accuracy	ROC-AUC Score
Logistic Regression	0	0.67	0.40	0.50	5	0.8	0.67
	1	0.82	0.63	0.88	15		
Random Forest	0	0.75	0.75	0.75	4	0.9230	0.84
	1	0.94	0.94	0.94	16		
Stratified k-fold cross validation	0	0.75	0.75	0.75	4	0.9	0.84
	1	0.94	0.94	0.94	16		

Logistic Regression	**Random Forest**	**Stratified k-fold cross validation**



Overall Conclusion:

From graphical representation we observed that 30% women have risk of preterm birth and 70% do not have the risk of preterm birth.where as overweight BMI category of women have high risk of preterm birth as compared to underweight and normal category.The proportion of C-section deliveries is significantly higher than normal deliveries.And the highest number of C-Section deliveries is observed in 22-25 age group and the highest number normal deliveries is observed in 30-33 age group.The average age, weight, BMI for women with preterm delivery was observed to be 30, 76.05, 27.20 and 27.72, 49.15,21.72 for no preterm respectively.The average age, weight, BMI, TSH level for women with normal delivery was observed to be 30.56, 70.80, 27.39, 2.74 and 26.20, 68.23,28.78, 2.97 for C-section respectively.There is a significant association between BMI Category and Preterm birth.There is a statistically significant difference in the mean age between the two modes of delivery is either normal or c-section.The findings revealed that 33% of women who underwent caesarean deliveries experienced anemia, while 20% of those who had normal deliveries were found to be anemic. In the model based on logistic regression model performs excellent than random Forest for detecting risk of preterm birth.The model based on random forest performs well than Stratified k- fold cross validation and logistic regression for prediction of modes of delivery.

References:

1. Betran AP, Ye J, Moller A-B, et al. Trends and projections of caesarean section rates: global and regional estimates. BMJ Glob Health 2021; 6: e005671.
2. National family health survey Date accessed May 7, 2024
3. Walani SR. Global burden of preterm birth. Int J Gynaecol Obstet. 2020;150(1):31–3
4. Singer L. T., Salvator A., and Guo S., "Maternal psychological distress and parenting stress after the birth of a very low-birth-weight infant, *JAMA*. (1999) **281**, no. 9, 799–805,
5. Bruno, A.M., Metz, T.D., Grobman, W.A. and Silver, R.M. (2022) Defining a Cesarean Delivery Rate for Optimizing Maternal and Neonatal Outcomes. Obstetrics & Gynecology, 140, 399-407.
6. Woday A, Muluneh DM, Sherif S. Determinants of preterm birth among mothers who gave birth at public hospitals in the Amhara region, Ethiopia: a case-control study, Plos One, pp. 1–12, 2019.
7. Tigist B, Abdela A, Kahsay ZG. Preterm Birth and Associated Factors among mothers who gave Birth in Debremarkos Town Health Institutions, 2013 Institutional Based Cross Sectional Study. Gynecol Obstet. 2015;5(5):1–5.
8. Pusdekar Y, Archana P, Kunal K, Savita B, Vanessa T. Rates and risk factors for preterm birth and low birthweight in the global network sites in six low- and low middle-income countries. Reprod Health. 2020;17(3):1–16. doi: 10.1186/s12978-020-01029-z.
9. Ayebare E, Ntuyo P, Malande OO. Maternal, reproductive and obstetric factors associated with preterm births in Mulago Hospital, Kampala, Uganda: a case control study. Pan Afr Med J. 2018;30:1–8. doi: 10.11604/pamj.2018.30.272.13531.
10. Taha Z, Hassan AA, Ludmilla W-S, Dimitrios P. Factors Associated with Preterm Birth and Low Birth Weight in Abu Dhabi, the United Arab Emirates. Int J Environ Res Public Health. 2020;17:1–10. doi: 10.3390/ijerph17041382.