

“Education for Knowledge, Science and Culture”

-Shikshanmaharshi Dr. Bapuji Salunkhe



(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE KOLHAPUR

(Empowered Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT

on

“STATISTICAL ANALYSIS OF CAESAREAN SECTION FOR
PREGANANT WOMEN”

Submitted by

Ms. Pradnya Bharat Kamble

Ms. Lalita Ramdas Koli

Mr. Ketan Kiran Patil

*in partial fulfillment for the award of
the degree of*

MASTER OF SCIENCE

in

STATISTICS

2023-24

CERTIFICATE

This is to Certify that,

Sr. No.	Name	Roll No.
1	Ms. Kamble Pradnya Bharat	1409
2	Ms. Koli Lalita Ramdas	1410
3	Mr. Patil Ketan Kiran	1417

Have satisfactorily completed the project work on “**Statistical Analysis of Caesarean Section For Pregnant Women**” as a part of practical evaluation course for **M.Sc. II**, prescribed by the Department of Statistics, *Vivekanand college, Kolhapur (Empowered autonomous)* in the academic year **2023-24**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Date:

Place: Kolhapur

*for
Pradnya*

Project Guide

(Ms. Bhosale. A. K.)

*Ashok
29/05/24*
Examiner

V. V. Pawar
Head

(Ms. Pawar V. V.)

HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(EMPOWERED AUTONOMOUS)

Acknowledgement

I am privileged to express my sincere thanks with great respect and gratitude to Ms. Pawar V.V. (Head, Department of Statistics), Ms. Bhosale A.K. and all staff members for their aspiring guidance. They all helped with kind of co-operation and constant encouragement. I am grateful to thank them for providing me with all necessary facilities.

Also, I would like to thank all the non-teaching staff of the department for their help and co-operation. I thank all my friends and the teaching staff for their co-operation and help which I received from them during the work throughout. I am indebted to my parents for their encouragement and patience throughout my study and for the trust they have on me

Yours Sincerely,

M.Sc.II

Department of Statistics

INDEX

Sr. No.	Content	Page No.
1	Introduction	5
2	Objectives	6
3	Data Collection	7
4	Methodology	8
5	Graphical Representation	10
6	Classification	12
7	Conclusion	19
9	Reference	20
10	Limitation and Scope	21

INTRODUCTION

Caesarean section (c-section) delivery a surgical procedure involving the incision of the abdomen and uterus to deliver a body has witnesses a substantial increase in rates globally, raising pertinent questions about its prevalence, implications and associated factors. This introduction delves into the statistical landscape of c-section deliveries, aiming to provide a comprehensive overview of this evolving trend childbirth practices.

Several factors contribute to the escalating prevalence of c-section deliveries, including medical and cultural influences. Medically, the procedure is often deemed necessary in cases of maternal – fetal complications, such as breech presentation, fetal distress, or placental abnormalities. However, the expanding scope of indications has extended beyond medical necessity to encompass elective & planned c-section. Maternal request, fear of labor complication, and the convenience of scheduling deliveries contribute to the increased frequency of elective c-section.

Caesarean section is one of the most commonly performed surgical operations in the world. Postpartum infection morbidity, one of the most common complications after caesarean section, not only poses a serious problem to maternal physiological and psychological wellbeing but also imposes significant burden to the national economy.

Objectives

- 1) To see caesarean delivery is suitable according to antenatal & postnatal condition.
- 2) To determine which reasons are most responsible for caesarean Section delivery.
- 3) To identify the trends and factors associated with caesarean section.
- 4) To check the classification of caesarean section in women's using various machine learning algorithm.

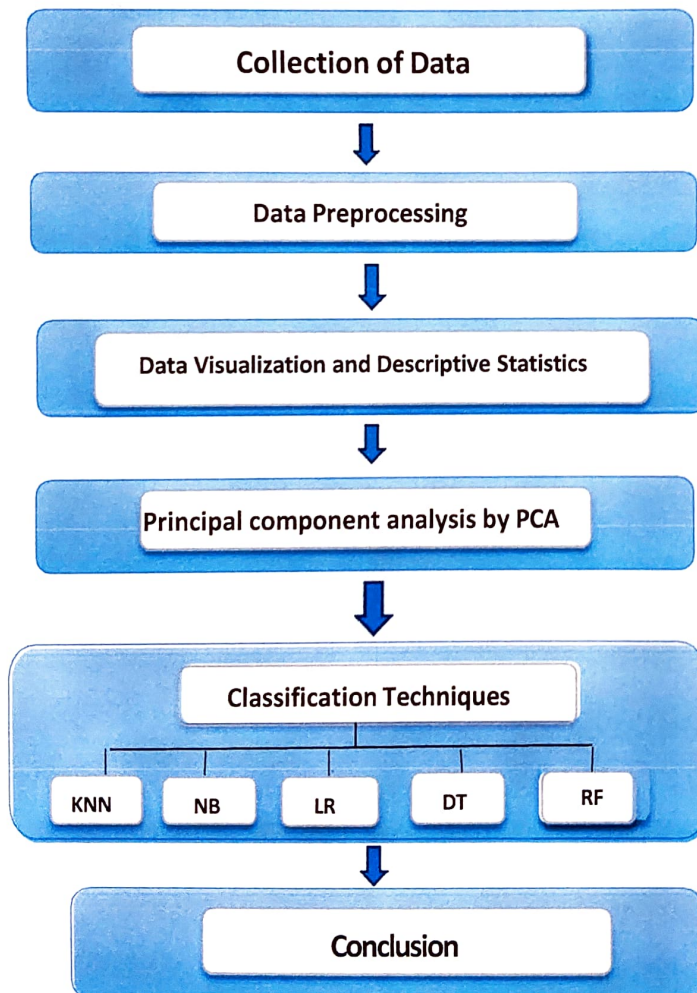
Data collection

For this project, we have collected primary data

- Target population: pregnant women in age group 18-40.
- Data consists total 370 number of sample observations which are 283 for c-section and 87 for normal delivery.
- Among 283 with c-section surgery/delivery.
- For conducting dataset, we prepared well-structured questionnaire which consists 26 numbers of questions related to pregnant women

✚ Methodology

In this study, machine learning classification technique has been proposed, trained and tested aims to differentiate between C-section delivery and normal delivery. The variables have been used as the input of the study from which the suggested method would determine whether or not the women's having C-section delivery or Normal delivery. Along with this we performed comparative analysis to identify which classification technique is well suited for our data. The framework of the methodology used in this research is illustrated below:



Statistical tools

❖ **Exploratory Data Analysis: -**

- Bar charts, Pie Charts
- Principle Component Analysis
- Classification Report.

❖ **Machine Learning Algorithms (Data Mining Classifiers):-**

Random Forest, K-Nearest Neighborhood, Logistic Regression, Decision Tree, Naïve bayes classifier.

❖ **Statistical Software: -**

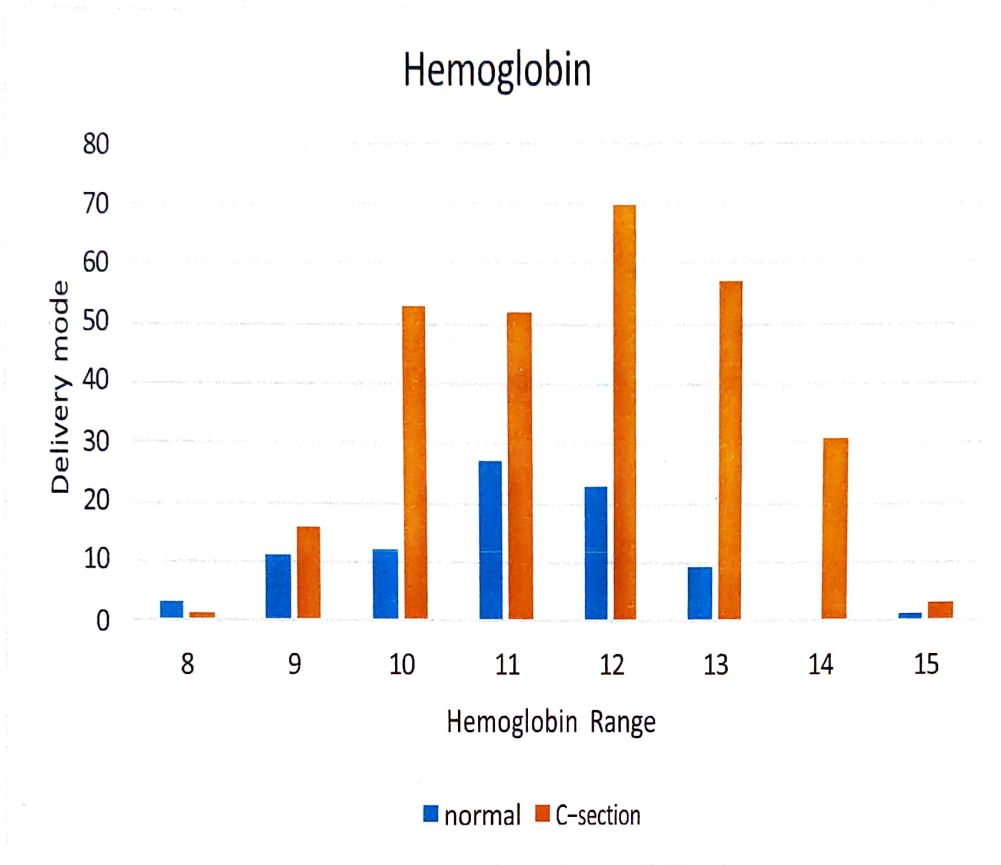
MS-Excel



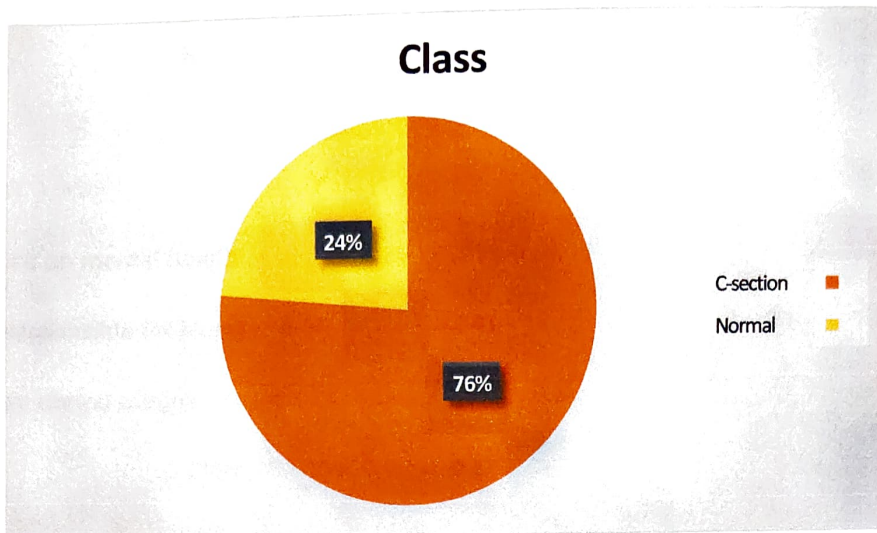
Python



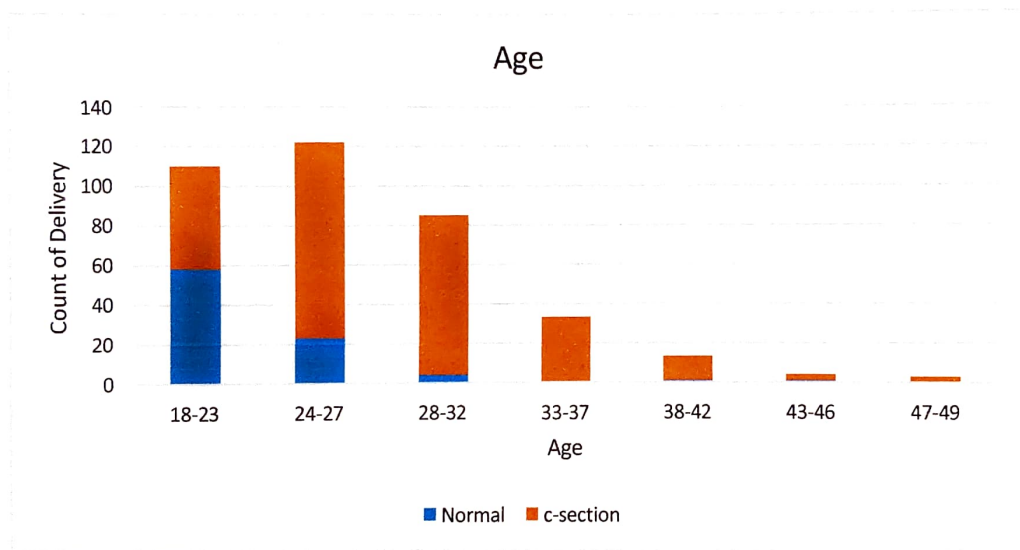
Graphical representation



Conclusion: Most of the women's having Low Hemoglobin range in normal- section.

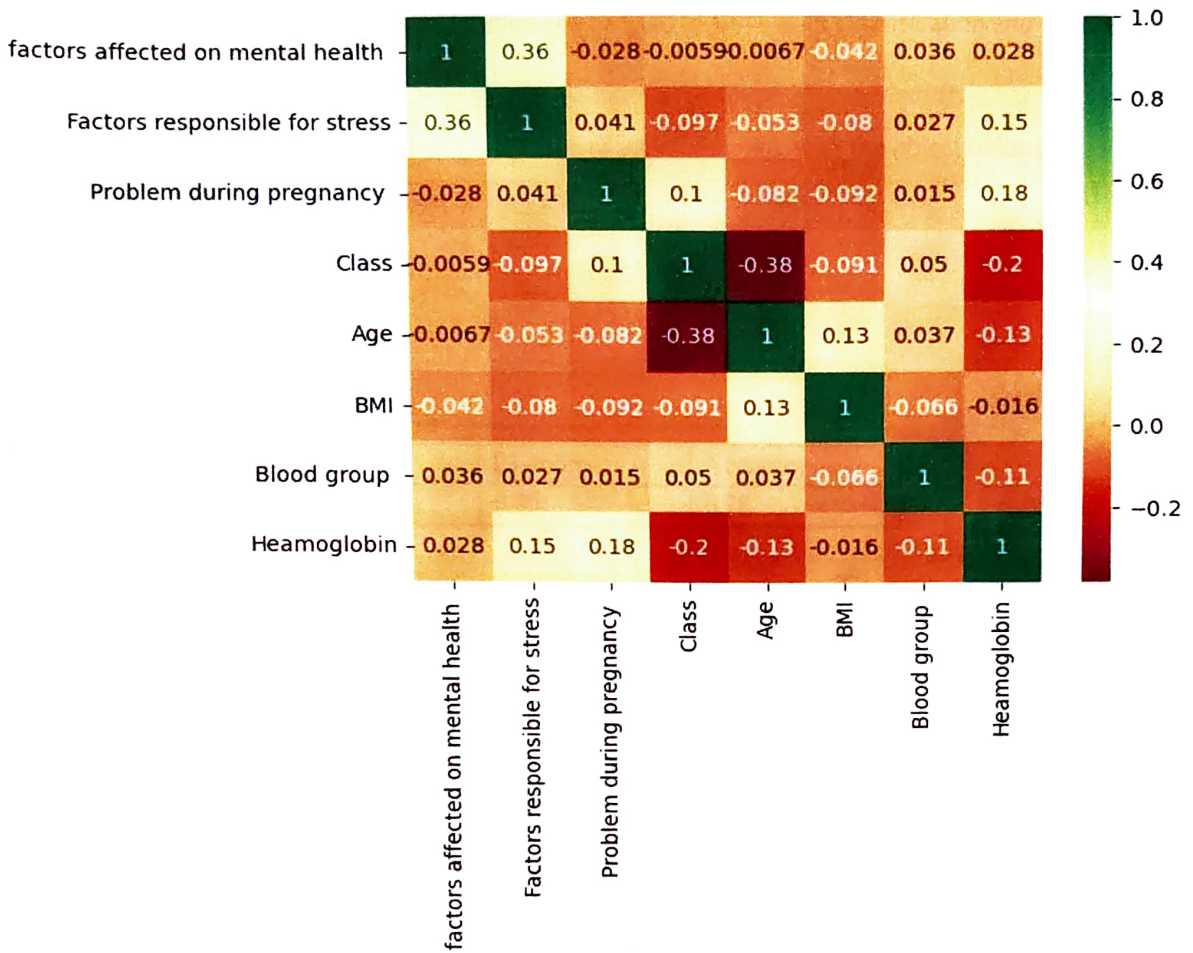


Conclusion: This graph shows 76% of C-section Rate and 24% of normal Rate.



Conclusion: Above graph shows the age group 24-27 and 28-32 high C-Section rate and age group 18-23 high normal rate.

✚ Correlation Heatmap



Conclusion: There is a relation between factors affected on mental health and factor responsible for stress.

Classification

K-Nearest Neighbourhood (KNN) Classifier

The k-nearest neighbours (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common among its k nearest neighbours.

Classification Report:

Precision	Recall	F1-score	Accuracy
0.82	0.78	0.80	0.78

Confusion matrix :

Predicted/Actual	C-Section	Normal
C-section	49	11
Normal	5	9

Interpretation: 78% predicted values are correctly classified with 22% misclassification rate by the KNN classifier.

Naive Bayes Classifier

Naïve Bayes Algorithm is a classification technique based on Bayes Theorem. Naïve Bayes is easy to build and very much useful for large datasets. By using the Naïve Bayes equation, we can find the future probability. The Equation is as follows:

$$P(x|c) = P(c|x).P(c) / P(x)$$

Where, $P(c|x)$ is future probability of class (c, target), $P(c)$ is the prior probability of the class. $P(x|c)$ is the likelihood which is the probability of predictor given class, $P(x)$ is the prior probability of predictor.

Classification Report:

Precision	Recall	F1-score	Accuracy
0.77	0.69	0.72	0.68

Confusion matrix :

Predicted/Actual	C-Section	Normal
C-section	43	17
Normal	6	8

Interpretation: 82% predicted values are correctly classified with 18% misclassification rate by the naïve bayes classifier.

Decision Tree

Decision tree is a Supervised Machine Learning Technique since, both values of independent and target variable Decision tree is very flexible used for both the classification and regression techniques. It can be used when the dependent variable is not normally distributed. The goal of using a Decision Tree is to create a training model that can use to predict the class or value of the target variable by learning simple decision rules inferred from prior data.

In Decision Trees, for predicting a class label for a record starting from a root node. Then comparing the values of the root attribute with the record's attribute. On the basis of comparison, following the branch corresponding to that value and jump to the next node.

The decision tree is a structure read from top to bottom. The topmost node is the root node. Each subsequent node in the tree is either a decision node where data is split into subsets based on some conditions or a leaf node where a data is not split up further.

Classification Report:

Precision	Recall	F1-score	Accuracy
0.85	0.86	0.84	0.83

Confusion matrix:

Predicted/Actual	C-Section	Normal
C-section	52	8
Normal	4	10

Interpretation: 83% predicted values are correctly classified with 17% misclassification rate by the Decision Tree classifier.

Random Forest

Random forest is an ensemble technique used for both the classification and regression problems. It is collection of decision trees and uses a decision tree algorithm in randomized pattern. During the classification each one tree holds a class and most popular class is assigned to newly specified tuples.

It corrects the overfitting habit of decision tree, in particular trees that are grown very deep tends to learn highly irregular patterns, they overfit their training sets (low bias but very high variance). Random forests are a way of averaging the multiple deep decision trees, with the goal reducing the variance. Not quite similar, forests give the effects of a k-fold cross.

Classification result:

Precision	Recall	F1-score	Accuracy
0.83	0.82	0.83	0.82

Confusion matrix :

Predicted/Actual	C-Section	Normal
C-section	58	2
Normal	11	3

Interpretation: 82% predicted values are correctly classified with 18% misclassification rate by the Random forest classifier.

Logistic Regression

Logistic Regression is a Supervised Machine Learning Classification Algorithm used to predict the target variable. As the target variable is binary, which means there are two possible classes. Linear regression technique cannot be used so Logistic Regression is preferred as it deals with dichotomous target variable.

Classification report:

Precision	Recall	F1-score	Accuracy
0.84	0.82	0.83	0.82

Confusion matrix :

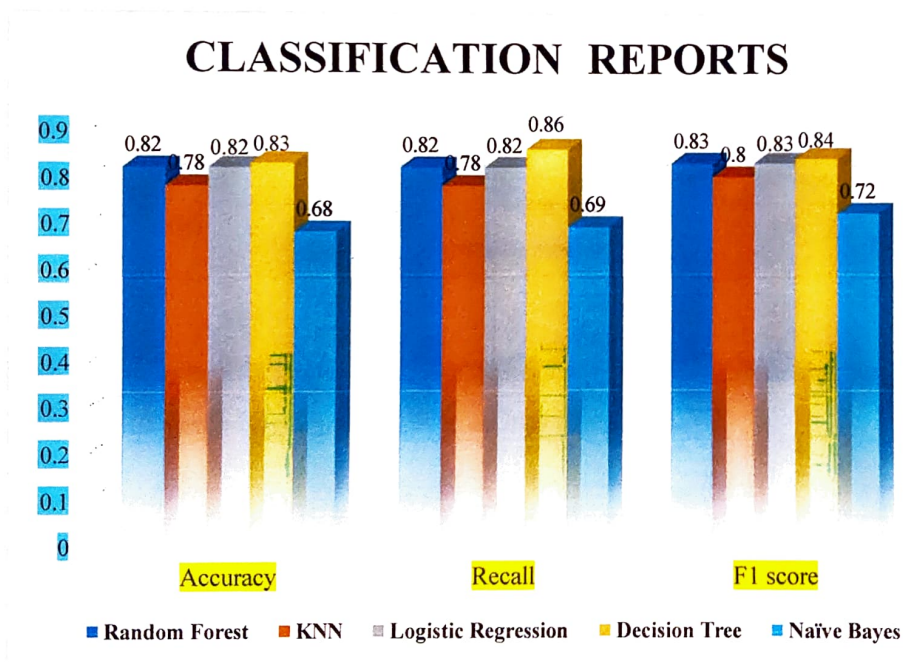
Predicted/Actual	C-Section	Normal
C-section	59	1
Normal	8	6

Interpretation: 82% predicted values are correctly classified with 18% misclassification rate by the Logistic Regression classifier.

Classification Report

To classify women's with C-section delivery and Normal delivery from given variables, we applied following techniques.

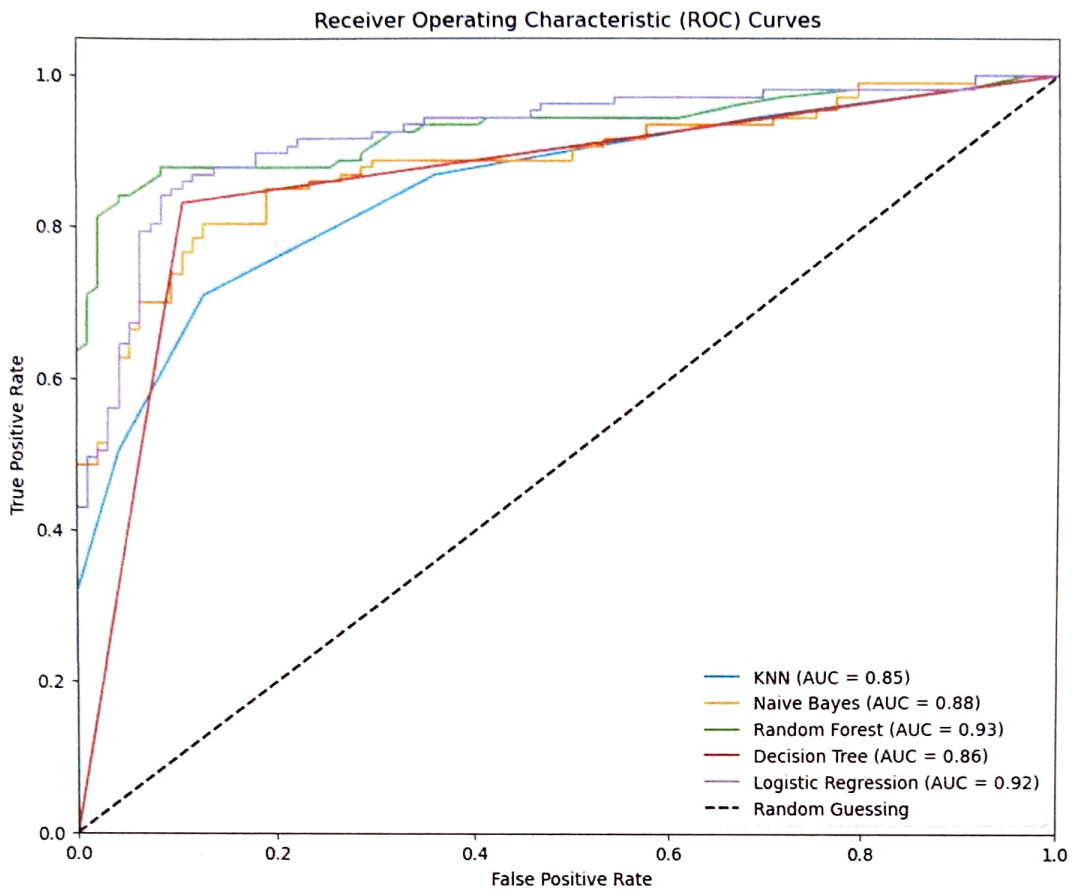
Performance			
Classifiers	Accuracy	Recall	F1 score
Random Forest	0.82	0.82	0.83
KNN	0.78	0.78	0.80
Logistic Regression	0.82	0.82	0.83
Decision Tree	0.83	0.86	0.84
Naïve Bayes	0.68	0.69	0.72



Interpretation : From above figure we conclude that Decision Tree model gives highest (83%) accuracy of classification.

Receiver Operating characteristic (ROC) Curve

ROC curve is commonly used to visualize the performance of the classifier.



Conclusion: From above ROC curve, the AUC is 0.93 which indicates that random forest is best model which correctly classified observation into categories.

Major findings

- ❖ Maximum number of Women's experienced problems in delivery due to Age .
- ❖ Approximately 85% of women need caesarean delivery due to advancing Age.
- ❖ Decision tree model gives highest (83%) accuracy of classification, whereas random forest model gives 82% accuracy, the AUC is 0.93 which indicates that discriminant is fair.

Reference

- J. Cleary-Goldman et al., “Impact of maternal age on obstetric outcome,” *Obstetrics Gynecol.*, vol. 105, no. 5, pp. 983–990, 2005.
- J. F. Ludvigsson and J. Ludvigsson, “Coeliac disease in the father affectsthe newborn,” *Gut*, vol. 49, no. 2, pp. 169–175, 2001.
- R. Robu and Ş. Holban, “The analysis and classification of birth data,” *Acta Polytechn. Hungarica*, vol. 12, no. 4, pp. 1–20, 2015
- T. Leone, “Demand and supply factors affecting the rising overmedicalization of birth in India,” *Int. J. Gynaecol. Obstetrics*, vol. 127, no. 2, pp. 157–162, 2014.
- Sana, S. Razaq, and J. Ferzund, “Automated diagnosis and cause anal-ysis of cesarean section using machine learning techniques,” *Int. J. Mach.Learn. Comput.*, vol. 2, no. 5, p. 677, 2012.

Limitations of the study

- Tools applied to the data can change their performance if we change the data
- We develop model only with available variables but if add other important variables then we expect that our models give better result

QUESTIONNAIRE

❖ For caesarean Delivery

1. Mother's Name: _____
2. Mobile Number: _____
3. Age: _____
4. Height: _____
5. Weight: _____
6. Blood Group: _____
7. Haemoglobin: _____
8. Place Of residence:
 - Urban
 - Rural
9. Are you vegetarian or non-vegetarian?
 - Vegetarian
 - Non-vegetarian
 - Both
10. Which of the following foods do you consume the most?
 - Fast food
 - Balanced diet
 - Oily/Spicy Foods
11. Are you addicted to any kind of addiction?
 - Smoking
 - Wine
 - Masher
 - None
12. What kind of exercise do you do?
 - Walking
 - Yoga
 - No

13. How many hours do you sleep?

- 4-6 hrs
- 6-8hrs
- 8-10hrs

14. What do you do?

- Housewife
- Service
- Wages

15. What factors affect your mental health?

- Family
- Educational
- Relationship
- Financial situation
- None

16. What factors are responsible for stress if you are employed?

- Heavy work stress
- Time management
- Lack of support
- None

17. Was there any problem during pregnancy?

- Blood pressure
- Diabetes
- Thyroid
- None

18. What was the doctor's advice?

- Baby position
- Umbilical cord wrapping
- Mandibular Depression
- Dehydration
- Obesity
- Baby have high heart beats
- Baby have high weight

19. Having a normal delivery but having a caesarean surgery out of fear?

- Yes
- No

20. Has the baby's grandmother had a caesarean surgery?

- Yes
- No

21. How many children do you have in total?

- 1
- 2
- 3
- 4

22. How many children have had caesarean surgery?

- 1
- 2
- 3
- All

23. In what month was the baby born?

24. Have you had any of the following diseases?

- HIV
- Tuberculosis
- Blood pressure
- Sugar
- Thyroid
- None

25. The hospital where the baby was born?

- Government
- Private

26. How much did caesarean surgery cost?

27. What complications are caused by Normal delivery?

- Body ache/Backpain
- swelling
- Abdominal girth increases
- No

❖ **For Normal Delivery**

1. Mother's Name: _____

2. Mobile Number: _____

3. Age: _____

4. Height: _____

5. Weight: _____

6. Blood Group: _____

7. Haemoglobin: _____

8. Place Of residence:

- Urban
- Rural

9. Are you vegetarian or non-vegetarian?

- Vegetarian
- Non-vegetarian
- Both

10. Which of the following foods do you consume the most?

- Fast food
- Balanced diet
- Oily/Spicy Foods

11. Are you addicted to any kind of addiction?

- Smoking
- Wine
- Masher

- None

12. What kind of exercise do you do?

- Walking
- Yoga
- No

13. How many hours do you sleep?

- 4-6 hrs
- 6-8hrs
- 8-10hrs

14. What do you do?

- Housewife
- Service
- Wages

15. What factors affect your mental health?

- Family
- Educational
- Relationship
- Financial situation
- None

16. What factors are responsible for stress if you are employed?

- Heavy work stress
- Time management
- Lack of support
- None

17. Was there any problem during pregnancy?

- Blood pressure
- Diabetes
- Thyroid
- None

18. Having a normal delivery but having a caesarean surgery out of fear?

- Yes
- No

19. Has the baby's grandmother had a caesarean surgery?

- Yes
- No

20. How many children do you have in total?

- 1
- 2
- 3
- 4

21. How many children have had caesarean surgery?

- 1
- 2
- 3
- All

22. In what month was the baby born?

23. Have you had any of the following diseases?

- HIV
- Tuberculosis
- Blood pressure
- Sugar
- Thyroid
- None

24. The hospital where the baby was born?

- Government
- Private

25. How much did caesarean surgery cost?

26. What complications are caused by Normal delivery?