MATERNAL HEALTH RISK PREDICTION USING MACHINE LEARNING

ROHITH NARAYANAN S

Register No. – 210419106092

Under the supervision of

PREMLATHA T

JULY 2022

Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported us throughout the course of this group project. We are thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work. I am sincerely grateful to them for sharing their truthful and illuminating views on a number of issues related to the project.

Further, I am fortunate to have **PREMLATHA T** as my mentor. She has readily shared her immense knowledge in data analytics and guided me in a manner that the outcome resulted in enhancing our data skills.

I wish to thank all the faculties, as this project utilized knowledge gained from every course that formed the PGP in Data Analytics by Imarticus Learning Institute.

We certify that the work done by us for conceptualizing and completing this project is original and authentic.

Date: July 18, 2022 Name: **Rohith Narayanan S**

Place: Chennai

Certificate of Completion

I hereby certify that the project titled **Maternal Health Risk Prediction Using Machine Learning** was undertaken and completed (July 18, 2022)

Mentor: **PREMLATHA T**

Date: 22 July 2022

Place – Chennai

Table of Contents

1. Abstract

Maternal mortality is a major, but often overlooked, public health problem and is unacceptably big nowadays. About 2,95,000 women died during and following pregnancy and childbirth in 2017 alone. Most of these deaths (94%) occurred in low-resource settings, and most could have been prevented with proper care.

Sub-Saharan Africa and Southern Asia accounted for approximately 86% (2,54,000) of the estimated global maternal deaths in 2017. Sub-Saharan Africa alone accounted for roughly two-thirds (1,96,000) of maternal deaths, while Southern Asia accounted for nearly one-fifth (58,000).

At the same time, between 2000 and 2017, Southern Asia achieved the greatest overall reduction in MMR: a decline of almost 60% (from an MMR of 384 down to 157). Despite its very high MMR in 2017, sub-Saharan Africa as a sub-region also achieved a substantial reduction in MMR of nearly 40% since 2000. Additionally, four other sub-regions roughly halved their MMRs during this period: Central Asia, Eastern Asia, Europe and Northern Africa. Overall, the maternal mortality ratio (MMR) in less-developed countries declined by just under 50%

2. Problem Statement

Since maternal risk and mortality is such a devastating problem, what can be done to decrease these numbers and save both maternal and fetal health? We will be solving this question of how to predict maternal health outcomes based on CTG data and prevent the loss of an unborn child.

3. Business Value Proposition

We predict the Level of Risk of Miscarriage faced by All Pregnant Women using Important Attributes of Maternity measured from the women themselves by a CTG test and prevent the loss of an unborn child.

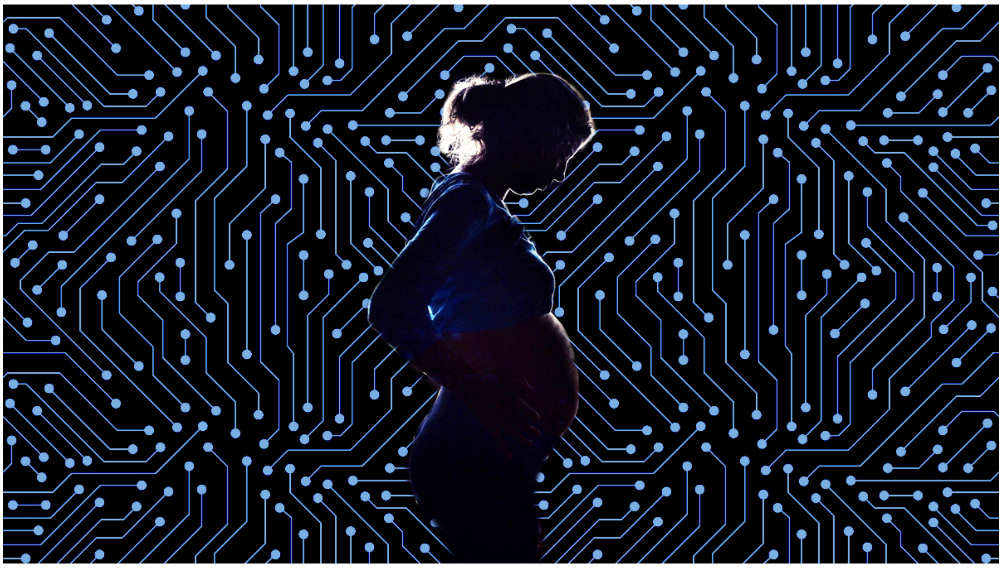
4. Objective

- **Primary Objective:**
  Is to Predict the Level of Risk of Miscarriage faced by Pregnant Women.
- **Secondary Objective:**
  Is to calculate the prediction at an earlier stage to take preventive measures to save the fetus.

5. Related Research

- Interventions to Prevent Perinatal Depression US Preventive Services Task Force Recommendation Statement
  US Preventive Task Force | February 2019
- Gestational Weight Gain and Severe Maternal Morbidity at Delivery Hospitalization
  Obstetrics & Gynecology | March 2019
- Pregnancy Outcomes in US Prisons, 2016–2017
  American Public Health Association | March 2019
- The role of male partner in utilization of maternal health care services in Ethiopia: a community-based couple study
  BMC Pregnancy and Childbirth | January 2019
- Prior dengue virus infection and risk of Zika: A pediatric cohort in Nicaragua
  PLOS Medicine | January 2019
- Association of Maternal Social Relationships With Cognitive Development in Early Childhood
  JAMA Network Open | January 2019
- Introduction of genomics into prenatal diagnostics
  The Lancet | January 2019

6. Significance of Study



The high level of maternal mortality continues to be a major problem in the world one that has only been exacerbated by the Covid-19 pandemic. Researchers and their research suggest that it can be reduced by a strategy consisting of three elements: 1) Using electronic health records and artificial intelligence (AI) and Machine Learning (ML) to predict which pregnant women are at high risk of experiencing complications while giving birth; 2) Employing digital technology to better monitor patients during their pregnancies and improve their access to both routine and high-acuity care during their pregnancies; 3) Following the guidelines of the American College of Obstetricians and Gynecologists and referring them to hospitals that offer higher levels of maternal care.

This project is completed by utilizing the advancements and advantages of Machine Learning Algorithms and an efficient solution is created as a result.

7. Scope of Study

- Due to the limitation of domain-specific data the model's whole responses would be based on the train data.
- It will be able to generate much accuracy or relevant results for the given data that the model hasn't been trained on.
- Sometimes the model could predict the effective result even when the given data has some errors.

8. Data Source

UCI Machine Learning Repository: Cardiotocography Data Set -

It consists of the Cardiotocography (CTG) test results of the fetus of pregnant women.

9. Research Methodology

The flow of the entire research project will be discussed below.

9.1. Introduction

The research/project will be carried out in various stages Basic parts are Dataset loading and preprocessing, Finetuning machine learning models using the given dataset, Hyperparameter optimization, Model evaluation, model consumption using API with UI connection and deployment is also being done here.

9.2. Dataset Description

The dataset consists of the CTG test results of the fetus's heart. It has the following attributes:

- baseline value - Baseline Fetal Heart Rate (FHR) (beats per minute)
- accelerations - Number of accelerations per second
- fetal_movement - Number of fetal movements per second
- uterine_contractions - Number of uterine contractions per second
- light_decelerations - Number of light decelerations per second
- severe_decelerations - Number of severe decelerations per second
- prolongued_decelerations - Number of prolonged decelerations per second
- abnormal_short_term_variability - Percentage of time with abnormal short-term variability
- mean_value_of_short_term_variability - Mean value of short-term variability
- percentage_of_time_with_abnormal_long_term_variability - Percentage of time with abnormal long-term variability
- mean_value_of_long_term_variability - Mean value of long-term variability
- histogram_width - Width of FHR histogram (generated from test)
- histogram_min - Minimum of FHR histogram (generated from the test)
- histogram_max - Maximum of FHR histogram (generated from the test)
- histogram_number_of_peaks - Number of FHR histogram peaks (generated from the test)
- histogram_number_of_zeroes - Number of FHR histogram zeroes (generated from the test)
- histogram_mode - Mode of FHR histogram (generated from the test)

- histogram_mean - Mean of FHR histogram (generated from the test)
- histogram_median - Median of FHR histogram (generated from the test)
- histogram_variance - Variance of FHR histogram (generated from the test)
- histogram_tendency - Tendency of FHR histogram (generated from the test)
- fetal_health - Fetal health as assessed by an expert obstetrician. 1 - Normal, 2 - Suspect, 3 - Pathological

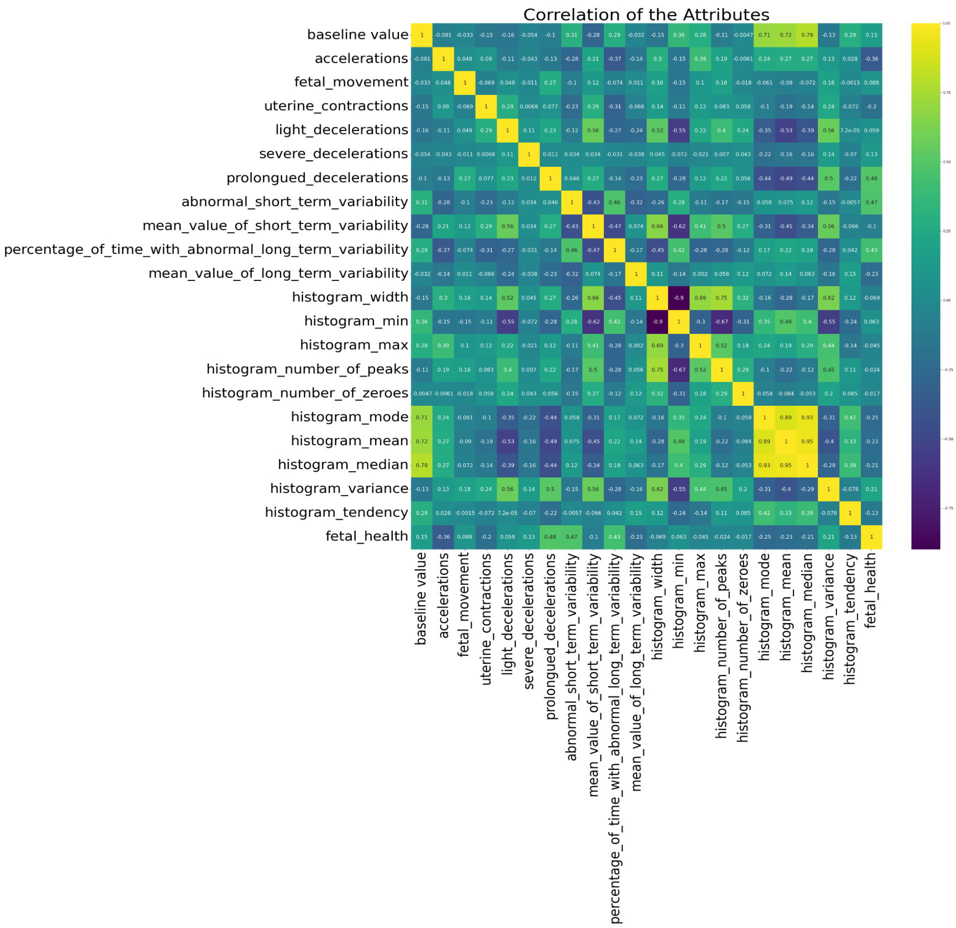
9.3. Data Pre-processing

We will be cleaning the above dataset, checking and treating various aspects before the data is used for modelling.

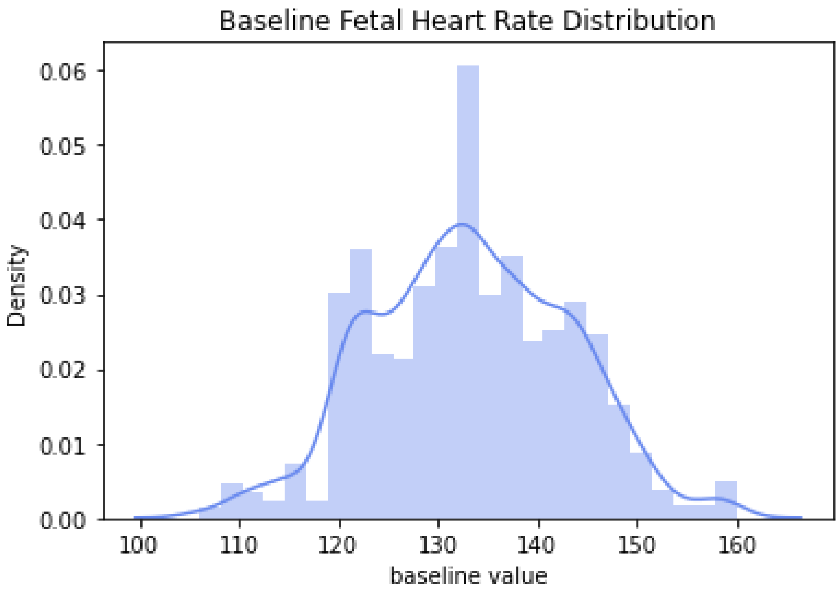
9.4. Exploratory Data Analysis

On doing Some **Exploratory Data Analysis** we can see that:

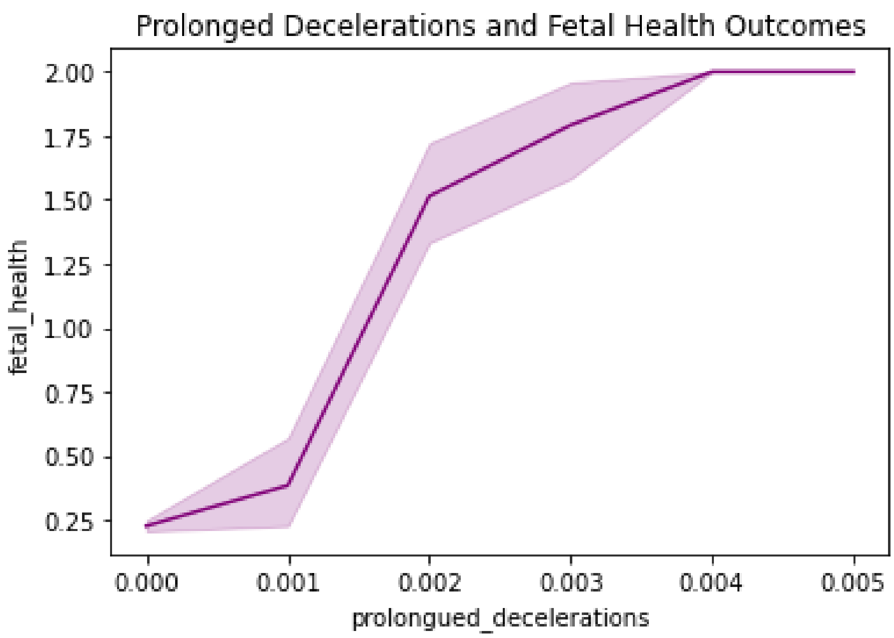
- **Data correlation using the heatmap**



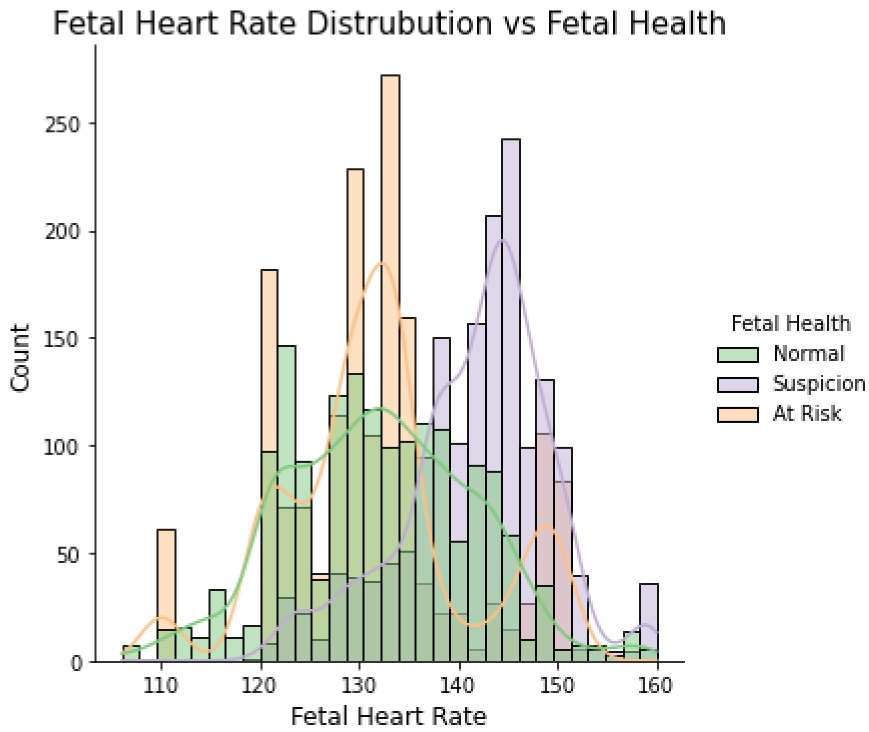
- **The Baseline Fetal Heart Rate Distribution**



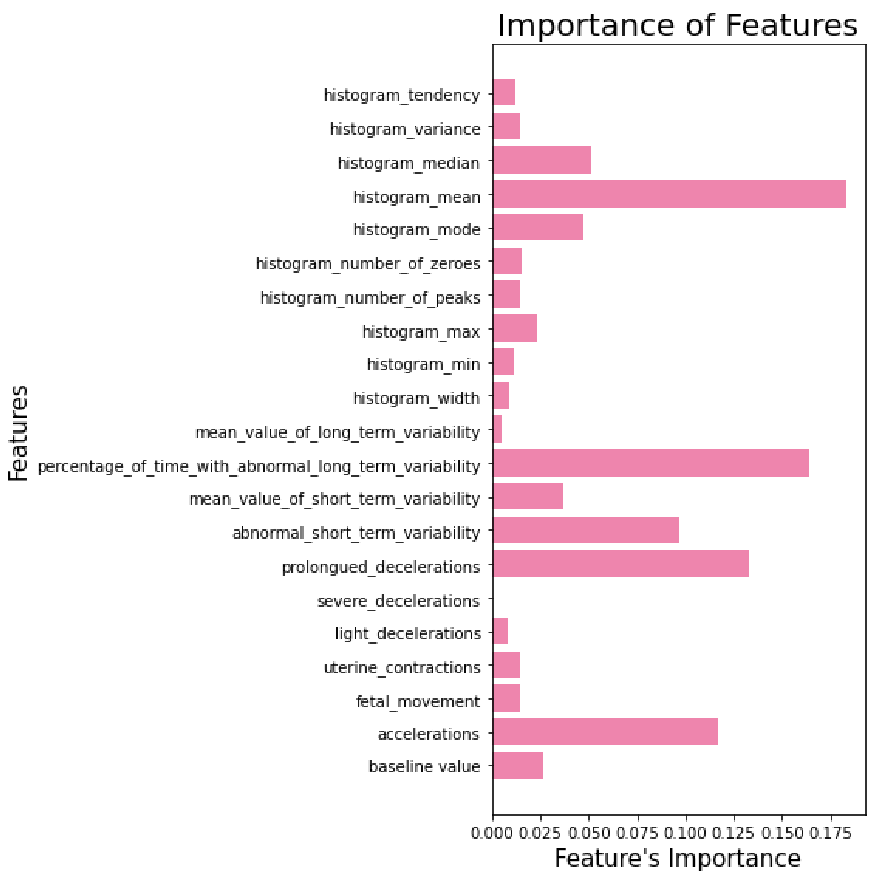
- **Relationship between Prolonged Decelerations & Fetal Health Outcomes**



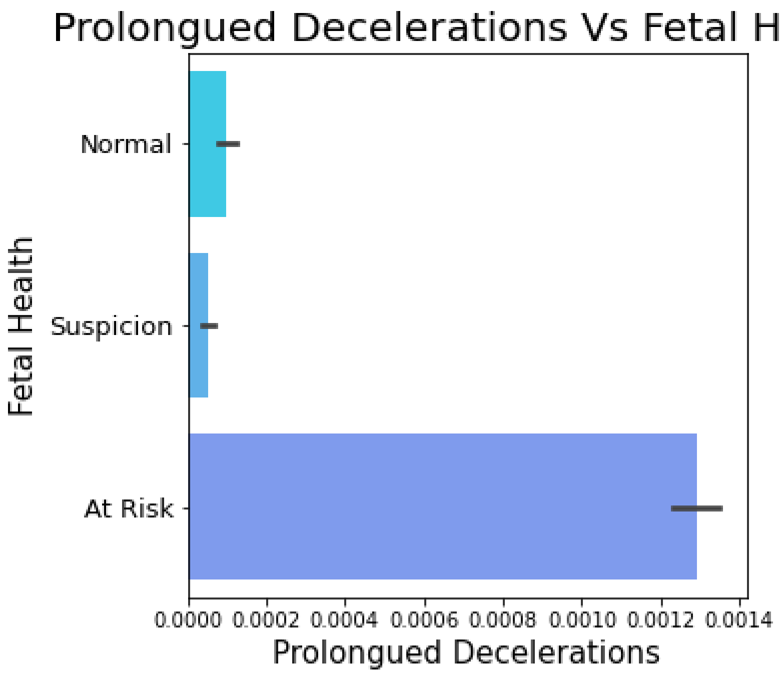
- **The relationship between Fetal Heart Rate Distribution & Fetal Health**



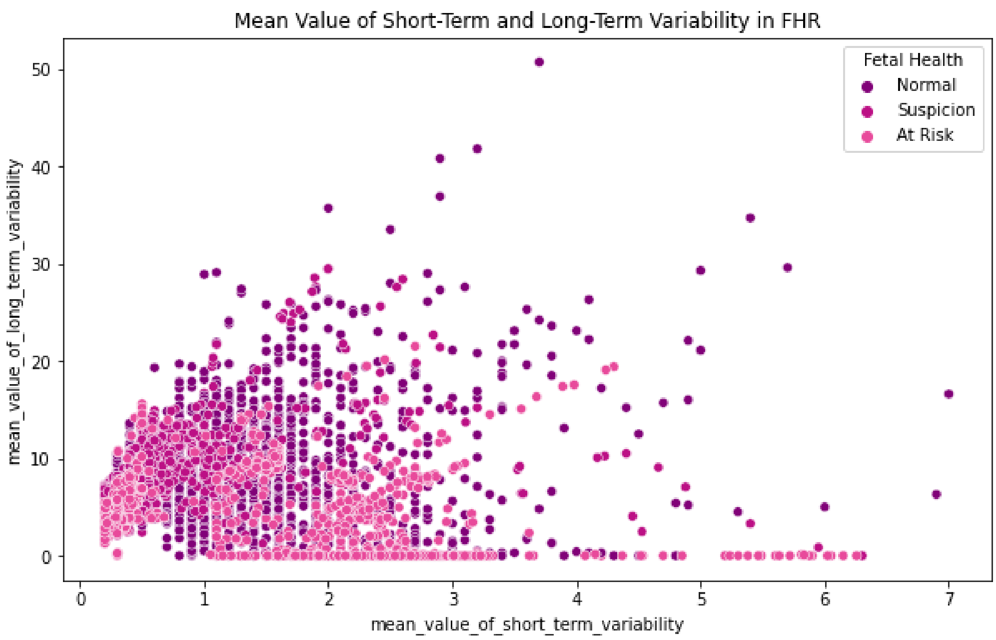
- **Importance of Features**



- **The relationship between Prolongued Decelerations & Fetal Health**



- **The Relationship between Mean Value of Short-Term and Long-Term Variability in FHR**



- **And many more data understanding were done**

9.5. Model Building

Here we will be using several models to calculate and predict the best model.

Here we use:

- **Decision Tree Classifier Model:**

  A decision tree classifier is a useful algorithm for classifying data. The goal of a classification tree is to separate data into discrete classes. (For example, whether a select group of people have heart disease or not.) In a classification tree, a categorical question is used to separate data and based on a true/false or yes/no answer, the data are classified into two groups.

  Here we got an accuracy of 0.9164149043303121.

- **Random Forest Classifier Model:**

  The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees. The Random forest classifier creates a set of decision trees from a randomly selected subset of the training set.

  Here we got an accuracy of 0.9828801611278952.

.

- **Logistic Regressor Model**

  The logistic Regressor Model is a statistical model (also known as the *logit model*) is often used for classification and predictive analytics. Logistic regression estimates the probability of an event occurring, such as voted or didn't vote, based on a given dataset of independent variables. Since the outcome is a probability, the dependent variable is bounded between 0 and 1.

  Here we got an accuracy of 0.7915407854984894.

- **XGBoost Classifier Model**

  XGBoost Model for Classification XGBoost is short for Extreme Gradient Boosting and is an efficient implementation of the stochastic gradient boosting machine learning algorithm. The stochastic gradient boosting algorithm, also called gradient boosting machines or tree boosting, is a powerful machine learning technique that performs well or even best on a wide range of challenging machine learning problems.

  Here we got an accuracy of 0.9848942598187311.

- **Gaussian Naive Bayes Model**

  Gaussian Naive Bayes Naïve Bayes is a probabilistic machine learning algorithm used for many classification functions and is based on the Bayes theorem. Gaussian Naïve Bayes is the extension of naïve Bayes.

  Here we got an accuracy of 0.770392749244713.

9.6. Model Evaluation

Model evaluation is will be done by considering the following parameters:

- Precision: Of the predictions, the model made for this class, what proportions were correct.
- Recall: Out of all of the instances of this class in the test dataset, how many did the model identify.
- F1-Score: An average metric that takes both precision and recall into account.
- Support: How many instances of this class are there in the test dataset.
- ROC (Receiver Operating Characteristic): curve will be more effective, as it does not need such a threshold.
- AUC (Area Under ROC Curve): is also shown in the above graph, which is simply used as a single value to be compared between different models for evaluating their performance.
- Confusion Matrix: It is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known.

9.7. Model Selection

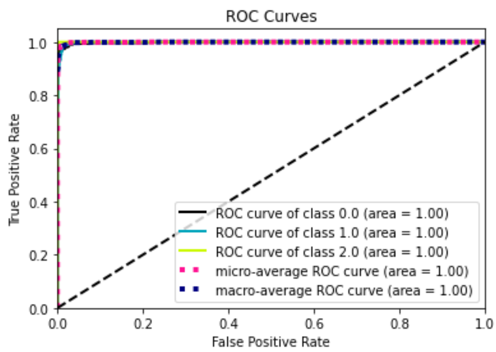
Since models, Random Forest Model & XGBoost Classifier Model have close to similar scores and are the highest performing models, we will evaluate them and select the best.

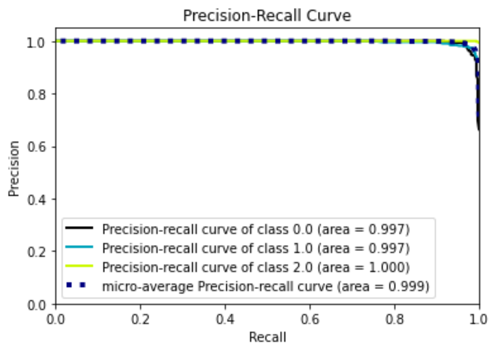
- **Random Forest Model**

Random Forest Model Score = 98.3%

Classification Report:

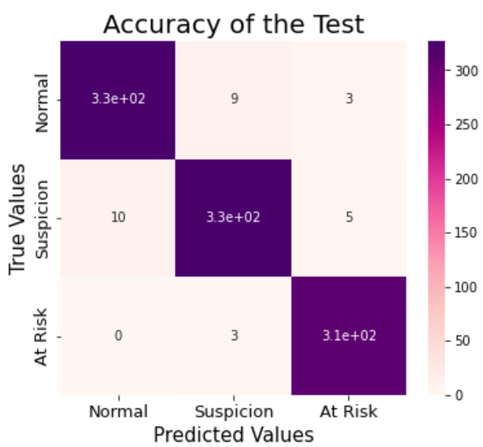
	precision	recall	f1-score	support
0.0	0.98	0.97	0.98	320
1.0	0.97	0.98	0.98	338
2.0	0.99	1.00	1.00	335
accuracy			0.98	993
macro avg	0.98	0.98	0.98	993
weighted avg	0.98	0.98	0.98	993





Final Accuracy Score: 0.9828801611278952

Confusion Matrix:

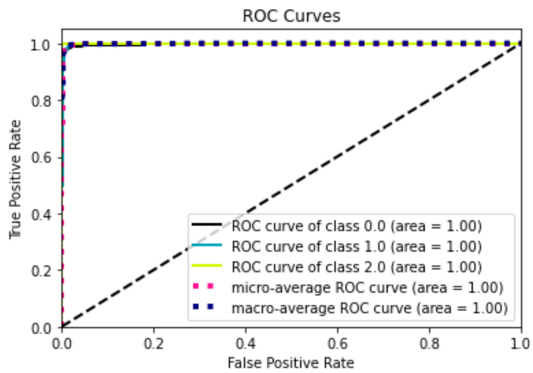


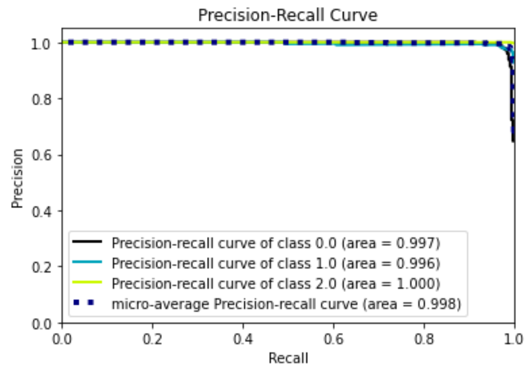
- **XGBoost Classifier Model**

XGBoost Classifier Model Score = 98.5%

Classification Report:

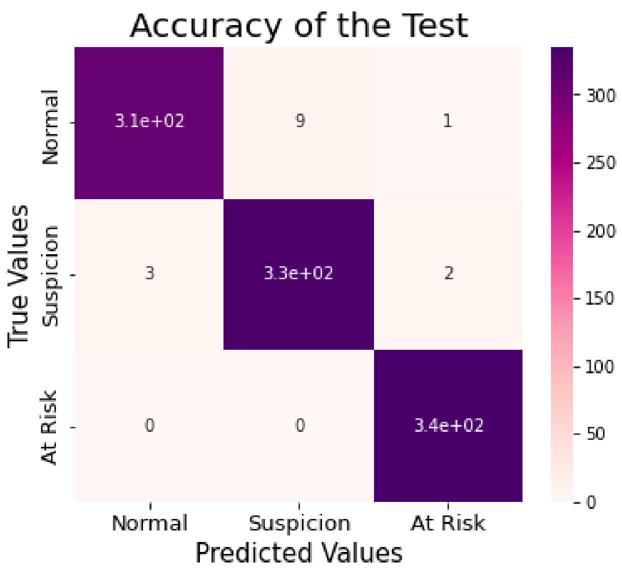
	precision	recall	f1-score	support
0.0	0.99	0.97	0.98	320
1.0	0.97	0.99	0.98	338
2.0	0.99	1.00	1.00	335
accuracy			0.98	993
macro avg	0.99	0.98	0.98	993
weighted avg	0.98	0.98	0.98	993





Final Accuracy Score: 0.9848942598187311

Confusion Matrix:



Upon considering many relevant parameters we decide to go with the **XGBoost Classifier Model** since it better performs in the **confusion matrix**, thereby predicting efficient results when comes to the **critical stage of the Fetus**.

9.8. Model Deployment

Once the best model has been selected we will be using the PICKLE library to store and create a website or an app to deploy it for field testing purposes for final deployment for commercial use.

10. Required Resources

The research will need the below hardware and software resources throughout the implementation.

10.1. Software Requirements

- Package Manager: Anaconda Navigator 1.9.12
- Presentation Layer: Jupyter lab 0.35.4
- Language: Python 3.6.X
- Tools: CudaToolkit 10.2
- Python Libraries for machine learning: Pandas and NumPy for data processing;, scikitplot, xgboost, and sk-learn for model building and evaluating; pickle for model saving and other necessary libraries.

10.2. Hardware Requirements

A laptop with the below configuration will be needed.

- Operating System. Windows 10: 64-bit
- Processor: Intel® Core™ i5 8th gen and above
- Memory(RAM): 4 GB
- Basic GPU is enough

11. References

- UCI Machine Learning Repository: Cardiotocography Data Set
- How AI Could Help Doctors Reduce Maternal Mortality (hbr.org)
- Introduction of genomics into prenatal diagnostics
- Gynaecological Surgery and Machine Learning: Complications and Length of Stay Prediction.
- Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study.
- Using Automated Machine Learning to Predict the Mortality of Patients With COVID-19: Prediction Model Development Study - PMC (nih.gov)
- (PDF) Early Prediction of Severe Maternal Morbidity Using Machine Learning Techniques (researchgate.net)
- Development of a Maternal Early Warning System Using Machine Learning - Duke Institute for Health Innovation (dihi.org)
- Machine learning from fetal flow waveforms to predict adverse perinatal outcomes: a study protocol - PMC (nih.gov)
- Journal of Medical Internet Research - Prediction of Maternal Hemorrhage Using Machine Learning: Retrospective Cohort Study (jmir.org)