

**DEVELOPMENT OF A PREDICTIVE SYSTEM FOR PREDICTING  
PREGNANCY COMPLICATIONS IN WOMEN USING MACHINE LEARNING  
ALGORITHMS**

**By**

**AYODELE OLUWATOBI OLUWANIFEMI**

**18010301063**

**A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE  
AND MATHEMATICS, COLLEGE OF BASIC AND APPLIED SCIENCES, IN  
PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF  
DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE.**

**2022**

**DECLARATION**

I hereby declare that this project has been written by me and is a record of my own research work. It has not been presented in any previous application for a higher degree of this or any other University. All citations and sources of information are clearly acknowledged by means of reference.

---

**AYODELE, OLUWATOBI OLUWANIIFEMI**

---

**DATE**

## CERTIFICATION

This is to certify that the content of this project entitled '**DEVELOPMENT OF A PREDICTIVE SYSTEM FOR PREDICTING PREGNANCY COMPLICATIONS IN WOMEN USING MACHINE LEARNING ALGORITHMS**' was prepared and submitted by **AYODELE OLUWATOBI OLUWANIFEMI** with matriculation number 18010301063 in partial fulfillment of the requirements for the degree of **BACHELOR OF SCIENCE IN COMPUTER SCIENCE**. The original research work was carried out by him under by supervision and is hereby accepted.

\_\_\_\_\_ (Signature and Date)

Dr. (Mrs.) Chinwe P. Igiri

(Supervisor)

\_\_\_\_\_ (Signature and Date)

Dr. M.O. Adewole

Coordinator, Department of Computer Science and Mathematics.

## **DEDICATION**

I would like to dedicate this project to God Almighty, for being faithful and merciful, for seeing me through to the end of this project. I also dedicate this work to my father, Dr. Kolawole Ayodele and my mother, Mrs. Nike Ayodele for being a major source of support in every way.

## **ACKNOWLEDGEMENTS**

I owe my profound gratitude to God Almighty for the strength, wisdom, divine help, and provision he granted me throughout this work. Also, to my parents for their undying support, motivation, prayers, and unsolicited support throughout this work. I would also love to appreciate my Father Dr. kolawole M. Ayodele for his immense impact on this work.

I express gratitude to my supervisor, Dr. (Mrs.) Chinwe P. Igiri, for her teachings, guidance, counsel, and constant support in ensuring the successful completion of this research. God bless you Ma. My heartfelt gratitude goes to the Acting Dean, College of Basic and Applied Sciences – Dr. Ofudje, the Head of Department of the Department of Computer Science and Mathematics – Dr. M.O. Adewole, and all other members of staff of the Department of Computer Science and Mathematics. I also extend a heartfelt appreciation to all my friends for motivating me when needed and pushing me.

## **ABSTRACT**

This study is aimed at the application of predicting pregnancy complication with the use of different machine algorithms. The study achieved this by eliciting knowledge on the variables, collecting relevant data, formulating the classification model, simulating the model. The use of different articles/studies were used as a guiding tools in the means of gathering the necessary data and features needed.

The result of the study showed that out 11 variables, 7 variables were associated with the classification of predicting pregnancy complications. The result also identified a number of variables which had missing values from the data collected. The simulation showed that the higher the proportion of dataset used for training then the higher performance of the model on the testing dataset. So using 90% of the dataset for training and 10% for dataset for testing showed the highest performance.

## TABLE OF CONTENTS

TITLE PAGE	i
DECLARATION	1
CERTIFICATION	3
DEDICATION	4
ACKNOWLEDGEMENT	5
ABSTRACT	vi
TABLE OF CONTENTS	vii
LIST OF FIGURES	x
LIST OF TABLE	xi
<b>CHAPTER ONE: INTRODUCTION</b>	
1.1 BACKGROUND TO THE STUDY	1
1.2 STATEMENT OF THE PROBLEM	4
1.3 AIM AND OBJECTIVES OF THE STUDY	
1.4 PROPOSED METHODOLOGY	6
1.5 SCOPE AND LIMITATIONS OF THE STUDY	6
1.6 SIGNIFICANCE OF THE STUDY	7

## **CHAPTER TWO**

2.1 INTRODUCTION	9
<b>2.2 COMPLICATIONS PREGNANCY WOMEN CAN HAVE</b>	10
2.2.1 HIGH BLOOD PRESSURE IN PREGNANCY	10
2.2.2 PREECLAMPSIA IN PREGNANCY	11
2.2.3 ANEMIA IN PREGNANCY	11
2.2.4 INFECTIONS IN PREGNANCY	12
2.3.1 PREDICTIVE MODELING	12
2.3.2 TYPES OF PREDICTIVE MODELING	13
2.4 THE USE OF MACHINE LEARNING ALGORITHMS	14
2.4.1 FUNCTIONS OF MACHINE LEARNING ALGORITHMS	15
2.4.2 TYPES OF MACHINE LEARNING ALGORITHMS	15
2.5 RELATED WORKS	17
2.6 SUMMARY OF RELATED WORK	21

## **CHAPTER THREE**

3.1 INTRODUCTION	23
3.2 SEARCH STRATEGY	23
<b>3.2.1 ANALYSIS OF THE PROPOSED PREDICTED SYSTEM</b>	24
3.2.2 PREPARATION OF ESTABLISHING PREGNANCY PREDICTION MODEL	24
3.2.3 DATA SET SELECTION/ COLLECTION	25



3.2.4 CLEANING OF DATA	25
3.2.5 TRAINING OF DATA	26
3.3 PREGNANCY COMPLICATION STUDIED	27
3.4 MACHINE LEARNING ALGORITHM APPLIED AND SOFTWARE	31
<b>CHAPTER FOUR</b>	
4.1 INTRODUCTION	32
4.2.1 SELECTION OF TASK	32
4.2.2 PREPARING OF TASK	34
4.2.3 SELECTING OF INPUT	35
4.2.4 SELECTING OF MODEL TYPE	36
4.3 RESULT OF DATA PREPROCESSING OF THE DATA COLLECTED	38
4.4 MACHINE LEARNING SIMULATOR	40
<b>CHAPTER FIVE</b>	
SUMMARY, CONCLUSION, AND RECOMMENDATIONS	41
5.1 SUMMARY	42
5.2 CONCLUSION	
5.3 CHALLENGES	42
REFERENCES	43

## LIST OF FIGURES

Figures	Page
4.2.1: The selecting of which task to perform	33
Fig 4.2.2: Preparing of target used	34
Fig 4.2.3: Selecting of input (dataset) used	35
Fig 4.2.4: Selecting of ML model to use	37
Fig 4.3.1: Result of data preprocessing of the data collected	38
Fig 4.3.2: AOC Curve for the machine learning model used	39
Fig 4.4: Decision tree simulator	40

## LIST OF TABLE

Table	Page
2.6 SUMMARY OF RELATED WORK	21
3.3.1: Identified variables for determining high blood pressure pregnancy Complications	28
3.3.2: Identified variables for determining Partum hemorrhage pregnancy Complications	29
3.3.2: Identified variables for determining Gestational diabetes pregnancy Complications	30

# CHAPTER ONE

## INTRODUCTION

### 1.1 BACKGROUND TO THE STUDY

Predictive analytics also similar to predictive modelling is the use of statistics and modelling techniques in making predictions about future outcome and performances. Predictive analytics examines current and historical data patterns to decide if those patterns would happen to come up again. In the current era many industries and companies such as insurance and marketing, make use of predictive analysis to make important decisions. So here is an outline of predictive analytic: So according to Moore's law which says that Every two years, technology should become half as expensive and twice as quick, according to this theory, which has been backed by the explosive growth in computer processing power. The predictive analytics methods, such as regression and machine learning technology were one byproduct of this development. These prediction models have altered our capacity for accurately processing and utilizing massive amounts of data. Thus, it can be presumed that models such as the ones above must have come around a few years after the creation of the first computer in the late 1930s, well even though these advanced techniques Were about sixty years old, the idea of humans analyzing the past to predict future characteristics actually dates back to the 17th century, which started the first of five phases of predictive analytics. The first stage was in 1689, when the Lloyds company employed predictive analytics to underwrite for sea cruises.

In order to assess the risk of these journey and forecast patterns of liability, Lloryd examined data sets from previous journeys. The idea has become standard practice in the insurance industry as a result of Lloryd's continuing usage of predictive models in all aspects of their insurance underwriting . The basic analytics (the Lloyd company) .Then move forward to the Mid 20<sup>th</sup> century, right after the invention of computers, they was a description phase link that linked the Lloyd company idea of what happened with the modern computers, to create the basic models of linear programming and computational modeling. This model have become the drivers of different business functions in most industry in the world. It was during the time predictive index founder, Arnold Daniel fought in world war two and his battalions would end the war with no casualties which made his commanders want to perform a study. Which they sent a psychologist to study the units which provided Daniel with the first exposure to psychometric testing. Which prompted him to create the first predictive index which was then called PI worldwide to be used in predictive analytics in his work place. This is the descriptive phase (what happened?). Then in the mid1970s predictive analysis began to evolve which enabled people to be able to answer the question why something happened instead of just what happened. The analysis looked for root causes and was the backbone of several innovations that would go to improve business functions which include optimization and maximization. This is the diagnostic phase (why did it happen?). Then in the late 20<sup>th</sup> century and the beginning of the 21<sup>st</sup> century saw the beginning of the predictive phase.

The development of mathematical models that utilized weighting techniques to score data sets were programmed into analytical programs. This allowed computers to make suggestions and predict outcomes based on a series of inputs. In Today world this techniques can be seen as predictive index, as applicant behavior and cognitive scores are weighted and ranked in accordance with the requirements of a position. These ranking abilities allowed computers to provide managers with data driven answer to business problems. This is called the predictive phase (what will happen?). So with the continually increasing processing speeds and innovation in programming language, machine learning and artificial intelligence have created a climate that can allow computer to increase accuracy levels in predictive software without human intervention. The predictive index uses machine learning to analyze job titles and provide users with accurate position characteristics based on samples from other real world positions. These machine learning systems allows the predictive index to continually improve and make a better responds to changes in the marketplace. (**predictive success coporation , 2019** ).

According to Geisser (1993) in a Journal titled “Predicting modeling” Another definition of predictive modeling which can be said to be the process in which a model is invented or chosen to try and possibly predict the probability of an outcome to happen. (**Encyclopedia of Bioniformatics and Computational Biology , 2019**) A predictive model is said not to be fixed, it is validated or checked regularly to add the changes in the underlying data. Which simply means it’s not a one and done prediction. Predictive models make assumptions based on what is gotten from what has happened in the past and what is happening now.

So which means that when new data shows changes in what is happening now , they would likely be an impact on what the future outcome would be. They are a lot of benefit of applying predictive modeling such as it helps in the reduction of time needed in doing a particular events, it helps to reduce the effort and cost in forecasting business and organizations outcomes. **(Ali, 2020)** .Predictive analysis is one of the branches of advanced analytics that is used in making predictions about future events or outcomes. It uses different techniques such as data mining, machine learning, Artificial Intelligence to analyze current and past data to make predictions about future outcomes. Predictive analysis model capture the various relationship among factors to assess the risk with a particular set of conditions to assign an answer or a score. Predictive analysis allow organization to take charge and forward looking in anticipating outcome and behaviors based on a hunch or assumptions. It also provides suggested actions that would benefit from the prediction.

## **1.2 STATEMENT OF THE PROBLEM**

In the country Nigeria, they are about close to 200 million people living in the country, which is often said to be one of Africa most populous country. But in the same country about 20% of all global maternal death occur there. In between 2005-2015, it was estimated that about 60,000 maternal death occurred and nothing less than 90,000 maternal near miss cases happened. It was also said that a Nigeria pregnant woman has a 1 in a 22 lifetime risk of dying during pregnancy or childbirth whereby in most developed countries, the lifetime risk is about 1 in 4900. **(World Health Organization , 2019)**

They are different pregnancy complications such as preeclampsia, preterm delivery, gestation diabetes, etc. which can lead to high risk in the delivery of the baby or the life of the mother. Complications of pregnancy are serious health problems that affect women during pregnancy or after childbirth which can put the life of the mother or child or even both at serious risk. It is possible that the woman can develop health problems before they become pregnant or even during the pregnancy stage.

### **1.3 AIM AND OBJECTIVES OF THE STUDY**

The project aim is to be able to create a predictive model that would be able to use the appropriate predictive algorithm to determine/ predict if a woman pregnancy would have any complication during her maternity or if it would be successful.

The objectives of this study include:

- To be able to identify the application and performance of machine learning Algorithms used to identify pregnancy complication.
- To gather data and identify the different symptoms associated with the safety of the pregnancy.
- To use the gathered data to build a model.
- To test the predictions accuracy of the model.



#### **1.4 PROPOSED METHODOLOGY**

The project aims at developing a predictive model that would be able to get some data/information about a pregnant women at an early stage and would be able to use those data to predict if the woman is likely to get any complication along her pregnancy period. In order to achieve the model that would be able to predict pregnancy complications a literature survey, which comprises of researching of articles, Journals, reviewing related works etc. would have to be carried out to be able to find and chose the appropriate maternal features that would be used as data for the predictive model, alongside finding the most suited predicting algorithms that is best appropriate. The technology used for this program would be rapid miner studio, Rapid Miner Studio is a visual design environment for rapidly building complete predictive analytic workflows. It provides a deep library of machine learning algorithms, data preparation and exploration functions, and model validation tools to support all your data science projects and use cases.

#### **1.5 SCOPE OF THE STUDY**

The study covers the identification and collection of required dataset needed for the creation of prediction model in determining pregnancy complications using machine learning algorithms.

## 1.6 SIGNIFICANCE OF THE STUDY

They has been a rapid growth in the use of AI, machine learning algorithms and predictive model in the modern healthcare sectors. The medical industry owes its fast development to digital transformation, which has provided numerous advantages for health care providers, doctors and patients. All this has helped in different ways such as

- (i) Improved diagnostics: In this case some complications have typical symptoms which qualified doctors can easily define them and prescribe cures according to the predefined treatment plan. However, each patients can have individual symptoms/ atypical symptoms that might point at a specific disease but then make diagnostics more complicated. With the presence of the atypical symptoms it makes the specialist rely on individual patient data and family history to determine whether the person needs hospitalization. Laboratory testing and diagnostics procedures like x-rays etc combined with predictive analytics provide value for further diagnosis and treatment plan choice. **(shybeko, 2021)**
- (ii) Personalized health care: Healthcare institutions tends to increase their efficiency by using the opportunities of precision medicine.
- (iii) Predictive modelling enhances patient centered care based on personal health records and aids in the development of the best treatment strategies designed for each patient. Any available data at the organizations allows predicting the effectiveness of procedures, manipulating, and laboratory tests and medications depending on the specifics of the person's anatomy and genes. **(shybeko, 2021)**

This will allow healthcare services to be able to choose the appropriate predictive model to determine and avoid complication of pregnancy in women.

## **1.7 DEFINITION OF TERMS**

1. **Pregnancy:** Pregnancy can be described as the term in a woman during the period which a fetus develops inside a woman's womb or uterus. Pregnancy normally lasts around 40 weeks, or just over 9 months, as measured from the last menstrual period to delivery.
2. **Pregnancy complications:** Complications of pregnancy are health problems that occur during pregnancy.
3. **Artificial intelligence:** Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems.
4. **Machine learning:** Machine learning (ML) is one of the different types of artificial intelligence (AI) that let software applications to become more accurate in predicting various outcomes without being explicitly programmed to do so.
5. **Predictive Model:** Predictive modeling can be defined as a mathematical process that is used to predict future events or outcomes by studying different patterns in a given set of input data.
6. **Literature survey:** This is a thorough summary of previous or earlier studies on a particular subject which is called a literature review. In a literature review, academic books, articles and other sources that are related to the certain field are studied. Which the review should describe and make sense of this prior research.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

#### **2.1 INTRODUCTION**

##### **Predictive analytics in health industry**

Predictive analytics in health care can be useful to help detect early signs of patient deterioration in the ICU and general ward.

In medicine it is mostly about anticipating and reducing risk based on current and past patient data, for example how likely is a brain tumor patients going to be affected by the complication if surgery is performed. What are the chances that a typhoid patient would be admitted to an intensive facility, Predictive analytic drives to alert doctors and medical personals about the possible events and outcomes before they occurs, which helps them to prevent as much as possible curable health issues.

Around the world about 60% of health care executives have seen the benefits of using predictive analytics and has adopted such methods, according to a survey in 2019 done by the society of actuaries, 42% has seen great improved patient satisfaction since using predictive analytics and also have saved costs.

Pregnancy can be described as the time or period during which one or more offspring develop inside a woman's womb. Pregnancy can be also said to be a complex vital period in a woman's life with a possible impact/effect on her physical and psychological health.

**(M.C.Romero-Ternero, 2020)**

Pregnancy occurs when a male sperm meet up with a female egg, it officially starts when a fertilized egg implant in the lining of the uterus. It might be difficult to adapt to the important physiological changes occurring during pregnancy, while seeking both the well-being of herself as well as the well-being of her fetus. There is a need to know what the well-being of a pregnant woman entails in the essential need to learn new things such as the lifestyle changes (nutrition, physical exercises, sleep, work, etc.), along with proper medical care and timely follow-ups. Another factor that can be a negative impact in pregnant woman psychological health is the potential risk of having health problems during pregnancy, especially if there is a high tendency of developing complications. While most pregnancies occur without complications some women develop complications before and after delivery .However some women who get pregnant experience complications that can involve both the health of the baby or the mother's health, they could be diseases or conditions mothers could have during or before pregnancy that could lead to complication, some occurs during delivery.

## **2.2 COMPLICATIONS PREGNANCY WOMEN CAN HAVE:**

Some common complications pregnant women faces doing pregnancy are as follows:

**2.2.5 High blood pressure:** This is also sometimes called hypertension, it occurs when the arteries carrying blood from the heart to the body organs becomes narrow and make pressure to increase in the arteries. It could cause serious health problems for anyone at any time in life.

If this happens in pregnancy, it can make it hard for blood to be able to reach the placenta, which usually provides nutrients and oxygen to the fetus, if this happens it reduces blood flow and slows the growth of the fetus and can place the mother health in danger and also make the mother in a place of greater risk of preterm labor and preeclampsia. High blood pressure normally doesn't cause symptoms. (**American college of obstetricans and Gynecologists. , 2020**)

#### **2.2.6 Preeclampsia in pregnancy:**

This can be described as a serious disorder that can affect all organs of the body of a pregnant woman. It normally start after 20 weeks of pregnancy, which are often in the third trimester. It is said that when it develops before 34 weeks of pregnancy it's called an early onset preeclampsia. preeclampsia can lead to lot of problems, such as conditions that can make a woman body have seizures and strokes, it can also lead to a HELLP syndrome which stands for hemolysis, elevated liver enzymes and low platelets count this damages /destroys red blood cells and interferes with the blood clotting, which may lead to chest pain, abdominal pains and bleeding in the liver. HELLP syndrome can lead to death and also a lifelong health problems gotten from the condition. (**American college of obstetricans and Gynecologists. , 2020**)

#### **2.2.3 Anemia in pregnancy:**

This means when a pregnant woman has a lower than normal number of red blood cells in their body. During pregnancy a woman body produces more blood to help support the growth of the baby.

If the woman isn't getting enough iron or certain nutrients, her body won't be able to produce the amount of red blood cells it needs to make the additional blood for the baby. (Uscher.Jen, 2020) This can cause them to feel more tired and weaker than usual and sometimes have pale skins, which usually calls for the attention of a doctor to treat the cause of the anemia. (cafasso, 2016) Anemia can lead to other more serious complications such as preterm labor.

#### **2.2.4 Infections in pregnancy:**

Infections are serious pregnancy complications, some which are sexually transmitted infections (STI) which can occur anytime during the stages of pregnancy and may lead to complications for the woman. Some infections can be passed from mother to infant during childbirth, when the infant passes through the birth canal, while other infections could infect a fetus during the pregnancy. Some infections can lead to birth defect, miscarriage, etc.

#### **2.3.1 Predictive modeling**

In the current era most organizations and industries put predictive modeling to the use when conducting predictive analysis, which they use in order to make wise business decisions. (Indeed Editorial Team, 2021)

In general cases the models helps those organizations to learn more about the customer needs and bases, account-related security alert issues. It is quite important to involve predictive modeling into your business if you're interested in learning how to ensure the success or improved functionality of your company. So Basically predictive modeling is the use of statistics and modelling techniques in making predictions about future outcome and performances.

### **2.3.2 Types of Predictive Modeling**

Predictive Modeling can be said to have two categories which are parametric and non-parametric. Hence, a model that uses a set of parameters such as discrete number is called parametric, while data that doesn't belong to a specific set of parameters are called non-parametric. They each have a specific use and both use different technique in their application. Some types of predictive model are:

1. Classification model: This type of predictive modeling is one of the most fundamental and often utilized models because it provides a straightforward answer to queries with a yes and no outcome .An extensive analysis of a query is produced by a classification model using historical data. Retail and finance businesses often use this because it quickly gathers and categorizes information to answer questions such as” is this application good”.
2. Forecasting model: In Forecasting models, these are one of the most common model types due to their versatility. These models produce numerical responses by analyzing historical data and estimating information based on that data.



A business such as an online retailer may use forecast modeling to estimate how many orders they may receive over the next week. These models can also successfully manage multiple parameters simultaneously.

3. Clustering model: A clustering model put aside different data into different groups based on similar characteristics. They use the data gotten from each group to determine large-scale outcomes for each cluster. This model functions by using two types of clustering. Which are Hard clustering and soft clustering: Hard clustering works by arranging data that determining whether each point completely belongs to a certain cluster and Soft clustering assigns a probability to each data point instead of separating them into distinct clusters.

## **2.4 The use of Machine Learning Algorithms**

Machine learning is a subfield of artificial intelligence that give computers/systems the ability to learn without explicitly being programmed. Machine learning can be defined as a subset of AI, which can be said as one of the rapidly growing Technical field in science. Machine learning start with data, such data can be; number, photos, text, repair records, time series from sensors etc. The data is then gathered and prepared to be used as trained data or the information the machine learning model will be trained on. To have a better program, it is better to gather a lot of data. We can use different machine learning algorithms/models, when you supply the data and let the computer train itself to find different patterns or to make its predictions. After some time we human can tweak the model, including changing the parameters to help get more accurate results.

Some data is held out from training data to be used as evaluation data, which is used to test how accurate the machine learning model is when it shows the new data. The result is the model that can be used in the future with different sets of data.

#### **2.4.1 FUNCTIONS OF MACHINE LEARNING SYSTEM**

- 1. Descriptive:** This means that the system uses the data gotten to explain what has happened.
- 2. Predictive:** This means that the system uses the data gotten to explain what will happen.
- 3. Prescriptive:** This means that the system uses the data gotten to explain suggestion about what actions to take. **(Brown, 2021)**

#### **2.4.2 Types of machine learning**

They are three main different subcategories of machine learning which are:

- (i) Supervised learning.
- (ii) Unsupervised learning.
- (iii) Reinforcement learning.

- 1. Supervised machine learning:** In this model, data are trained with labelled data sets, which then allows the model to grow more accurate over time. For example an algorithm can be trained with pictures of dogs and other things, all labelled by humans and the machine would learn different ways to identify pictures of dogs on its own. They are two types of supervised learning which are classification and regression.

This is the most commonly used learning model used today. Classification algorithms predicts the class label of new data points depending on how the model is supervised by training data .while regression algorithms identifies correlations between dependent and independent variables .Supervised machine learning is the most commonly used machine learning technique used in medicine and engineering. In this machine learning paradigm, the main goal is to infer a function,  $f$ ;

$$F: X \longrightarrow Y$$

The function,  $f$  is the model inferred by the supervised ML algorithm from a given sample data or trained set  $A_i$  which is composed of pairs of (inputs  $(X_i)$  and output  $(Y_i)$ ) such that  $X_i \in x$  and  $Y_i \in y$ :

$$A_j = ((X_1, Y_1), (X_1, Y_1), \dots, (X_n, Y_n) \in (x \times X \times y)$$

So basically, for regression problems  $x \in \mathbb{R}^d$  (where  $d$  is the dimension or number of features) of the vector,  $x$  )and  $Y \in \mathbb{R}$  for classification problems  $x$  and  $Y_i$  are discrete while for binary classification  $Y_i \in \{-1, +1\}$  . In the statistical learning framework, the first hypothesis is that the training data are independently and identically generated from an unknown but fixed joint probability distribution function  $p(X, Y)$ . The main goal of the learning algorithm is to find a function  $f$  attempting to model the dependency encoded in  $p(X, Y)$  between the input  $X$  and the output  $Y$ .  $H$  will denoted the set of function where the solution,  $f$  is ought such that  $f \in H$  where  $H$  is the set of all possible functions,  $f$ .

2. **Unsupervised machine learning:** The program looks for different patterns in an unlabeled data. Unsupervised machine learning can find patterns or trends that people aren't explicitly looking for. They are two types of unsupervised learning which are clustering and association.
3. **Reinforcement machine learning:** This train machines through trial and error to take the best action by establishing a reward system. This system can train models to play games when it made the right decision.

## 2.5 Related works

There has been a number of contribution of related works on the predictive modeling in the area of pregnancy complications in women along with the various complications women can have during pregnancy and other related knowledge or concern about it. Following a review of a number of related works on the subject matter of the development of predictive modelling for complication using machine learning algorithms.

**(Ayleen Bertini, 2022)** An article published which explained about how a study was carried out by some research scientists that works for Frontiers in bioengineering and biotechnology that focuses on using machine learning to predict complications in pregnancy: A Systematic review. The main objective of the research carried out was to identify the applicability and performance of machine learning methods used to identify pregnancy complications.

It emphasized about how AI helps medical professionals in making decisions, reducing medical errors and also helping in improving the accuracy of the interpretation of various diagnosis which thereby reduce the workload to which they are exposed. In the research they were different types of data used for the health learning models which includes electronic medical records (EMR), medical images, biological markers and biochemical parameters. The type of data collected determines the type of machine learning that would be used in diagnosis. .Due to the large quantity and complex nature of medical information, Machine Learning was found to be recognized as a promising method for supporting diagnosis or predicting clinical outcomes. . In the research they found out that the decision support systems remained a complex black boxes, which means that their internal logic were still hidden from the clinical team who could not fully understand the rationale behind their predictions, then they decided that Interpretability was important before any health-care team can increase their reliance on Machine Learning systems. So they focused on developing both interpretable models and explanatory methods. The machine learning models were validated using the train–test split or the cross-validation schemes. The models were fitted to a training data, a set fitted to a training data set. The pregnancy complications they tried to predict were divided into 16 main prediction categories which were: prematurity, preeclampsia, adverse delivery, size for gestational age, gestational diabetes mellitus, neonatal mortality,

fetal acidemia, fetal hypoxia, placental accreta, pulmonary diseases, cesarean section, placental invasion, congenital anomaly, severe maternal morbidity, spontaneous abortion, and trial of labor after cesarean.

Machine Learning methods that were mostly used in the prediction of perinatal complications were: The random forest, Logistic regression, neural networks and Support vector machine (SVM).

**(Dusseldorf, 2022)** Another article published by some group of researchers which spoke about improving the prediction of pregnancy complications using AI. It described about how one in ten pregnant women experiences different signs of preeclampsia which can be defined as the rising blood pressure and the increased protein in the urine during some point in a woman pregnancy. At the end only a few of the cases leads to having serious complications. In the article it stated how a scientist called Charite Universitätsmedizin Berlin gathered around 1,647 data from different preeclampsia patients and used different machine learning methods to create a functional model that would be able to predict the probability of a serious complication. **(Dusseldorf, 2022)** It also thoroughly explained the concepts of preeclampsia which was formerly called as toxemia. It stated that it has variety of signs and symptoms. It also stated that the blood pressure rises suddenly and quickly and excess protein can be found in the urine, while fluid accumulated in the arms or legs which is known as a condition called edema. Some of the signs and symptoms experienced were dizziness, headache, nausea or upper abdominal pain.

While using dataset such as blood pressure and urine protein measurement and biomarkers such as sFLt and PIGF they wanted to achieve more accuracy in their predictions, then decided that they wanted to explore the use of machine learning algorithms to analyze different tests results in order to get a better idea of who is at risk or not . When they trained the datasets with the algorithms, they compared the prediction accuracy of the AI model with the traditional prediction method based on the clinical parameters, they found out that their algorithms were clearly superior to the traditional predictions based on blood pressure, proteinuria and the sFLt-1/PIGF ratio. In which of notices was that the positive prediction value for example that the complication would occur was double accurate. The Digital Health Accelerator Program of the Berlin Institute of Health at Charité (BIH) stated that it is supporting further product development so that a solution based on the newly created algorithm can be rapidly translated into broad clinical practice, therefore making it widely used by patients and doctors. However, Future studies must first show the validity of the algorithm. "AI of course cannot replace physicians," stated by Verlohren, a physician himself, is convinced. "But when it comes to considering how preeclampsia treatment should proceed, AI certainly provides a better basis for making decisions.

## 2.6 SUMMARY OF RELATED WORK

S/N	Author(s)	Year	Title	Methodology
1	Ayleen Bertini, Rodrigo salas, Steren Chabert, Luis Sobrevia, Fabrian Pardo	January 2020	Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review	This study was carried out by researching the various pregnancy complications and the use of Twenty-one different Machine Learning methods used to predict different 16 perinatal complications. Machine Learning methods that were mostly used in the prediction of perinatal complications were: The random forest, Logistic regression, Neural networks and Support vector machine (SVM), using The four features: Electronic medical records (EMRs), Medical images (recordings, ecotomographs, ultrasound, resonance, etc.) Biological markers and Others (sensors and fetal heart rate)



<b>S/N</b>	<b>Author(s)</b>	<b>Year</b>	<b>Title</b>	<b>Methodology</b>
2	Dusseldorf,  Messe	2022	Improving the prediction of pregnancy complications using AI	The research was carried out based around 1,647 data from different preeclampsia patients and used different machine learning methods to create a functional model that would be able to predict the probability of a serious complication. They use dataset such as blood pressure and urine protein measurement and biomarkers such as sFLt and PIGF for their study, they wanted to test the accuracy and to explore the use of machine learning algorithms to analyze different tests results in order to get a better idea of who is at risk or not .

## **CHAPTER THREE**

### **RESEARCH METHODOLOGY**

#### **Introduction**

In this chapter, the methodology applied to this research work is clearly defined, this chapter start with a description of the framework for the research methodology, which explain the series of step taken: which start from search strategy, data identification and collection, training of data and model formulation. The different types of machine learning algorithms used for each complications were carefully selected based on the variables that were suited and stated how it was applied.

#### **3.1 Search strategy**

The use of different articles/studies were used as a guiding tools in the means of gathering the necessary data and features needed. So the use of different search engines were used such as Google scholar, science direct etc which were very useful.

Such articles were screened using some important keywords such as predicting pregnancy complications, using machine learning in pregnancy outcome, preterm birth predictions, vaginal(normal) delivery after c section or cesarean section and machine learning', 'cesarean section or c section prediction and machine learning', 'mode of childbirth prediction and machine learning', etc. This brought about different types of publications such as: Journals, conference articles etc.

### **3.2.1 ANALYSIS OF THE PROPOSED PREDICTED SYSTEM**

In building a predictive system there are five key stages in the predictive analytics process cycle which are:

1. Preparation for establishing pregnancy prediction model
2. Dataset selection/ collection.
3. Cleaning of data.
4. Training of dataset.
5. Model generation/evaluation.

### **3.2.2 Preparation for establishing pregnancy prediction model:**

The sole aim of the prediction model is to develop an accurate and useful pregnancy prediction model using multiple variables with the appropriate dataset. The first thing done would be to look for various research articles and questions that relate to the dataset selection and the approach or means to which the model would be developed. For example

1. What would be the target outcome (pregnancy complications) to predict
2. Who would be the target patient of the model (e.g pregnant women, babies)
3. Who would be the target user of the prediction model (e.g doctors etc)

So based on the answers to the above questions, the right datasets for the model can be chosen. The category of the target users will be used to determine the selection and handling of the multiple variables which would affect the structure of the model.

### **3.2.3 Dataset selection/ collection:**

Dataset is one of the most important aspect of the prediction model, it is ultimately determined by its quality and credibility. It is said that they are no general rules for determining the quality of data. Which can be said as no perfect data exists and a prefect model.so the most suitable dataset was collected and used. Different dataset would be used depending on the different purpose of the predictions. Due to the different dataset gathered, different models/algorithms were used based on their features. This section focuses on the method associated with determining and gathering all the information that would be used throughout the follow up on complications of women. Following consent by the head of medical doctors of hospital, the data needed for the development of the predictive model for complications of pregnant women were collected. Data consisting of information like each patient age, blood pressure etc. were collected from the health record, then converted and stored into spreadsheet format.

### **3.2.4 Cleaning of data:**

In this procedure, it's about the process of preparing data for analysis by either removing or editing data that are not correct, incomplete, Irrelevant or are duplicated. It entails about fixing spelling and errors, correcting mistake such as empty datasets and identifying duplicate data points. It is an important aspect because it can affect the accuracy of the model. Different ways of cleaning data

- (i) By fixing missing data
- (ii) Fixing structural errors

- (iii) Removing duplicated data
- (iv) Validating the data. **(Barkved, 2022)**

### **3.2.5 Training of dataset:**

Training of data is the initial dataset used to train the machine learning algorithms. It is this data that models refine their rules. It is this set of data samples that are used to fit into the parameters of the machine learning model. Machine learning can be said to be as good as the data they are trained on. It analyzes the dataset repeatedly to fully understand the characteristics and its features to be able to adjust itself for a better result. This is an important part of the model which help make accurate predictions. This tells us what expected output would look like.

Training of data can be divided into two which are

1. Labeled data: This are a group of data samples grouped with one or meaning labels, which identifies specific properties, classification. Labeled training data are used in supervised learning, it allows machine algorithms to learn the characteristics associated with the specific labels.
2. Unlabeled data: This is quite the opposite of label data, these are raw data that are not tagged with any labels. It is mainly used for unsupervised machine learning, in which the models have to find patterns in the data.

The trained data must contain the following for it to be good;

1. It must be relevant
2. It must be representative
3. It must be uniformed
4. It must be comprehensive.

### **3.3 Pregnancy complications studied**

They are different complication pregnant women can face before, during and after her pregnancy in which some cases can be avoided. Each pregnancy complication listed has been carefully studied and examined to be able to get some important features that can be used to create a predictive model. The pregnancy complications that would be used for the predicting of pregnancy complications were:

- High blood pressure
- Gestational diabetes
- Partum hemorrhage

**Table 3.3.1: Identified variables for determining high blood pressure pregnancy**

**Complications**

S/N	Variable names	Labels
1	Mother age	Numeric
2	Baby gender	Nominal
3	Mother education level	Nominal
4	Employment	Nominal
5	High blood pressure(bp) record	Numeric
6	History of high blood pressure	Nominal
7	History of diabetes	Nominal

**Table 3.3.2: Identified variables for determining Partum hemorrhage pregnancy**

**Complications**

S/N	Variable names	Labels
1	Mother age	Numeric
2	Baby gender	Nominal
3	Mother maternal height	Numeric
4	Mother maternal weight before pregnancy	Numeric
5	Mother maternal weight upon admission of pregnancy	Numeric
6	Birth weight of baby	Numeric
7	Mode of delivery	Nominal



**Table 3.3.3: Identified variables for determining Gestational diabetes pregnancy**

**Complications**

S/N	Variable names	Labels
1	Mother age	Numeric
2	Mother maternal height	Nominal
3	Pre pregnancy body max index(BMI)	Numeric
4	Mother maternal weight before pregnancy	Numeric
5	Mother maternal weight upon admission of pregnancy	Numeric
6	FPE	Numeric

### 3.4 Machine learning algorithms applied and software

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. There are so many different machine learning algorithms used in predicting outcome, especially based on the given variable and data set being used to predict. The variable used for predicting pregnancy complications in women was used to determine which machine algorithms would be best suited. The machine learning algorithms used were:

- (I) **High blood pressure-** Decision tree, random forest and logistics regression
- (II) **Partum hemorrhage-** Logistic regression, random forest and decision tree
- (III) **Gestational diabetes-** Logistic regression

To be able to predict the pregnancy complications in women, a suitable software application that would be able to perform the necessary task in predicting and making a model was used, the software used for this research was called Rapidminer. Rapidminer is a data science platform which was designed to support many analytics users across a broad AI lifecycle. Rapidminer is used to provide data mining and machine learning procedures which include things such as data loading and transformation, data preprocessing and visualization, predictive analytics etc. It has a graphical user interface tool that allows us to link elements on a palette to do data analysis.

## CHAPTER FOUR

### IMPLEMENTATION AND RESULT

#### 4.1 INTRODUCTION

In this section, the results and discussion of the development of a predictive model for predicting if a pregnancy would be successful or have any complication would be stated in this chapter.

##### 4.2.1 Selecting of task

This section deals with the selection of the required data set that would be worked on. In this section, you decide what type of problem you want to solve, they are three task to choose from:

**Predict:** This Select the task if you want to predict the values for one of the columns of your data, and identify the column by clicking on it. It will build a machine learning model which predicts the values of this column based on the values of the other columns.

The goal is to predict the values of one column based on the values of the other columns.

The general method is called *supervised learning*, because the data used to train the model has a known outcome.

Within the area of supervised learning, we can distinguish between the two methods, which are classification and regression, depending on the target. *Classification* is used when the target values are categorical. It answers questions like "Will the outcome be A or B?"

- Clusters: Select this task if you want to group your data into clusters. The goal here is not to predict the values of a single column, but to find sets of data points that are close together.
- Outliers: Select this task if you want to find unusual points in your data. The goal here is to find individual data points that are far away from all other data points, possibly because of errors in data collection or because of weird or unexpected behavior.

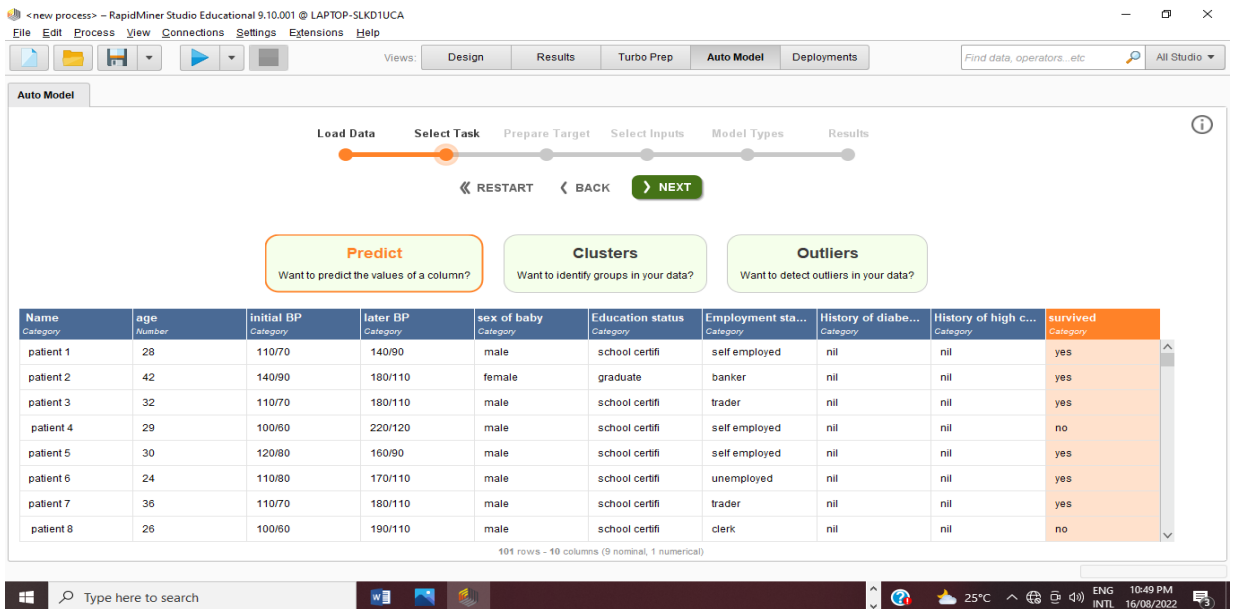


Fig 4.2.1: The selecting of which task to perform

## 4.2.2 Preparing a new Target

This section deals with solving a classification problem. A bar chart shows the number of data points in each class. At most 10 classes are shown, the 10 classes with the most data points. If these are the classes you want to predict, and everything looks okay, you then click the *Next* at the bottom of the screen.

If there are only two classes, you may choose which of the classes is of highest interest to you, and the performance measures for each model (displayed later, together with the results) will show specific performances for this class.

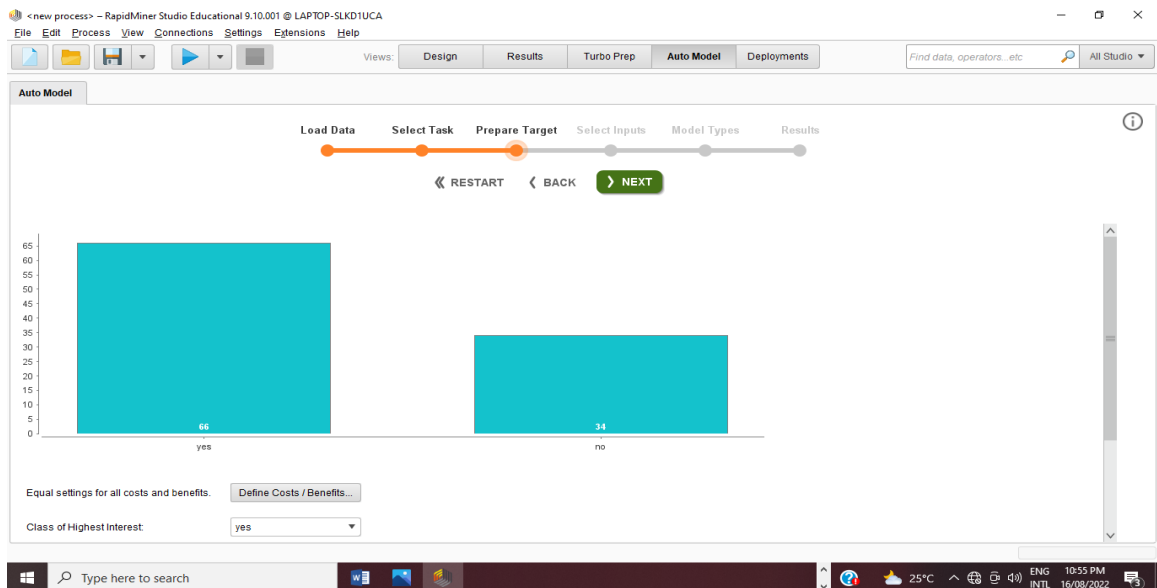


Fig 4.2.2: Preparing of target used

### 4.2.3 Selecting of input used

This section focuses on the quality of the data, specifically the quality of each column of data. It would be useful to consider discarding data columns (Attributes) that provide less value. How to know which Attributes are valuable, and which are worthless? A key point is that you're looking for patterns. Without some variation in the data and some discernible patterns, the data is not likely to be useful. To help in making a decision, it is indicated that the Attribute value with a color-coded status bubble (red / yellow / green).

The screenshot shows the 'Auto Model' workflow in RapidMiner Studio. The 'Select Inputs' step is currently active, showing a list of selected attributes. The table below displays the quality metrics for these attributes.

Selected	Status	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input checked="" type="checkbox"/>	Green		initial BP	1.70%	8.91%	33.00%	0.99%	5.67%
<input checked="" type="checkbox"/>	Green		later BP	1.77%	19.80%	15.00%	0.99%	9.69%
<input checked="" type="checkbox"/>	Green		Education status	0.53%	5.94%	47.00%	0.99%	22.40%
<input checked="" type="checkbox"/>	Green		Employment status	0.78%	26.73%	16.00%	0.99%	28.12%

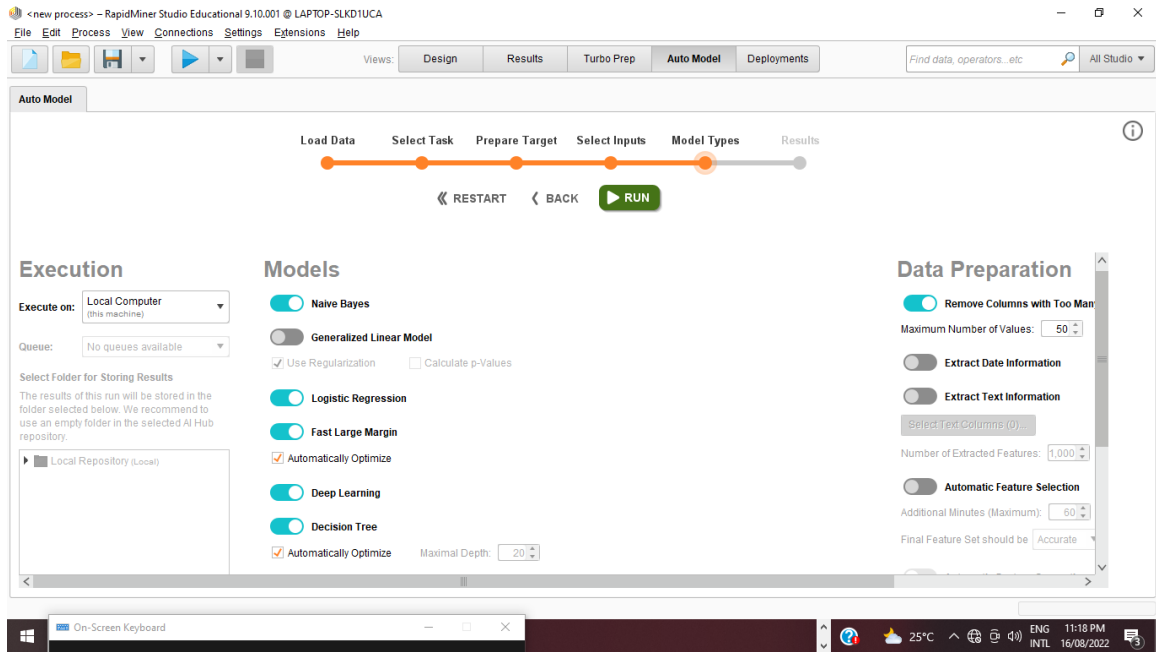
Fig 4.2.3: Selecting of input (dataset) used

#### 4.2.4 Selecting Model Type

Based on the data and your choices, relevant machine learning models are displayed. You can select more than one model, the results will include a performance comparison. Then you need to select at least one of the models before you can click on the *Run* button at the bottom of the screen. Depending on the data set, some models may be deselected, to avoid long runtime.

#### For Data Preparation

- **Extract Data Information:** if seasonal information should be extracted from date columns or if date columns should be ignored instead. If acticated, Auto Model extract numerical features from date columns and also calculates differences of dates among themselves as well as to today.
- **Extract Text Information:** if numerical information should be extracted from text columns, they can remove or add other columns if you want to. More extracted features can help to identify more subtle patterns in your data but will slow down modeling time. In addition, you can also extract generic sentiment and the language for each text column.
- **Automatic Feature Selection:** this identifies optimal column subsets and all optimal trade-offs between complexity and accuracy
- **Automatic Feature Generation:** this generates new columns in addition to feature selection



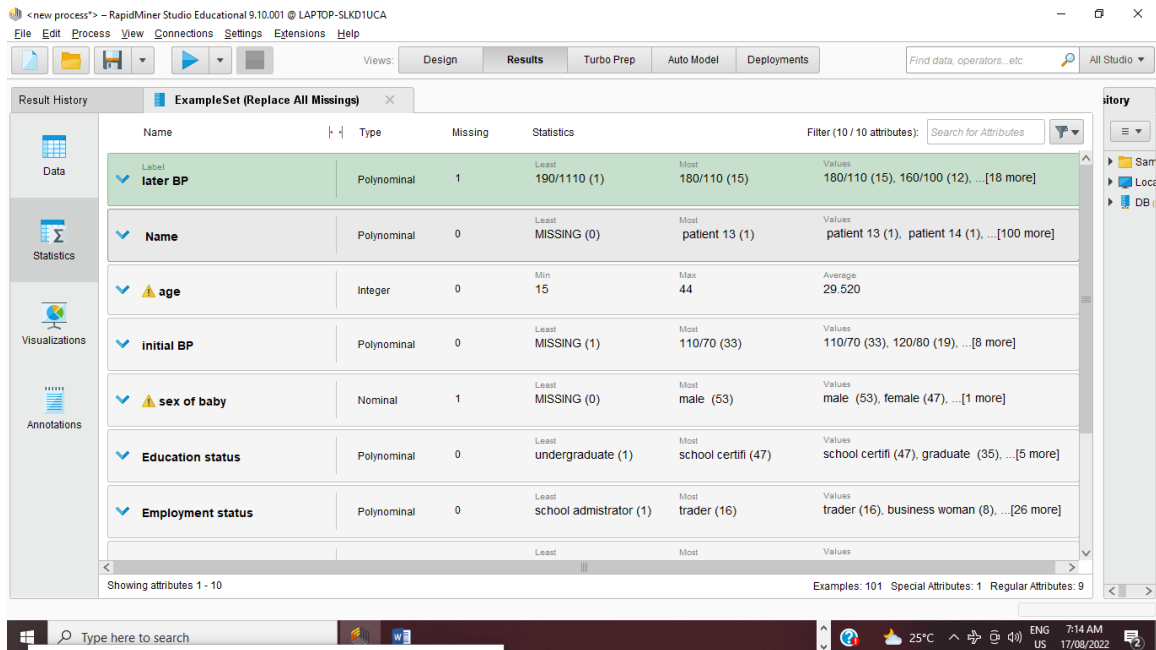
**Fig 4.2.4: selecting of ML model to use**



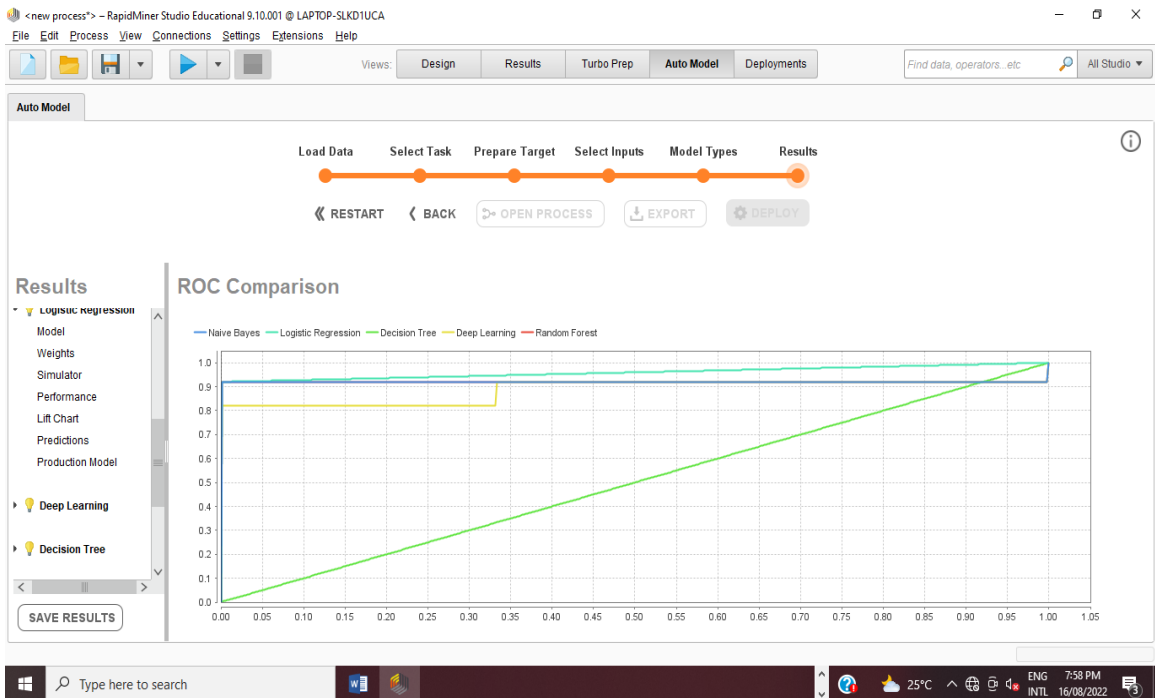
### 4.3 Result of data preprocessing of the data collected

After the collection of the data sets required for this study were collected from hospitals around lagos, the data was preprocessed using rapid miner with the use of the auto model function. **Fig 4.3.1** shows the preprocessed data collected

ROC Comparisons: shows the ROC curves for all models, together on one chart. The closer a curve is to the top left corner, the better the model is. Only shown for two-class problem. **Fig 4.3.2** shows the AOC Curve for the machine learning model used



**Fig 4.3.1: Result of data preprocessing of the data collected**



**Fig 4.3.2: AOC Curve for the machine learning model used**

## 4.4 Machine learning Simulator

For predicting pregnancy complications using machine learning model, this shows an easy-to-use, real-time interface to change the inputs to a model and view the output. It shows predictions, confidences, and explanations for those inputs. So here you input the variable that would be used to predict the possible outcome based on the previous dataset that has been trained. Different machine learning can be chosen based on which one best suited to the user.

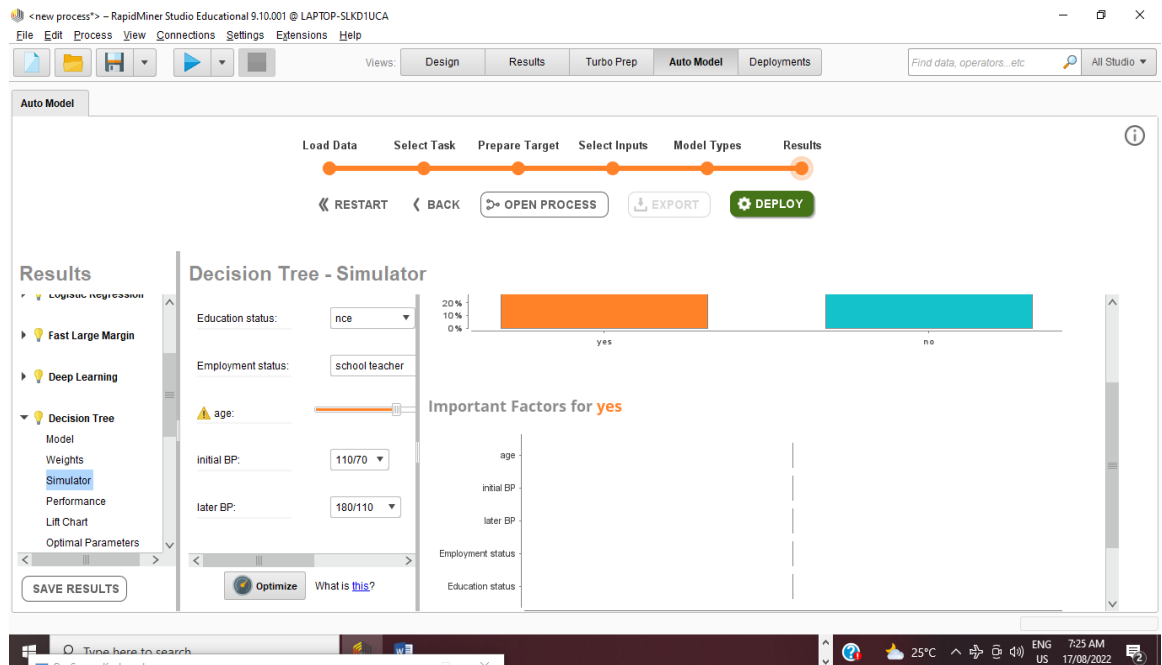


Fig 4.4: Decision tree simulator

## **CHAPTER FIVE**

### **SUMMARY, CONCLUSION, AND RECOMMENDATIONS**

#### **5.1 SUMMARY**

After going through most of the complications pregnant women can go through and the risk they can go experience, this proposed model is believed to help eradicate the problems associated with the complication issues experienced by women, which is aimed at detecting early signs or symptoms on whether a woman pregnancy would be successful or have some complications along the pregnancy.

#### **5.2 CONCLUSION**

This project work presents a predicting model that would be based on past record or data using different machine learning algorithm that would be best suited in predicting various outcome based on some given variable collected as input to get an output. This model can be applied in various hospital setting with an appropriated system to get the best use of the system. It is envisaged that by using this model, the problem of pregnancy related complications would be reduced in Nigeria.

### **5.3 CHALLENGES**

Some of the challenges encountered while carrying out this project include gathering and handling of large data set.

## References

- Ali, R. (2020, September 23). *Predictive Modeling: Types, Benefits, and Algorithms*. Retrieved from Netsuite.
- American college of obstetricans and Gynecologists. . (2020, December 30). *Preeclampsia and high blood pressure during pregnancy*. Retrieved from American college of obstetricans and Gynecologists: <http://www.acog/patients/FAQs/preeclampsia-and-high-blood-pressure>
- Brown, S. (2021, april 21). *MACHINE LEARNING,explained*. Retrieved from MIT MANAGEMENT.COM.
- cafasso, J. (2016, May 25). *Complications during pregnancy and delivery*. Retrieved from Healthline.com: <https://www.healthline.com/health/pregnancy/delivery-complications>
- Encyclopedia of Bioninformatics and Computational Biology . (2019). Predicting Modelling-an overview . *ScienceDirect Topics* , 1-2.
- Indeed Editorial Team. (2021, November 16). *10 Predictive Modeling Types (With Benefits and Uses)*. Retrieved from Indeed.com: <https://www.indeed.com/career-advice/career-development/predictive-modeling-types>
- M.C.Romero-Ternero, A. M. (2020, october). *Artificial Intelligence in pregnancy: A scoping review*. Retrieved from Researchgate.com: [www.researchgate.net](http://www.researchgate.net)
- predictive success coporation . (2019 , May 6). *A Brief History of Predictive analytics* . Retrieved from meduim.com: <https://meduim.com/>
- shybeko, A. (2021). *uses of predictive modelling in healthcare* . Retrieved from light-IT .net: <https://light-it.net/>
- Uscher.Jen. (2020, June 12). *Anemia in pregnancy*. Retrieved from Grow by WebMD: <https://www.webmd.com/baby/guide/anemia-in-pregnancy>
- World Health Organization . (2019, June 25). *who.int.com*. Retrieved from Maternal health in Nigeria: generating information for action: <https://www.who.int/news/item/25-06-2019-maternal-health-in-nigeria-generating-information-for-action>

















