FETAL BIRTH WEIGHT ESTIMATION

USING ADVANCED MACHINE LEARNING

A PROJECT REPORT

Submitted by

KIMAXY.M	**211422243158**
MADHESWARI G B	**211422243181**
MONIKA.N	**211422243200**

in partial fulfillment for the award of the

degree of

BACHELOR OF ENGINEERING

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

PANIMALAR ENGINEERING COLLEGE CHENNAI-600123

ANNA UNIVERSITY: CHENNAI-600 025

NOVEMBER,2024

BONAFIDE CERTIFICATE

Certified that this project report titled "**FETAL BIRTH WEIGHT ESTIMATION USING ADVANCED MACHINE LEARNING**" is the bonafide work of **KIMAXY M**, **MADHESWARI G B, MONIKA N** Register No. **211422243158, 211422243181, 211422243200** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

INTERNAL GUIDE
Mrs. LOGAPRIYA, M.E,
ASSITANT PROFESSOR,
Department of AI &DS.

HEAD OF THE DEPARTMENT
Dr. S.MALATHI M.E., Ph.D
Professor and Head,
Department of AI & DS

Certified that the candidate was examined in the Viva-Voce Examination held on ………………………

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

I also take this opportunity to thank all the Faculty and Non-Teaching Staff Members of Department of Artificial Intelligence and Data Science for their constant support. Finally I thank each and every one who helped me to complete this project. At the outset we would like to express our gratitude to our beloved respected Chairman, **Dr.Jeppiaar M.A.,Ph.D,** Our beloved correspondent and Secretary **Mr.P.Chinnadurai M.A., M.Phil., Ph.D.,** and our esteemed director for their support.

We would like to express thanks to our Principal, **Dr. K. Mani M.E., Ph.D.,** for having extended his guidance and cooperation.

We would also like to thank our Head of the Department, **Dr. S. Malathi M,E.,Ph.D.,** of Artificial Intelligence and Data Science for her encouragement.

Personally we thank, **Mrs. LOGAPRIYA, M.E,** Department of Artificial Intelligence and Data Science for the persistent motivation and support for this project, who at all times was the mentor of germination of the project from a small idea.

We express our thanks to the project coordinator **Ms.K.CHARULATHA M.E.,** Assistant Professor in Department of Artificial Intelligence and Data Science for their Valuable suggestions from time to time at every stage of our project.

Finally, we would like to take this opportunity to thank our family members, friends, and well-wishers who have helped us for the successful completion of our project.

We also take the opportunity to thank all faculty and non-teaching staff members in our department for their timely guidance in completing our project.

KIMAXY M,

MADHESWARI G B,

MONIKA N

ABSTRACT

This project aims to address the critical issue of low birth weight, as highlighted by recent UNICEF data indicating that over 7 out of every 100 babies born in 2015 weighed less than the recommended 2.5 kg. Low birth weight is associated with increased risks of infant mortality, juvenile illnesses, and developmental challenges, all of which can have lifelong impacts. Various factors influence birth weight, including genetic and environmental elements such as the baby's sex, maternal and paternal ages, maternal weight gain during pregnancy, and lifestyle habits like smoking and alcohol consumption. The goal of this project is to develop a predictive model using machine learning algorithms to estimate the birth weight of infants based on these factors, allowing for early identification of pregnancies at risk. By comparing a range of machine learning techniques—such as linear regression, decision trees, random forests, and gradient boosting—we will identify the most accurate algorithm for this purpose. Once the best model is established, it will not only provide an estimate of the baby's birth weight but also generate tailored dietary and lifestyle recommendations for the mother to potentially reduce low birth weight risks. This initiative will involve comprehensive data collection on maternal, paternal, and environmental variables and rigorous feature engineering to enhance predictive power. Through this project, we aim to support healthcare providers in delivering personalized prenatal care, thus promoting healthier outcomes for both mothers and infants. In addition to offering individualized health advice, the model can contribute to broader public health efforts by identifying common factors linked to low birth weight, helping inform preventive measures and nutritional programs. By integrating this predictive model into prenatal care, we can provide an evidence-based tool that mitigates risks associated with low birth weight, ultimately fostering improved maternal and infant health on a larger scale.

Table of Contents

LIST OF FIGURES

LIST OF ABBREVATIONS

LBW	-	Low Birth Weight
WHO	-	World Health Organization
SVM	-	Support Vector Machine
ML	-	Machine Learning
SVR	-	Support Vector Regression
KKN	-	K-Nearest Neighbor
UNC	-	Universal Naming Convention
RMSE	-	Root Mean Square Method
SCHS	-	State Center for Health Statistics

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

Every parent on the planet wishes for a healthy child. A baby's weight is the best indicator of his or her health. The prediction of birth weight in the normal range indicates to parents and doctors that the baby is healthy. A substantial deviation may indicate an abnormality. It could be due to a shift in the mother's habits or an unintentional alteration in the environment.

Low Birth Weight (LBW) is used to describe babies with birth weights less than 2,500 grams. The average birth weight is about 3500 grams. The overall rate of low birth weight has been increasing all over the world. This could be because of increasing multiple births, lifestyle, stress, education of the parent, mother’s health, etc.

Babies with LBW are at high risk for complications. Babies with LBW have a hard time eating, gaining weight, fighting infections, and staying warm in normal temperatures due to low body fat. In general, babies with LBW are also premature, and the lower the birth weight, the greater the risk for complications. Survival of infants with low birth weight depends largely on how much the baby weighs at birth, with the smallest babies (<500 grams) having the lowest survival rate.

Our study tests various machine learning algorithms to predict birth weight with minimal errors. Depending on the prediction received from the machine learning model, we will be able to suggest health plans to improve the health of the mother and baby if needed. This can lead to earlier detection of potential health risks, enabling timely interventions that can improve outcomes for both the mother and child. By incorporating such predictions into clinical practice, we hope to reduce the incidence of low birth weight and its associated complications.

1.2 SCOPE AND OBJECTIVE

SCOPE

Many pregnant women face significant barriers that hinder their access to crucial prenatal care, affecting both maternal and infant health outcomes. Financial constraints are a major factor, as many women cannot afford prenatal services, leading them to delay or skip essential checkups. Additionally, there is often a lack of awareness regarding the importance of prenatal care, with many women not fully understanding its role in supporting a healthy pregnancy. For women experiencing poverty, the stress of financial strain and limited resources compounds these issues, further deterring access to care. High stress levels, especially in high-risk women, have been linked to poor birth outcomes, including low birth weight and premature delivery. This project, therefore, not only aims to predict birth weight but also considers these socio-economic barriers, integrating them into a model that can better support personalized recommendations and improve outcomes for underserved women and their infants.

OBJECTIVE

The purpose of our study is to develop a predictive system that estimates an infant's birth weight prior to delivery using advanced machine learning algorithms. This system will help expectant mothers and healthcare providers identify cases at risk for low birth weight, a condition linked to various health complications for newborns. By analyzing maternal and environmental factors—such as maternal age, weight, lifestyle habits, and prenatal care patterns—the model will provide an estimated birth weight, giving users an early warning if the predicted weight falls below recommended thresholds. This early detection system aims to empower users to seek timely professional guidance and make informed decisions about their prenatal care.

CHAPTER 2
LITERATURE SURVEY

2.1 INTRODUCTION

Birth weight has a significant influence on chances of survival. Low birth weight (LBW) is becoming more of an issue, particularly in emerging countries. A major cause of neonatal death is low birth weight, defined as less than 2500 g. Babies born at a low birth weight are 25 times more likely to die than babies born at a normal birth weight. It's also a good indicator of a child's future health complications. Low birth weight affects one out of every seven newborns, accounting for about 14.6 percent of the babies born worldwide. The prevalence varies substantially by region, with rates of 7.2 percent in the More Developed Regions, 13.7 percent in Africa, and 17.3 percent in Asia, respectively. These newborns are more likely to die within the first month of birth or to experience long-term implications, including stunted growth, low IQ, overweight or obesity, developing heart disease, diabetes, and early death. We have reviewed research papers that discuss the effect of birth weight on mortality rates, the prediction of baby weight, the importance of birth weight, the relationship between prenatal care and birth weight, and the effect of maternal anemia on birth weight, highlighting the need for targeted interventions to address these issues effectively. This comprehensive understanding emphasizes the critical importance of improving maternal health and access to prenatal care to mitigate the risks associated with low birth weight.

2.2 LITERATURE SURVEY

1. **FetalCare: Enhancing S. P. Shetty, S. A. Patil, S. M. Jeelan, S. Pranav and C. Deepti, "FetalCare: Enhancing Prenatal Care through a Machine Learning Approach for Fetal Health Classification and Birth Weight Prediction,"** ***2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*****, Salem, India, 2024.**

The process of navigating the intricate landscape of maternal and infant health, allows the research work described in this paper to confront the formidable challenge of preterm birth risk through a sophisticated machine learning framework tailored for prenatal care. The focus on both fetal health classification and birth weight prediction allows this work to employ different algorithms—Support Vector Classifier, Random Forest, AdaBoost, and Decision Tree by
seamlessly integrating tem through a refined ensemble approach. This strategic fusion prioritizes precision in categorizing fetal health into "NORMAL," "SUSPECT," or "PATHOLOGICAL," reflecting an unwavering commitment to accurate identification, especially in critical health conditions. Rigorous evaluation metrics, notably the macro recall score, underscore the dedication to robust assessments essential in addressing the multifaceted nature of preterm birth risks. S imultaneously, the birth weight prediction component employs a combination of AdaBoost and Random Forest regressors, further enhancing the accuracy of estimations.

2. **Khan, W., Zaki, N., Masud, M.M. et al. Infant birth weight estimation and low birth weight classification in the United Arab Emirates using machine learning algorithms. Sci Rep 12, 12110 (2022).**

Correct prediction of a newborn's delivery weight (BW) is an essential determinant to evaluate the newborn's fitness and protection. toddlers with low BW (LBW) are at a higher threat of serious brief- and lengthy-time period health effects. during the last decade, system-getting-to-know (ML) strategies have shown a hit breakthrough in the field of scientific diagnostics. diverse computerized systems have been proposed that use maternal functions for LBW prediction. However, each proposed gadget uses extraordinary maternal features for the LBW category and estimation. Consequently, this paper affords a detailed setup for BW estimation and LBW category. more than one subset of features was mixed to carry out predictions with and without function choice techniques. Furthermore, the artificial minority oversampling method was employed to oversample the minority magnificence. The performance of 30 ML algorithms changed into evaluated for each little one BW estimation and LBW classification. Experiments were carried out on a self-created dataset with 88 capabilities. The dataset was obtained from 821 women from three hospitals in the United Arab Emirates. distinctive overall performance metrics, which imply absolute blunders and mean absolute percentage mistakes, were used for BW estimation. Accuracy, precision, consider, F-scores, and confusion matrices were used for the LBW type. Sizable experiments carried out with the usage of 5-folds go validation show that the best weight estimation turned into received the usage of Random forest algorithm with suggesting absolute blunders of 294.53 g even as the excellent class overall performance became acquired the usage of Logistic Regression with SMOTE oversampling strategies that done accuracy, precision, consider and F1 rating of 90.24%, 87.6%, 90.2% and 0.89, respectively. The consequences additionally propose that functions consisting of diabetes, hypertension, and gestational age, play a critical function in LBW class.

3. **Abdullah Zahirzada, Kittichai Lavangnananda . " Implementing Predictive Model for Low Birth Weigh" 2021 13th International Conference on Knowledge and Smart Technology (KST) DOI: 10.1109/KST51265.2021.9415792.**

Start weight is a massive determinant of the chance of survival of an infant. Low beginning Weight (LBW) has emerged as a more and more common hassle, mainly in growing andunderdeveloped international locations. Therefore, the capacity to predict the LBW is beneficial and an awesome safety measure. It's also a very good indicator of the destiny health risks of that infant. This study is concerned with the implementation of predictive LBW models for rural and concrete regions of Afghanistan based totally on the information acquired from the Afghanistan Demographic and fitness Survey 2015. the main goal of the observation is to pick out the most suitable system mastering techniques (i.e. classifiers) and a number of the five popular ones. These are ok-Nearest Neighbor (okay-NN), Naïve Bayes, Neural community, Random forest, and Support Vector system (SVM). prior to imposing the predictive models, information preprocessing is done. that is achieved by using records cleansing and characteristic selection. The Correlation based totally feature selection set of rules (CFS) is employed to choose the pinnacle ten attributes. The classifier in this examination accommodates categories, normal and LBW. The education of the dataset is carefully accomplished to make certain properly-balanced samples in each class. The study well-known shows that Random wooded area is the high-quality classifier with an accuracy of 84.70% and 85.2% and with the vicinity below the Curve (AUC) of 91.0% and 92.1% for both rural and urban regions, respectively.

4. **N. S. Borson, M. R. Kabir, Z. Zamal and R. M. Rahman, "Correlationanalysis of demographic factors on low birth weight and prediction modeling using machine learning techniques," 2020 Fourth World Conference on Smart Trends in Systems, Security, and Sustainability (WorldS4), 2020, pp. 169-173, DOI: 10.1109/World S450073.2020.9210338.**

In keeping with a national survey in Bangladesh, a south Asian united states of America, about 22.6 percent of the brand newborn infants are born with low start weight (below 2.5 kg or 2500 grams). There are a few key factors regarding low start weight which are clinically diagnosed but apart from the clinical attitude some other health and demographic factors additionally play a vital role in this phenomenon which can be directly or circuitously associated. The purpose of thisstudy is to make use of the capacity machine studying algorithms to assemble a predictive model for the low beginning weight given some health and demographic data associated with neonatal fitness situations within the context of Bangladesh. For the predictive evaluation, algorithms like Logistic Regression, Naïve Bayes, Random woodland, okay-Nearest Neighbor, aid Vector device, and Neural network models were utilized in the look at. The findings of this examination may bea guiding principle for fitness professionals as well as researchers for reading low beginning weight toddlers that could help human beings in mass to understand and take important precautions to keep away from any such event in which a child is born with much less weightthan the common.

5. **Olli-Pekka Rinta-Koski, Simo Sarakka, Jaakko Hollmen,Markus Leskinen, Krista Rantakari, Sature Andersson . " Prediction of Major complications affecting very low birth weight infants" Published in 2017 IEEE Life Sciences Conference (LSC) DOI: 10.1109/LSC.2017.8268174.**

Bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), and retinopathy of prematurity (ROP) are excessive complications affecting Very Low Beginning Weight (VLBW) infants. Our findings display that information collected inside the intensive care unit at some point during the primary 24 or 72 hours of care can be used to predict whether a VLBW little oneis liable to developing BPD. Using the Gaussian procedure class, we finished classification outcomes with areas underneath the receiver operator feature curve of 0.85 (preferred mistakes (SE) 0.05) for 24h and 0.87 (SE 0.06) for 72h BPD facts. This compares favorably withconsequences carried out using the scientific fashionable SNAP-II and SNAPPE-II scores. Sensitivity for BPD became 0.52 (SE 0.06). Sensitivity for NEC and ROP became near zero, suggesting that NEC and ROP cannot be reliably expected with this approach from our information set.

6. **M. Abdollahian and N. Gunaratne, "Low Birth Weight Prediction Based on Maternal and Fetal Characteristics," 2015 12th International Conference on Information Technology - New Generations, 2015, pp. 646-650, DOI: 10.1109 /ITNG.2015.108.**

The newborn size is an important indicator of little one survival and childhood morbidity and seems to be related to a subsequent danger of kind 2 diabetes, high blood pressure, cardiovascular ailment, and other disorders consequently, many studies have attempted to perceive assets of the variant in newborn length. The purpose of this study is to decide whether the correct prediction of time period delivery weight is possible based totally on maternal and fetal characteristics mechanically measured far off from the time period. more than one linear regression is deployed to outline which mixtures of those variables are substantial the use of actual records accrued in a maternity medical and start weight prediction equations are advanced.The fashions are then used to predict the shipping weight for the Low start Weight (LBW) babies. The efficacy of the prediction models is assessed and compared based on their suggestions and trendy blunders of the predicted weights. The paper proposes two regression models based on a few measurable traits of each mother and fetus. The fashions can explain 62.9% and 59.4% of the shipping weight version for the low birth weight infants. The proposed fashions had been then used to estimate the recorded weights collectively with their corresponding 95% confidence and prediction durations for the LBW infants. The consequences suggest that the maximum vast factors for the decreased regression version are head circumference, gestation age, and fetal length. The decreased model can explain 59.4% of the delivery weight version for the low birth weight infants. while the regression version primarily based on the above predictors as well as mom hemoglobin degree, chest circumference, andmom top and BMI can give an explanation for 62.9% of the transport weight variant for the low birth weight infants.

7. **Agarwal K, Agarwal A, Agrawal VK, Agrawal P, Chaudhary V. Prevalence and determinants of " low birth weight" among institutional deliveries. Ann Nigerian Med. 2011;5(2):48-52.**

Beginning weight is a crucial determinant of baby survival and improvement. it's also difficult for medical and epidemiological investigations. This observation was deliberate to discover the epidemiological factors associated with low birth weight (LBW) among institutional deliveries so that appropriate advice may be made to save you LBW targets: The gift have a look at turned

into therefore undertaken to find out some maternal factors that can have their affiliation if any with LBW.This move-sectional look at became done at a tertiary care hospital among 350 moms turning in stray-born neonates to have a look at the location. All toddlers had been weighed within 24 hours after the delivery. The babies had been weighed on a beam kind of weighing machine with up to 20 g accuracy. LBW became described as having a birth weight of <2500 grams. All moms had been examined and interviewed within 24 hours after transport andfindings had been recorded. The analysis became executed with the use of an Epi information bundle, in this study, 40.0% of moms brought LBW toddlers. Findings suggest thatgestational age less than 37 weeks (76.5%), maternal age less than 20 years (58.5%), abnormal antenatal checkup (70.5%), mother′s top much less than a hundred and 50 cm (68.5%), mom′s weight less than 50 kg (76.1%), hemoglobin much less than 10 gm/dl (60.5%), severe bodily paintings (78%), and tobacco chewing (58.5%) are extensive determinants of LBW.Our observation indicates that gestational age, maternal age, regular antenatal checkup, mother′s peak, mom′s weight, anemia, bodily paintings, and tobacco chewing are considerable determinants of LBW. occurrence of LBW can be reduced through increasing the gestational age, ordinary antenatal checkups, a balanced weight loss plan in the course of antenatal length, ok relaxation at some stage in antenatal length, and heading off the tobacco chewing.

8. **Olli-Pekka Rinta-Koski, Simo Sarakka, Jaakko Hollmen,Markus Leskinen, Krista Rantakari, Sature Andersson . " Prediction of Major complications affecting very low birth weight infants" Published in 2017 IEEE Life Sciences Conference (LSC) DOI: 10.1109/LSC.2017.8268174.**

Bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), and retinopathy of prematurity (ROP) are excessive complications affecting Very Low Beginning Weight (VLBW) infants. Our findings display that information collected inside the intensive care unit at some point during the primary 24 or 72 hours of care can be used to predict whether a VLBW little oneis liable to developing BPD. Using the Gaussian procedure class, we finished classification outcomes with areas underneath the receiver operator feature curve of 0.85 (preferred mistakes (SE) 0.05) for 24h and 0.87 (SE 0.06) for 72h BPD facts. This compares favorably withconsequences carried out using the scientific fashionable SNAP-II and SNAPPE-II scores. Sensitivity for BPD became 0.52 (SE 0.06). Sensitivity for NEC and ROP became near zero, suggesting that NEC and ROP cannot be reliably expected with this approach from our information set.

CHAPTER 3

SYSTEM DESIGN

3.1 INTRODUCTION

Design is a multi-step that focuses on data structure software architecture, procedural details, algorithms, and the interface between modules. The design process also translates the requirements into the presentation of software that can be accessed for quality before coding begins.

This study of ML flows first displays the data source, pre-processing stage, feature selection for the classifier, parameter tuning, and classifier's functioning technique for the classification.

3.2 EXISTING SYSTEM

- ❖ Complicated math equations are used to predict birth weight.
- ❖ Fetal Growth Calculator by WHO

DISADVANTAGES

- ❖ The accuracy is less comparatively.
- ❖ No suggestion on how to improve health if required
- ❖ Mathematical equations should be done manually.

3.3 PROPOSED SYSTEM

- ❖ To build a simple machine learning-based webpage which predicts the birth weight of the baby based on the values given by the users.

- Suggests seeking professional help when required
- The various machine learning algorithms are tested.
- The algorithm with the highest accuracy score is used to build the machine learning model for the web page.

ADVANTAGES

- The proposed system provides more than 90% accuracy.
- Our model uses python libraries for implementation which can be accessed with any system configuration.
- The model is thereby, faster and lighter saving the time of the users.
- User-friendly UI for the website.

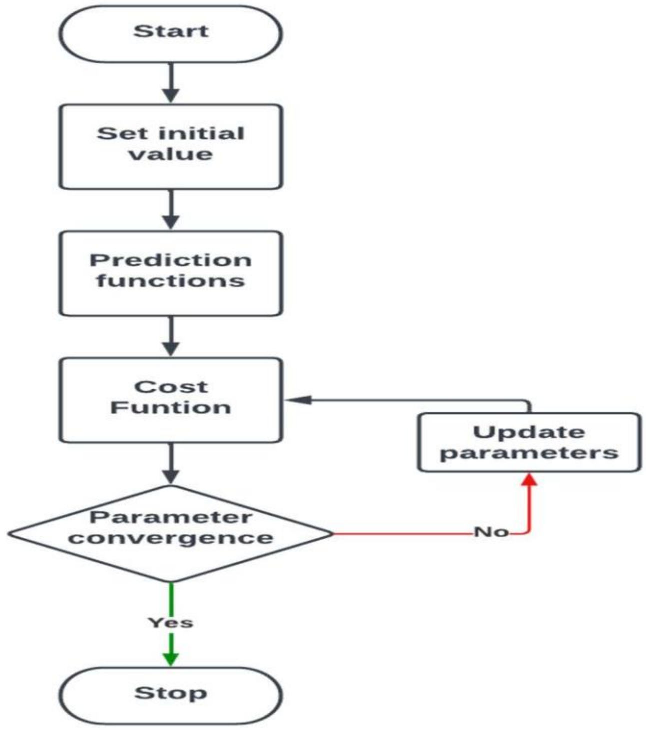
3.4 ALGORITHM

Various machine learning algorithms were tested to see which was best suited and gave accurate results for the proposed system.

Algorithms used (Linear Regression, RandomForest Regressor, Support vector Regression, Lasso Regression, Decision Tree Regression, and KNN model)

Linear Regression

It is one of the machine learning algorithms based on supervised learning. Linear regression is used to predict the dependent variable by the use of values of one or more variables. Through this, we are able to understand the procedure for the estimation of the data in a simple manner.



\

Fig 3.4.1 Linear Regression

Random Forest Regressor

Random Forest Regressor is a powerful, versatile machine learning algorithm that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting, making it suitable for both regression and classification tasks. It uses an ensemble learning approach, aggregating predictions from multiple decision trees to produce a more reliable and robust output. This approach helps to handle complex data relationships and large datasets more effectively by mitigating biases and variance commonly associated with single decision trees. Additionally, it can naturally handle missing values and provide feature importance scores, aiding in model interpretability.

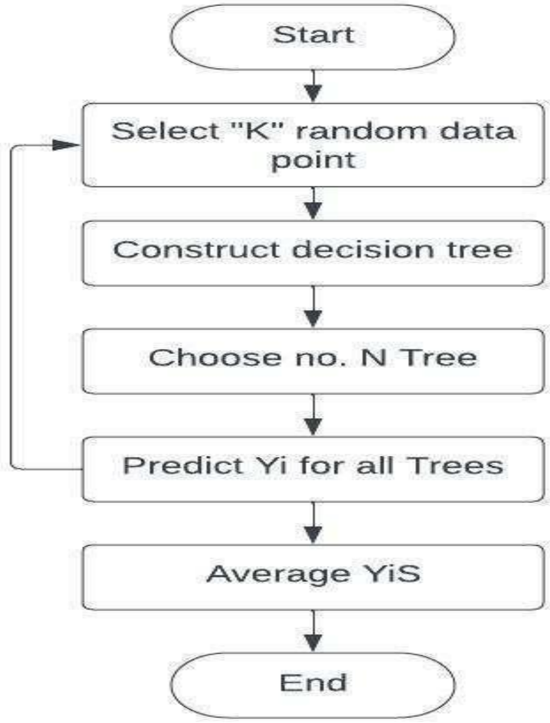


Fig 3.4.2 Random Forest Regressor

Support Vector Regression

Support vector Regression is based on a supervised machine-learning algorithm. It is used to predict discrete values and it is a regression algorithm that supports both linear and non-linear regressions. It has the same principles as the SVMs. We use SVR to find the best-fit line within a threshold value. In this, the best-fit line is the hyperplane which has the maximum number of points.

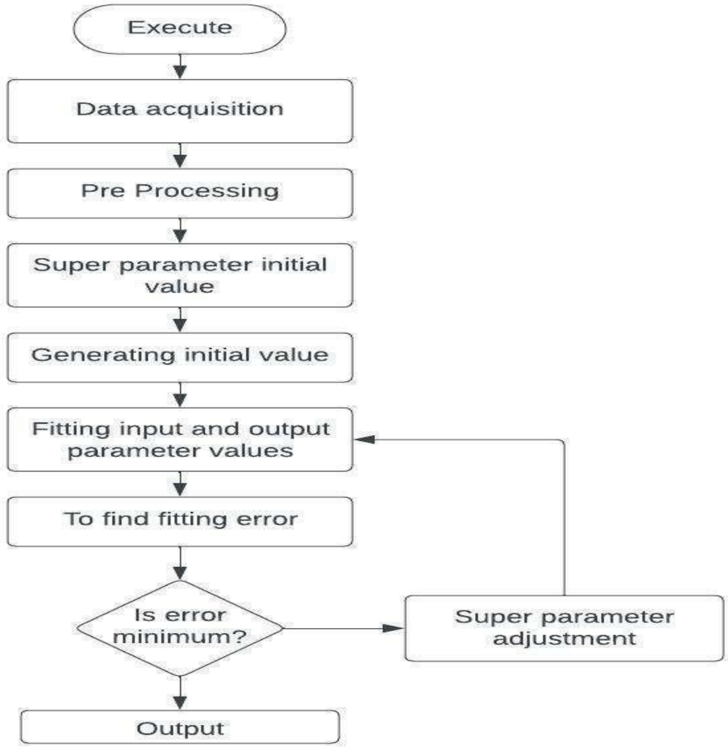


Fig 3.4.3 Support Vector Regression

Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) is a linear regression technique that applies L1 regularization to both shrink model coefficients and perform feature selection, making it useful for improving model interpretability and predictive accuracy, especially in high-dimensional datasets. By adding a penalty proportional to the absolute values of the coefficients, Lasso encourages sparsity, driving some coefficients to zero, effectively removing less important features from the model. This makes Lasso particularly valuable in scenarios where there are many predictors, helping to reduce overfitting and improve generalization. Compared to Ridge regression, which applies L2 regularization and retains all features with smaller coefficients, Lasso can exclude irrelevant variables entirely. Elastic Net combines the strengths of both Lasso and Ridge, making it effective when predictors are highly correlated.

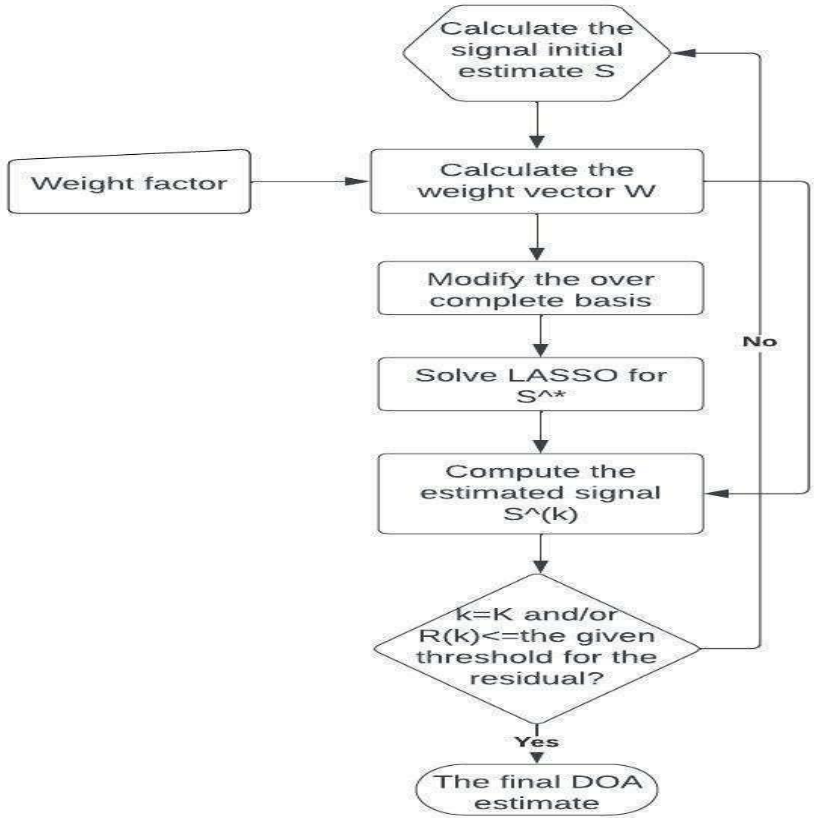


Fig 3.4.4 Lasso Regression

Decision Tree Regression

Decision tree regression trains a model in the form of a tree to predict data in the future and generate useful continuous output by observing the properties of an item. Continuous output denotes the absence of discrete output, i.e., output that is not only represented by a discrete, well-known set of numbers or values. This method allows for capturing complex patterns in data, making it suitable for tasks such as predicting prices, temperatures, or any variable where outcomes can take on a range of values. Additionally, the interpretability of decision trees provides valuable insights into the decision-making process, enabling users to understand how different features contribute to the final prediction.

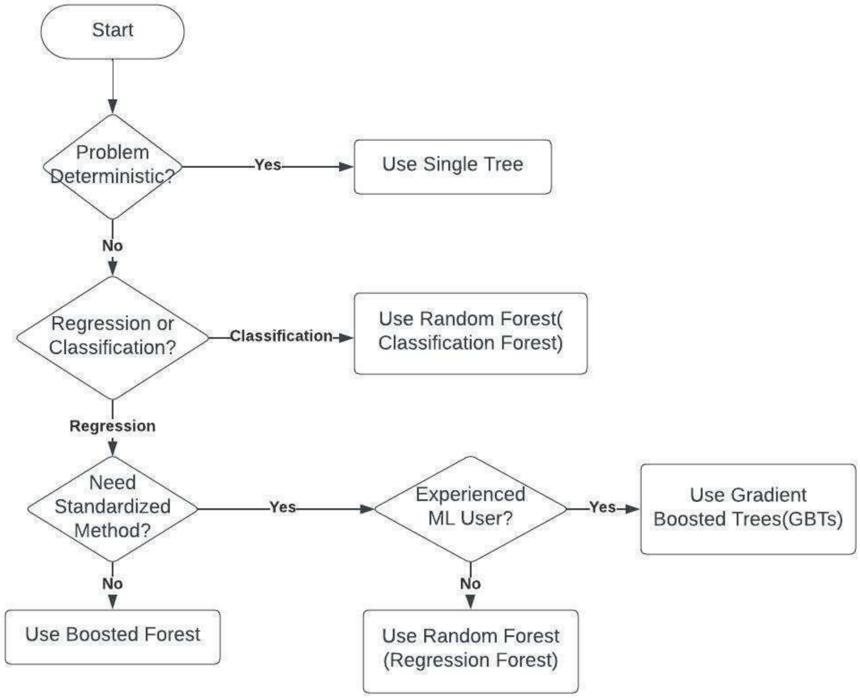


Fig 3.4.5 Decision Tree Regression

KNN model

KNN (K-Nearest Neighbors) is a machine learning algorithm based on supervised learning techniques, employed for both regression and classification problems. It is an easy-to-implement method that can effectively handle datasets, allowing for the imputation of missing values by identifying and utilizing the closest available data points. The algorithm works by storing all available data and classifying new data points based on the majority class or average value of their k-nearest neighbors. Additionally, KNN is versatile and can adapt to various types of data, making it suitable for applications in fields such as finance, healthcare, and marketing. However, its performance can be influenced by the choice of k, the distance metric, and the presence of irrelevant features in the dataset.

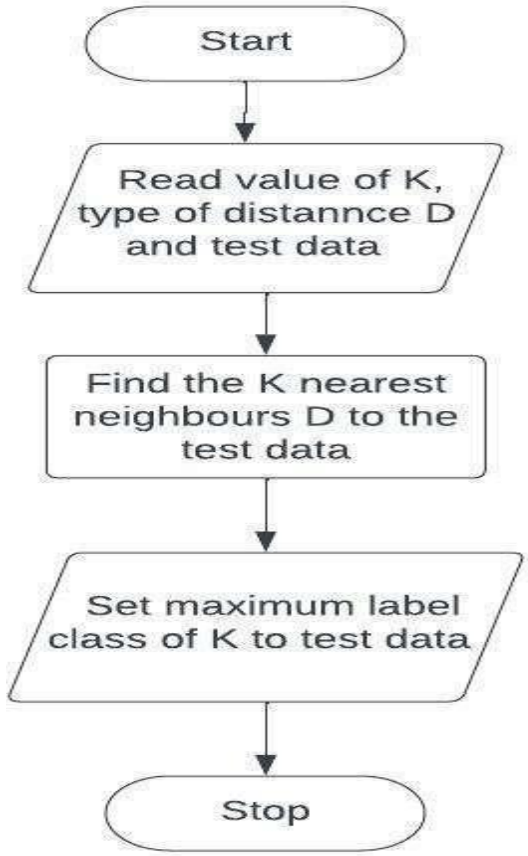


Fig 3.4.6 KKN Model

3.5 SYSTEM ARCHITECTURE DIAGRAM

The website provides instructions to the user on what the website does and how to use it. After that, the users can move on to providing the requested information. Then the prediction is made based on machine learning algorithms and the dataset provided. The predicted weight is printed and suggestions are made accordingly.

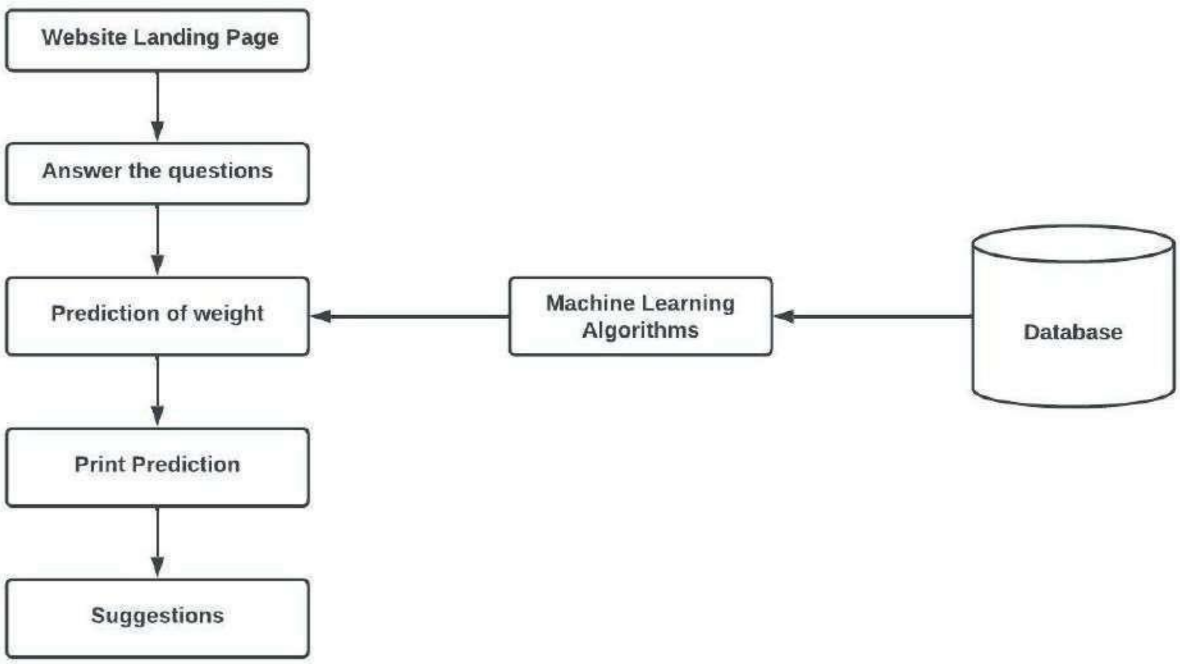


Fig:3.5.1 Proposed Method

3.6 SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS:

- System - Pentium-IV
- Speed - 2.4 GHZ
- Hard disk - 40 GB
- Monitor - 15 VGA color
- RAM - 512 MB

SOFTWARE REQUIREMENTS:

- Operating System - Windows 10
- Angular
- Coding Language - Python
- IDE - Jupyter Notebook/ Google collab

CHAPTER 4

IMPLEMENTATION AND ANALYSIS

4.1 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and the business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis, the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

The feasibility study investigates the problem and the information needs of the stakeholders. It seeks to determine the resources required to provide an information systems solution, the cost and benefits of such a solution, and the feasibility of such a solution.

The goal of the feasibility study is to consider alternative information systems solutions, evaluate their feasibility, and propose the alternative most suitable to the organization. The feasibility of a proposed solution is evaluated in terms of its components.

4.1.1 ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. Thus the developed system requires an IDE and a person for data collection. All the software used to develop the system is open source. This makes the system easily accessible for anyone with an internet connection and a smart gadget.

4.1.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is the Technical requirements of the system. Any system developed must not have a high demand for the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. Thus the developed system requires a browser to run the webpage or IDE such as Visual Studio Code to evaluate datasets, check the code, and collection of datasets for different environmental conditions requires a person to follow the instructions.

4.1.3 SOCIAL FEASIBILITY

The aspect of the study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. Pregnancy is a sensitive topic and if the accuracy of the model is not correct it might stir up unnecessary concerns. Every mother wants their child to be safe so they will easily adapt to this system.

4.2 DATASET DESCRIPTION

The Dataset is obtained from Vital Statistics Data on Births and related parameters from the UNC Odum Institute. We acknowledge the State Center for Health Statistics (SCHS) and the Howard W. Odum Institute for Research in Social Science at UNC at Chapel Hill as the source of data. The current dataset is part of a huge dataset of vital statistics details across the years. The dataset has a majority of records for the state of North Carolina and a significant number of outer state cases. It contains 101400 cases and similar data is available for the past 20 years across 125 variables. We used the following dataset for initial analysis and prediction. The following are some interesting variables observed in the dataset:

ID: Unique Identification number of a baby

SEX : Sex of the baby (1-boys, 2-girls)

MARITAL: Marital status of its parents

FAGE : Age of father

GAINED : Weight gained during pregnancy

VISITS : Number of prenatal visits

MAGE : Age of mother

FEDUC : Father's years of education

MEDUC : Mother's yearsof education

TOTALP : Total pregnancies

BDEAD : number of children born alive now dead

TERMS : Number of other terminations

LOUTCOME : Outcome of last delivery

WEEKS : Completed weeks of gestation

RACEMOM : Race of mother/child

HISPMOM : Hispanic

HISPDAD : Is dad hispanic?

CIGNUM : Average number of cigarettes used daily (Mother)

DRINKNUM: Average number of drinks used daily (mother)

ANEMIA : Mother has/had anemia

CARDIAC : Mother has/had cardiac disease

ACLUNG : Mother has/had acute or chronic lung disease

DIABETES : Mother has/had diabetes

HERPES : Mother has/had genital herpes

HYDRAM : Mother has/had hydramnios/Oligohydramnios

HEMOGLOB : Mother has/had hemoglobinopathy

HYPERCH : Mother has/had chronic hypertension

HYPERPR : mother has/had pregnancy hypertension

ECLAMP : Mother has/had Eclampsia

CERVIX : Mother has/had incompetent cervix

PINFANT : Mother had/had previous infant 4000+ grams

PRETERM : Mother has/had previous preterm/small infant

RENAL : Mother has/had renal disease

RHSEN : Mother has/had Rh sensitization

UTERINE : Mother has/had uterine bleeding

BWEIGHT : Baby's weight at birth

Understanding Relations Between Variables in the Dataset:

The following graph shows that the median weight of the babies according to the dataset is 7.375 pounds.

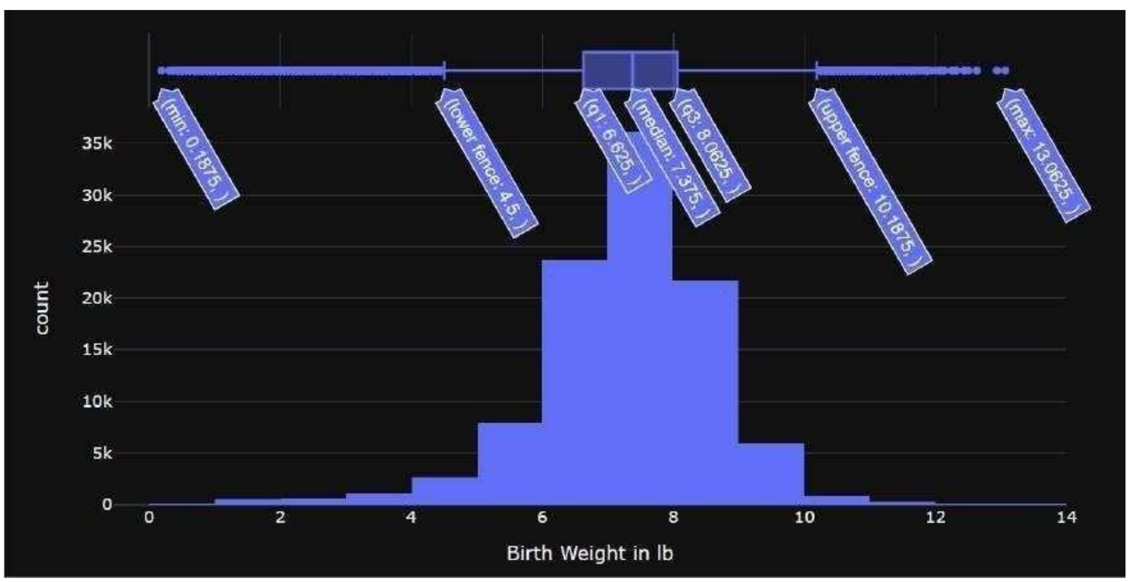


Fig:4.2.1 Birth Weight vs Number of Infants

Graph depicting how the gender of the infant affects the birth weight. The graph shows that boys have a slightly higher probability of being heavier than girl babies.

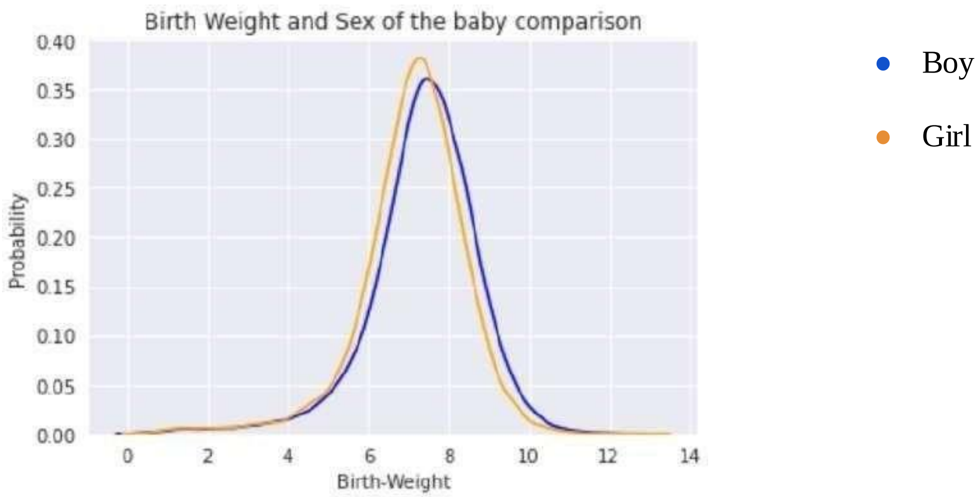


Fig:4.2.2 Birth Weight VS Sex

Graph showing the effect of smoking on birth weight. The graph shows us how Non- smokers have a higher probability of having heavier babies.

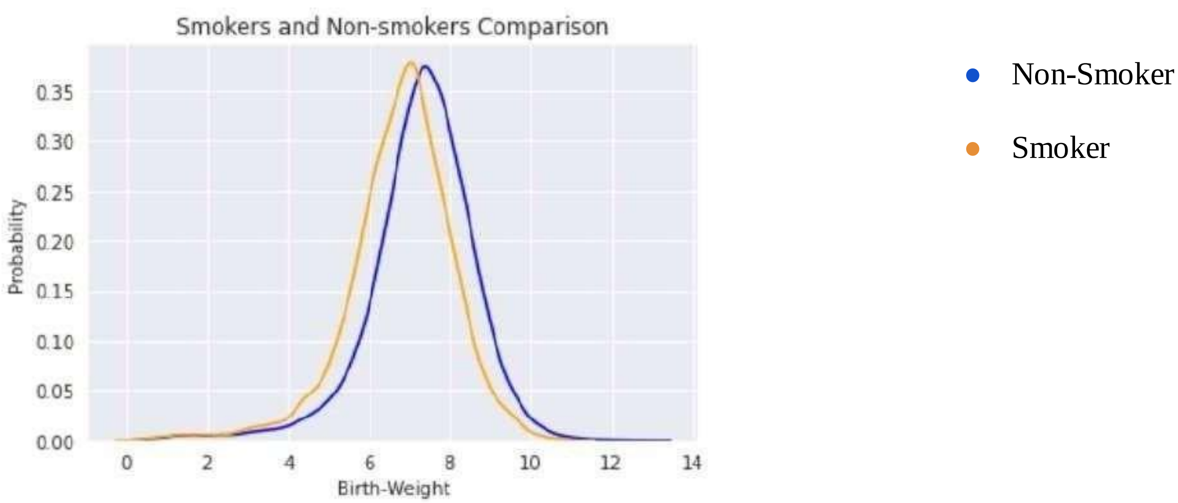


Fig:4.2.3 Birth Weight vs Smoking

Graph showing the effect of drinking on birth weight. The below plot confirms that the birth weight of the baby of a drinker mother tends to be less than that of a non-drinker mother.

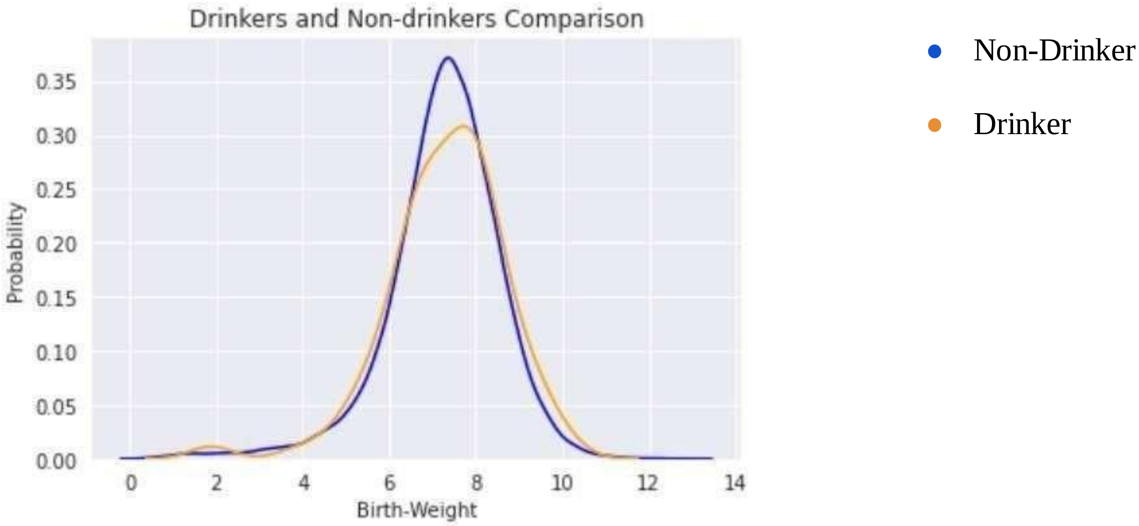


Fig:4.2.4 Birth Weight Vs Drinking

4.3 MODULE DESCRIPTION

Modules:

- Webpage
- Machine Learning

Implementation:

Uploading of Dataset: The huge dataset obtained in .csv file format. A small python code has been implemented in order to read the dataset to perform prediction.

Data Loading

```
[ ] from google.colab import drive
    drive.mount('/content/drive')

    Mounted at /content/drive
```

```
[ ] bweight_data = pd.read_csv("drive/MyDrive/Final Yr Project/birth-weight-prediction/baby-weights-dataset.csv")
```

Fig:4.3.1 Uploading Dataset

Preparing the Dataset: For predictions to be accurate we reduce the null values in the dataset. As the number of null values was comparatively low, the rows containing null values were just dropped.

```
bweight_data.isnull().sum().sum()

5
```

```
[ ] bweight_data = bweight_data.dropna(axis=0)
    bweight_data.isnull().sum()

ID          0
SEX         0
MARITAL     0
FAGE        0
GAINED      0
VISITS      0
MAGE        0
FEDUC       0
MEDUC       0
TOTALP      0
BDEAD       0
TERMS       0
LOUTCOME    0
WEEKS       0
RACEMOM     0
RACEDAD     0
HISPMOM     0
```

Fig:4.3.2 Preparing Dataset

The columns HISPMOM and HISPDAD are not used for predictions and thereby dropped.

```
dummy = pd.get_dummies(bweight_data,columns=['HISPMOM','HISPDAD'],drop_first=True)
```

Fig:4.3.3 Dropping Columns

Splitting Data: We need to split a dataset into train and test sets to evaluate how well our machine learning model performs. Here training dataset is 80% and testing dataset is 20%

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(train, balanced_targets, test_size=0.2)
```

Fig:4.3.4 Splitting Dataset

Evaluation: Training data is used in model training, or in other words, it's the data used to fit the model. On the contrary, test data is used to evaluate the performance or accuracy of the model. It's a sample of data used to make an unbiased evaluation of the final model fit on the training data.

4.4 RESULT

The Dataset is obtained from Vital Statistics Data on Births and related parameters from the UNC Odum Institute. We acknowledge the State Center for Health Statistics (SCHS) and the Howard W. Odum Institute for Research in Social Science at UNC at Chapel Hill as the source of data.

As seen in Previous Sections various machine learning algorithms and their accuracies were tested. The below graphs show their RMSE and R2 values.

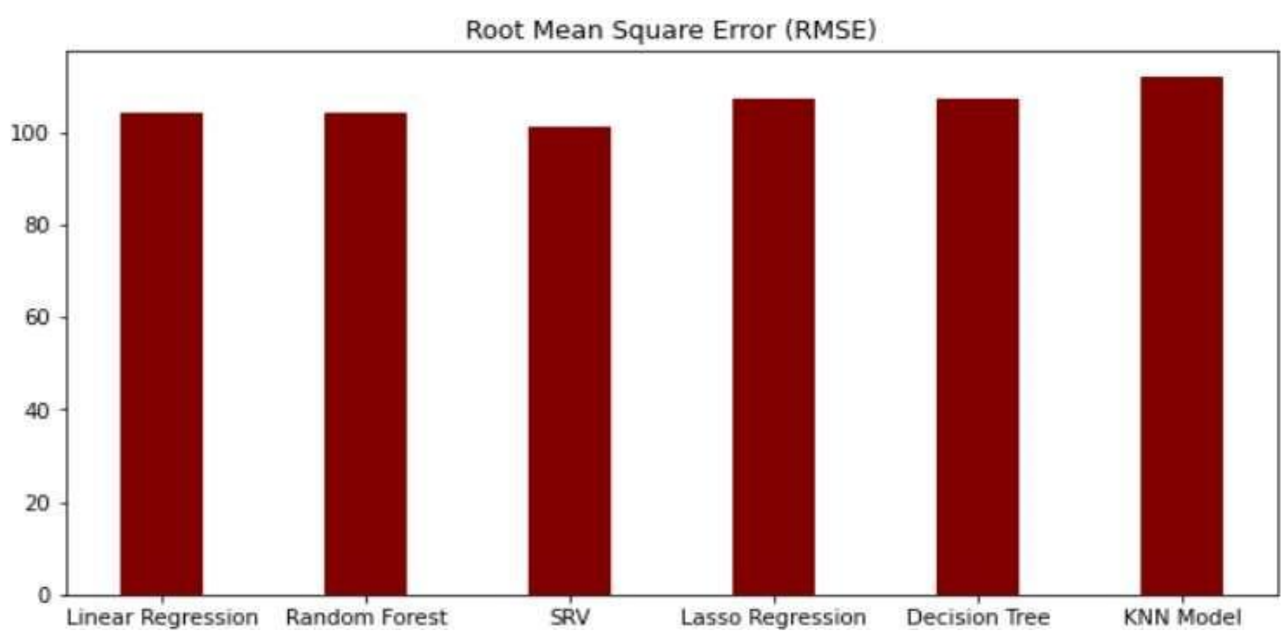


Fig:4.4.1 RMSE

One of these algorithms will be used to build the model for the proposed system by comparing the accuracies of root mean squared value and r-squared values.

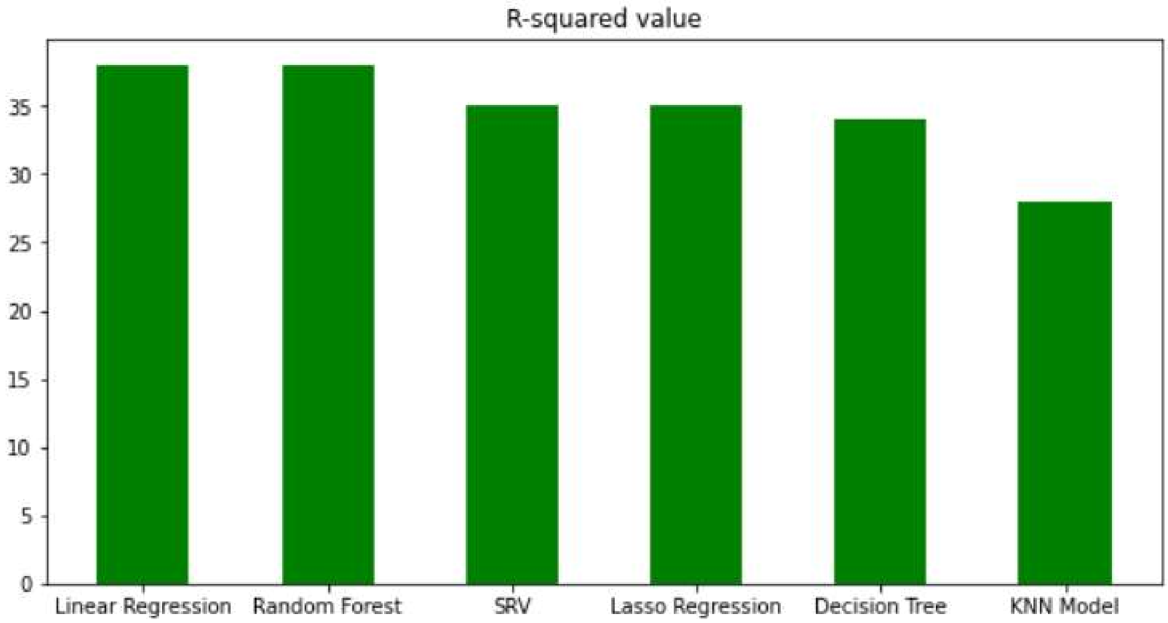


Fig:4.4.2 R-Squared Value

Upon inspection of the results, the SVR, random forest, and linear regression algorithms all had similar RMSE values. However, the R2 value was highest in both the linear regression and random forest models. Considering the general merits of the algorithms, the random forest regressor was chosen for the proposed system.

Overall, random forests offer a more flexible, robust, and accurate method for predicting outcomes than linear regression models. For these reasons, they are often preferable to linear regression when dealing with complex datasets.

Using a Random forest regressor to predict a real-time case:

```python
print("Enter the sex of the baby (1-boys, 2-girls)")
sex=int(input())
lst.append(sex)
print("Enter the marital status of the couple (1-single,2-married)")
marital=int(input())
lst.append(marital)
print("Enter the father's age ")
fage=int(input())
lst.append(fage)
print("Enter weight gained by mother during pregnancy")
gained=int(input())
lst.append(gained)
print("Enter number of prenatal vistis")
visits=int(input())
lst.append(visits)
print("Enter the mother's age ")
mage=int(input())
lst.append(mage)
print("Enter the Father's years of education")
fedu=int(input())
lst.append(fedu)
print("Enter the Mother's years of education")
medu=int(input())
lst.append(medu)
```

Fig 4.4.3 Questions to be filled

```
[35] print(nivi)

[1, 2, 35, 10, 15, 30, 18, 16, 1, 0, 0, 1, 37, 4, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

rfr.predict([nivi])

/usr/local/lib/python3.8/dist-packages/sklearn/base.py:450: UserWarning: X does not have valid feature names, bu
  warnings.warn(
array([6.80621604])
```

Fig 4.4.4 Prediction

In real time the baby's weight was 3.145 kg but the prediction showed us 3.087 so the difference is just 0.058 kg which makes the model highly efficient.

CHAPTER 5

CONCLUSION

5.1 CONCLUSION

This application will be able to predict birth weight within a minute. It will help in making prior arrangements for doctors and facilities before the baby's birth when needed. The system will be available all the time. This application is helpful mainly for hospitals and pregnant ladies. It is highly scalable and provides high performance and maintainability.

The prediction of birth weight is made based on many factors so that the algorithm will be able to give the best result. The web application will also provide suggestions to prevent low birth weight and help mothers take special care during pregnancy.

“Prevention is better than cure”, by predicting the birth weight beforehand we will be able to take the necessary precaution.

5.2 FUTURE WORK

At present, the machine learning model of the system is being developed. Therefore the next step of the proposed system will be to complete the model and develop a user-friendly web application. This webpage should be able to predict the birth weight of the infant from the data provided and provide suggestions when required.

APPENDIX A

SAMPLE

CODING

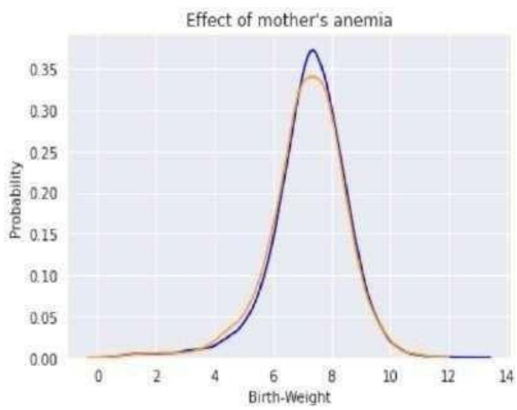
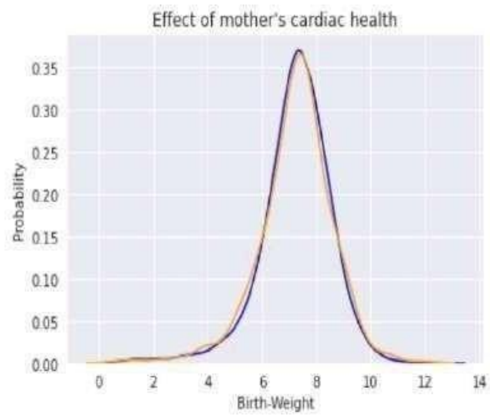
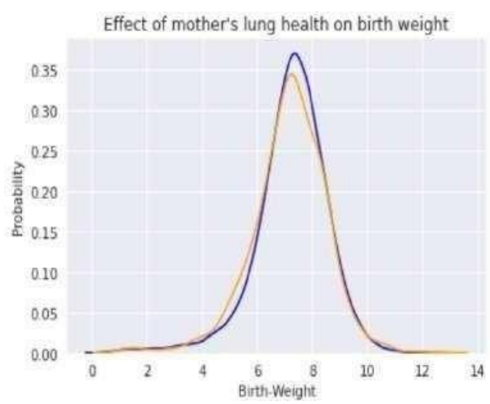
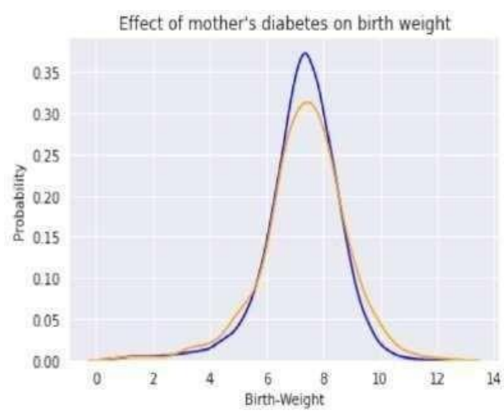
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats
from sklearn.model_selection import train_test_split
import plotly.graph_objects as go
import plotly.express as px

sns.set_style("darkgrid")
pd.options.plotting.backend = "plotly"
bweight_data = pd.read_csv("dataset.csv")
bweight_data.head()


bweight_data.BWEIGHT.describe()
fig=px.histogram(bweight_data, x =bweight_data["BWEIGHT"],nbins=20,
marginal="box",template = "plotly_dark",
                 labels={"BWEIGHT":"Birth Weight in lb"}, title =
"<b>Birth Weight And it count</b>")
fig.show()

sns.distplot(bweight_data['FAGE'], hist=False, color="blue")
sns.distplot(bweight_data['MAGE'], hist=False,color="orange")
plt.title('Father and Mother Age Comparison')
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.show()

sns.distplot(bweight_data[bweight_data['SEX'] == 1]['BWEIGHT'],
hist=False, color="blue",label="Boy")
sns.distplot(bweight_data[bweight_data['SEX'] == 2]['BWEIGHT'],
hist=False, color="orange",label="Girl")
plt.title('Birth Weight and Sex of the baby comparison')
```

```
plt.ylabel('Probability')
plt.xlabel('Birth-Weight')
plt.show()

target = bweight_data['BWEIGHT'].map(lambda x:1 if x >= 5.5 else 0)
inputs = bweight_data.drop(['BWEIGHT'], axis=1)

bweight_data.describe().T

indices_to_remove = []
balanced_inputs = inputs.drop(indices_to_remove, axis=0)
balanced_targets = target.drop(indices_to_remove,
axis=0)

#reset indices
reset_inputs = balanced_inputs.reset_index(drop=True)
reset_targets =
balanced_targets.reset_index(drop=True)

print("Inputs after balancing data:", reset_inputs.shape[0])
print("Targets after balancing data:",
reset_targets.shape[0])

balanced_inputs.head()

dummy =
pd.get_dummies(bweight_data,columns=['HISPMOM','HISPDAD'],drop_first=True)

T = dummy[dummy.columns[:]].corr()['BWEIGHT']
i = 0
train = pd.DataFrame()
for col in dummy:
    if(T[i] >= 0.04 or T[i] <=-0.04):
        t = dummy[col]
        train[col] = t
    i += 1
train = train.drop(columns = ['BWEIGHT'])

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(train,
balanced_targets, test_size=0.2)
```

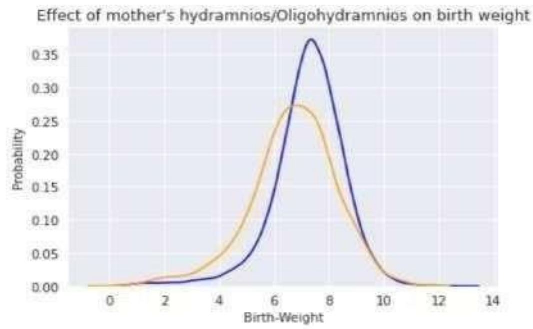
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
lr = LinearRegression()
lr.fit(X_train, y_train)

predicted = lr.predict(X_test)

RMSE = np.sqrt(mean_squared_error(y_test, predicted))
r2 = r2_score(y_test, predicted)

print('Root mean squared error: ', RMSE)
print("r2: ", r2)


from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn import metrics
from sklearn.metrics import roc_curve

logit = LogisticRegression()
logit.fit(X_train, y_train)

predicted_logit = logit.predict(X_test)

LogisticRegressionScore = accuracy_score(predicted_logit, y_test)


plt.figure()
metrics.plot_roc_curve(logit, X_test, y_test)
plt.title("Receiver Operating Characteristic (ROC)")
plt.show()

print("Logistic Regression score: ", LogisticRegressionScore)
```

APPENDIX B

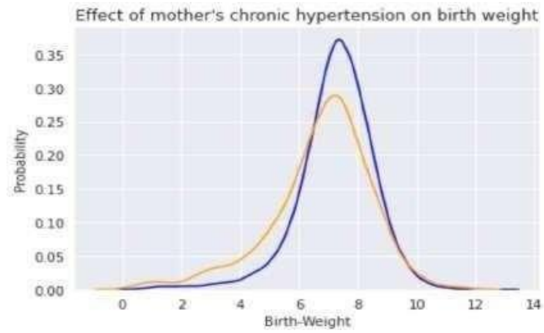
Sample Output

Effect of mother's anemia on birth weight	Effect of cardiac effect on birth weight
	
Effect of mother's lung disorder on birth weight	Effect of Diabetes on Birth Weight
	

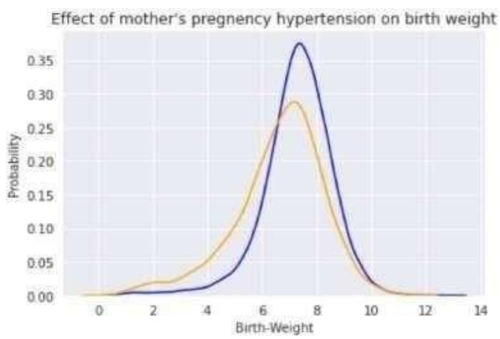
Effect of hydramnios/Oligohydramnios on Birth Weight



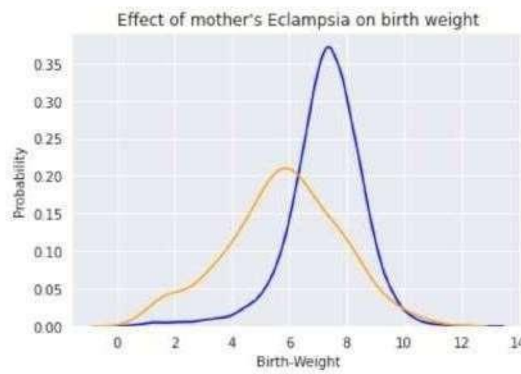
Effect of chronic hypertension on Birth Weight



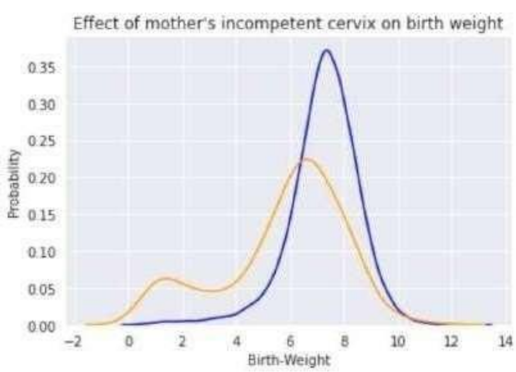
Effect of pregnancy hypertension on Birth Weight



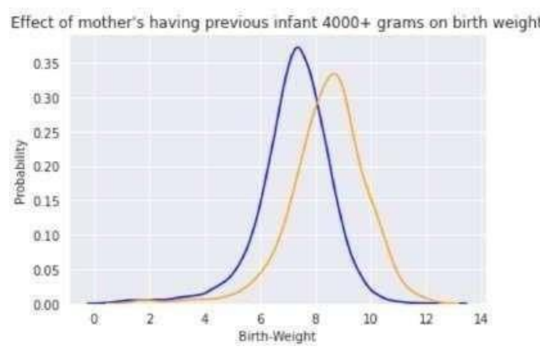
Effect of Eclampsia on Birth Weight



Effect of the incompetent cervix on Birth Weight



Effect of mother's having previous infant 4000+ grams on Birth Weight



REFERENCES:

[1] Kidanto HL, Mogren I, Lindmark G, Massawe S, Nystrom L. “Risks for preterm delivery and low birth weight are independently increased by the severity of maternal anemia”. S Afr Med J. 2009 Feb;99(2):98-102. PMID: 19418670.

[2] M. Abdollahian and N. Gunaratne, "Low Birth Weight Prediction Based on Maternal and Fetal Characteristics," 2015 12th International Conference on Information Technology - New Generations, 2015, pp. 646-650, DOI: 10.1109/ITNG.2015.108.

[3] Khan, W., Zaki, N., Masud, M.M. et al. “Infant birth weight estimation and low birth weight classification in the United Arab Emirates using machine learning algorithms”. Sci Rep 12, 12110 (2022).

[4]. N. S. Borson, M. R. Kabir, Z. Zamal and R. M. Rahman, "Correlation analysis of demographic factors on low birth weight and prediction modeling using machine learning techniques," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 169-173, DOI: 10.1109/WorldS450073.2020.9210338.

[5] Agarwal K, Agarwal A, Agrawal VK, Agrawal P, Chaudhary V. “Prevalence and determinants of low birth weight" among institutional deliveries. Ann Nigerian Med.2011;5(2):48-52.

[6] Abdullah Zahirzada, Kittichai Lavangnananda . “ Implementing Predictive Model for Low Birth Weigh” 2021 13th International Conference on Knowledge and Smart Technology (KST) DOI: 10.1109/KST51265.2021.9415792.

[7] Olli-Pekka Rinta-Koski, Simo Sarakka, Jaakko Hollmen,Markus Leskinen, Krista Rantakari, Sature Andersson . “ Prediction of Major complications affecting very low birth weight infants” Published in 2017 IEEE Life Sciences Conference (LSC) DOI: 10.1109/LSC.2017.8268174.

[8] Mario W. L. Moreira; Joel J. P. C. Rodrigues; Vasco Furtado; Constandinos X. Mavromoustakis; Neeraj Kumar; Isaac Woungang. “Fetal Birth Weight Estimation in High-Risk Pregnancies Through Machine Learning Techniques” Published in ICC 2019 - 2019 IEEE International Conference on Communications (ICC) DOI: 10.1109/ICC.2019.8761985.

[9]. Najmus Sakib Borson; Md. Riftabin Kabir; Zaisha Zamal; Rashedur M. Rahman. “Correlation analysis of demographic factors on low birth weight and prediction modeling using machine learning techniques” Published in 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) DOI: 10.1109/WorldS450073.2020.9210338.

[10.]Alfensi Faruk, E. Cahyono. “Prediction and Classification of Low Birth Weight Data Using Machine Learning Techniques” Published on 10 April 2018 DOI: 10.17509/JOST.V311.10799.

[11] R. Czabanski, M. Jezewski, J. Wrobel, T. Kupka, J. Łęski, J. Jezewski .” THE PREDICTION OF THE LOW FETAL BIRTH WEIGHT BASED ON QUANTITATIVE DESCRIPTION OF CARDIOTOCOGRAPHIC SIGNALS” Published 2008 Journal of Medical Informatics and Technologies Corpus ID: 41953394.

[12] Uzapi Hange, Rajalakshmi Selvaraj, Malatsi Galani, Keletso Letsholo. “**A** Data-Mining Model for Predicting Low Birth Weight with a High AUC” Published in 2018, 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017.

[13] Ghalib Ahmed Tahir, Tooba Samad, Zongying Liu, Sundus Abra r, Murtaza Ashraf, Hammad Qureshi. **“**Prediction of Child Birth Weight Using Kernel Extreme Reservoir Machine and QPSO for Optimization” Published in May 2021SN Computer Science

[14] Mehri rejali , Marjan Mansourian, Zohre Babaei, Babak Eshrati. “Prediction of Low Birth Weight Delivery by Maternal Status and Its Validation: Decision Curve Analysis” Published in 2017 Jul 25

[15] Yang Ren, Dezhi Wu, Ana Lopez-De Fede. “Identification and Prediction of Low-Birthweight Baby Outcomes and Mom Risk Factors” Conference: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)