Low Body Weight Prediction of Newborn Baby using Machine Learning Algorithms

Sreeja Vijayakumar
Department of Computer Applications
Amal Jyothi College of Engineering,
Kanjirappally, India
sreejavijayakumar@mca.ajce.in

Meera Rose Mathew
Department of Computer Application
Amal Jyothi College of Engineering,
Kanjirappally, India
meerarosemathew@amaljyothi.ac.in

***Abstract* — Babies who are born weighing less than 2,500 grams are referred to as having a low birth weight .A average newborn typically weighs around 8 pounds. A suitable medical intervention plan can be implemented if a case of possible LBW is discovered early in the pregnancy based on maternal variables in order to ultimately prevent the LBW. Low birth weight is the one factor that most significantly affects an infant's likelihood of surviving (LBW).The goal of this analysis were to identify factors that influence low birth weight in newborns, increase awareness among expectant moms, and develop a machine learning algorithm to predict low birth weight kids. The purpose of this study was to compare and choose the most suitable model for LBW prediction.**

Keywords— Machine Learning , Logistic Regression, SVM, Decision tree , Random forest.

I. INTRODUCTION

Lower body weight refers to a newborn who weighs less than 2500 grams, at birth. Although LBW infants are more likely to experience health issues, not all of them will require interventions or special care, especially those who are near to the 5 pound weight cut-off.

Main reasons

- Babies were born prematurely or at the proper time but did not develop properly during pregnancy
- Prematurity, preeclampsia, drug or alcohol misuse, repeated pregnancies, inadequate nutrition during pregnancy, and illness in the mother or the foetus before birth.

Since low birth weight makes newborns' bodies more brittle, it makes it more difficult for them to feed, breathe, grow, maintain their internal temperature, or fight infections. Given that low birth weight can cause a variety of illnesses in children, it is important to consider the maternal, nutritional, and socioeconomic factors while evaluating low birth weight.Early identification or prediction of LBW baby output during pregnancy can help to stop LBW or raise the possibility that LBW infants will survive by giving the mother extra prenatal care. Here, LBW is predicted using a variety of machine learning algorithms. Gradient boosting classifier, support vector classifier, decision tree classifier, K nearest neighbor, and logistic regression were used as machine learning approaches. Decision Tree Classifier is best performer as it has a test accuracy of 0.8682, sensitivity 0.9612.Data analysis, methodology description, results acquired by using various models and conclusion are steps followed in the paper. We verify the most accurate algorithm and value.

II. LITERATURE REVIEW

MAC Akmal Jahan , AM Razmy [1] based on maternal characteristics, a system predicts the incidence of low birth weight in babies. For this classification task, a variety of classifiers including Random Tree, C4.5, Random Forest, Decision Stump, Logistic Model Tree, REP Tree, and BF (best-first) tree are evaluated.

Wondesen teshome Bekele [2] by using data from the Ethiopia Demographic and Health Survey 2016 as a foundation, this study constructed predictive LBW models.The models used completely different other classifiers.

Shashi Nandar Kumar, Pallavi Saxena, Rachana Patel, Arun Sharma, [3] In this study, a thorough effort was undertaken to predict LBW offspring early on using maternal sociodemographic characteristics and PAH concentration. Both traditional regression and machine learning techniques were used to make the prediction. Both approaches revealed that work, cooking habits, and pregnancy weight gain are better indicators of LBW. Additionally, the research reveals notable PAHs that distinguish LBW children.

Machine learning algorithms were created by Anisha R. Yarlapati, Sudeepa Roy Dey, and Snehanshu Saha[4] in order to gather important data from health indicators of pregnant women for the early identification of suspected LBW cases. By employing the Bayes' minimal error rate classifier to divide the dataset into two groups, they transform the LBW detection problem into a binary machine classification problem. The proposed model had a 96.77% accuracy rate.

Md Menhazul Abedin ,S M Ashikul Islam Pollob , Md Touhidul Islam , Md Merajul Islam , Md Maniruzzaman [5]. They use data from the Bangladesh Demographic and Health Survey, 2017–18, which included 2351 respondents, on low birth weight. To define and forecast low birth weight, two machine learning-based classifiers were used. Their findings show that Bangladesh needs an effective, affordable, and complimentary complementary technique to reduce and accurately forecast low birth weight newborns.

III .MACHINE LERANING ALGORITHMS

Automated analytical model building is accomplished by the use of machine learning, a technique for data analysis. It is a part of AI that was developed on the premise that computers

DOI: 10.5281/zenodo.7962143
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

can see patterns in data, learn from it, and make decisions mostly independently of human input. Machine learning is a rapidly developing trend in the healthcare industry, as seen by the emergence of wearable computing and sensors that may use data to analyse a patient's health in real time. Machine learning technology can help medical practitioners analyse data and look for patterns or warning indications that could point to a more precise diagnosis and course of therapy.

Logistic Retrogression,it is the go-to method for binary classification problems. It is used to calculate double values like 0/1 and discriminational values.It uses data to relate to a logit function to determine the culpability of an event. Therefore, another name for it is logit regression. It is characterised as a probability forecasting model because it predicts probabilities that have values between 0 and 1.

The extreme vectors and points that help create the hyperplane are chosen via SVM. The SVM approach is based on support vectors, which are utilised to represent these extreme situations.

For classification and regression, a decision tree is a non-parametric supervised learning method. The aim is to learn simple decision rules produced from the data attributes to build a model that predicts the value of a target variable.

The supervised machine learning method is used to grow and combine different decision trees to form the Random Forest. A large number of simple decision trees are constructed during training using the random forest approach, and a majority vote (mode) across them is used to classify data.

III METHODOLOGY

Using a ML model, the data from collected dataset should move through a series of processes in order to identify LBW babies. It takes time for it to blend into the algorithm's single silent and filtered input.

1) Data Collection
2) Data Analysing
3) Data Cleaning
4) Feature Selection
5) Data Visualization
6) Model fitting and generating Predictions
7) Evaluvation of models
8) Selection of best model

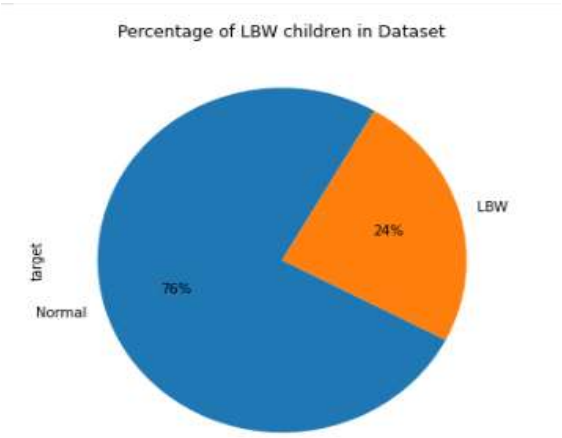
a)Data Collection

Data is collected from different platforms. It contains samples,that are collected from various hospitals in Kerala, India. Dataset conatins various features that affect LBW.

b) Data Analysing

This dataset consists of 75 features. All are numeric variables. Data analysing based on the value of the target. Analysis is reliant on Bad obsteric history, Workload,Injection details, Iron tonic or not, Convulsion, Economic status Asthma, fever, Mother's age, Mother's weight, Fundal height, Mother's occupation

Figure.1



c) Data Cleaning

Errors in the dataset are found and fixed through the process of data cleaning. A prediction model may suffer if errors are present. To find and fix data issues, a variety of actions and activities are performed using the data cleaning process.

c) Feature Selection

In the dataset have 74 features , to enhance the model's functionality and reduce computational costs we flittering out the unnecessary columns & store the new dataset .The following factors have come as significant relationship with the goal

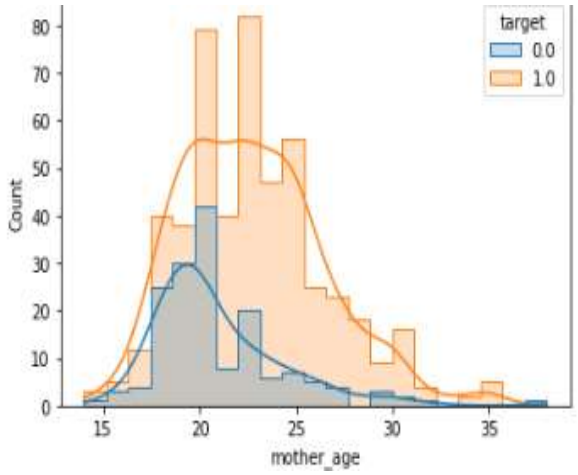
Features
Parity Or Order Of Gravida
Target
Mother's Age
Mother's Height
Workload
Asthma
Bad Obsteric History
Injection
Mid-Arm Circumference
Iron
Bleed
Fever
Body Mass

The target variable, or logit, is the one we need to forecast. One indicates a low birth rate while zero indicates a normal birth rate.

DOI: 10.5281/zenodo.7962143
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Some features Analysis

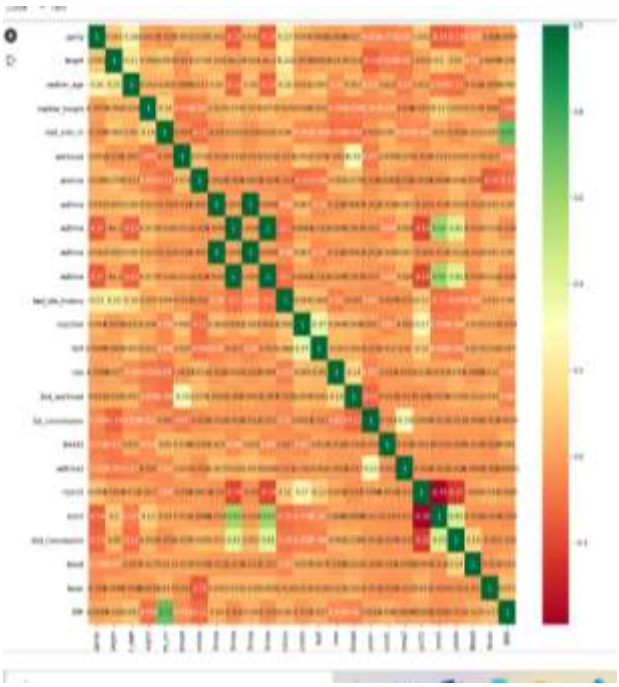
Figure.2



e) Data Visualization

Data is visualized using different graphs and identify the corelation between significant features.Data visualization helps to identify features nature.

Figure.3



f) Fitting in to model and Making Predictions

Here we uses;
Mathew Correlation coefficient Its purpose is to evaluate or quantify the difference between the expected values and actual values, which is equivalent to chi-square statistics for a 2 x 2 contingency table.

Log Loss- A classification model's performance is measured by logarithmic loss when the prediction input is a probability value between 0 and 1. The log loss in a perfect model would be 0.
F1 Score = 2*(Recall* Precision) / (Recall + Precision)

A. Model building and evaluation

I.RandomForest

Accuracy	0.868263
Precision	0.85517241
Sensitivity	0.96124031
Specificity	0.4473684
F1 Score	0.9051094
ROC	0.7043043
Log_Loss	5.3773950
mathew_corrcoef	0.5065227

Figure.4

II.Decision Tree

Accuracy	0.8622754
Precision	0.9140625
Sensitivity	0.90697674
Specificity	0.710526315
F1 Score	0.91050583
ROC	0.80875152
Log_Loss	4.75689013
mathew_corrcoef	0.611911

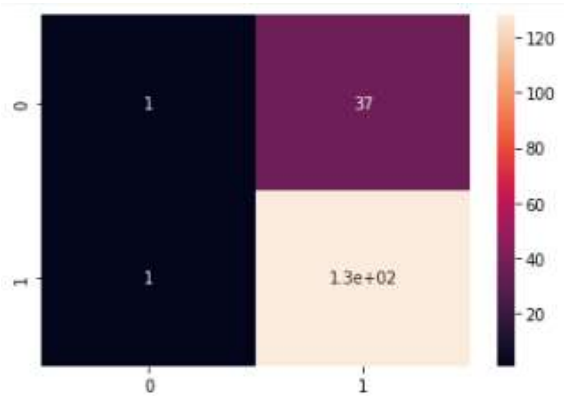
Figure.5

DOI: 10.5281/zenodo.7962143
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

III. Logistic Regression

Accuracy	0.7724550
Precision	0.7757575
Sensitivity	0.992248062
Specificity	0.02631578
F1 Score	0.8707482
ROC	0.50928192
Log_Loss	7.85929992
mathew_corrcoef	0.0715480

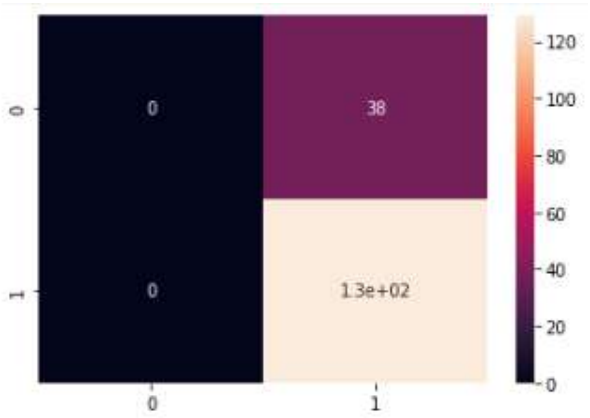
Figure.6



IV. SVM

Accuracy	0.772455
Precision	0.7724550
Sensitivity	1.0
Specificity	0.0
F1 Score	0.871621
ROC	0.5
Log_Loss	7.8593047
mathew_corrcoef	0.0

Figure.6

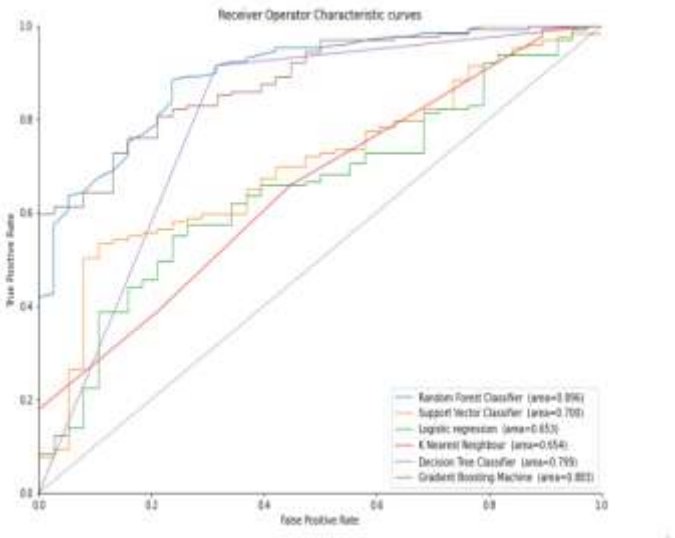


VI.ACCURACY OF DIFFERENT MODELS

Models	**Accuracy**
Logistic regression	0.772455
RandomForest	0.868263
SVC	0.7724550
Decision Tree	0.8622754

RandomForest Classifier has the highest accuracy.

Figure.7



V.CONCLUSION

In this study, we used a variety of machine learning techniques to predict low body weights in newborns based on the collected data.Using several machine learning methods, we were able to develop various classifier models.Among all the models , the Random Forest had the highest accuracy.But we analyse the models on the basis of accuracy,sensivity,f1score and log loss.Therefore, Decision Tree Classifier is best performer as it has highest test accuracy of 0.8682, sensitivity 0.9612 and highest f1-score of 0.9185 and lowest Log Loss of 4.5501.So we successfully detemined the best model for prediction of LBW.

DOI: 10.5281/zenodo.7962143
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam

REFERENCES

[1] Decision tree based automated prediction of infant low birth weight (2020) by MAC Akmal Jahan and AM Razmy. Faculty of Applied Sciences, South Eastern University of Sri Lanka.

[2] Bekele, W.T. Machine learning algorithms for predicting low birth weight in Ethiopia. BMC Med Inform Decis Mak 22, 232 (2022). https://doi.org/10.1186/s12911-022-01981-9

[3] Shashi Nandar Kumar, Pallavi Saxena, Rachana Patel, Arun Sharma. Predicting risk of low birth weight offspring from maternal features and blood polycyclic aromatic hydrocarbonconcentration(2020). ISSN 0890-6238,https://doi.org/10.1016/j.reprotox.2020.03.009.

[4] Yarlapati, Anisha R.; Roy Dey, Sudeepa; Saha, Snehanshu (2017). Early Prediction of LBW Cases via Minimum Error Rate classifier: A Statistical Machine Learning Approach.

[5] Predicting risks of low birth weight in Bangladesh with machine learning by S M Ashikul Islam Pollob , Md Menhazul Abedin , Md Touhidul Islam , Md Merajul Islam , Md Maniruzzaman Published: May 26, 2022https://doi.org/10.1371/journal.pone.0267190

DOI: 10.5281/zenodo.7962143
ISBN: 978-93-5906-046-0@2023, Dept. of Computer Applications, Amal Jyothi College of Engineering Kanjirappally, Kottayam