Predicting Low Birth Weight (LBW) Cases in Early Pregnancy using Machine Learning Approaches Top of Form

Rallapalle Kavitha

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— LBW, indicating newborn health issues, is linked to infant mortality and long-term health concerns. This study utilizes machine learning to detect potential LBW cases early by analyzing maternal health indicators. It frames the problem as a binary classification task and achieves improved accuracy. Decision rules derived from Indian healthcare data aid in predictive healthcare for smart cities, with a screening tool developed for Obstetrics and Gynecology professionals.

Keywords: Low Birth weight (LBW), Smart health informatics, Predictive analytics, Machine Learning (ML).

I.   INTRODUCTION

The WHO Maternal Health Program 1992 highlights the rising concern of Low Birth Weight (LBW), expected to increase by 12% annually. LBW contributes to neonatal deaths globally, with significant health risks for affected children. LBW, defined as babies under 2500g, poses challenges worldwide, particularly in developing countries like India. Maternal health strongly influences birth weight, emphasizing the importance of early detection and intervention to mitigate LBW risks.

II.   LITERATURE REVIEW

2.1   Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. Bull World Health Organ. 1987; 65(5):663-737. PMID: 3322602; PMCID: PMC2491072.

The existence and magnitude of a causal effect on birth weight, gestational age, and prematurity and intrauterine growth retardation were determined by a set of methodological standards. In developed countries, the most important factor was cigarette smoking, followed by nutrition and pre-pregnancy weight. In developing countries, the major determinants were racial origin, nutrition, low pre-pregnancy weight, short maternal stature, and malaria. Pre-pregnancy weight, prior premature birth or miscarriage, diethylstilbestrol exposure and smoking were major determinants of gestational duration, but the majority of prematurity was unexplained in both developed and developing countries.

2.2   Vega J, Sáez G, Smith M, Agurto M, Morris NM. Factores de riesgo para bajo peso al nacer y retardo de crecimientointrauterino en Santiago de Chile [Risk factors for low birth weight and intrauterine growth retardation in Santiago, Chile]. Rev Med Chil. 1993 Oct; 121(10):1210-9. Spanish. PMID: 8191127.

An epidemiologic case-control study to ascertain the determinants of low birthweight was carried out in Santiago, Chile, from January to December 1989. The cases were defined as livebirths < 2500 g. The controls were livebirths > or = 2500 g of birthweight. All cases and a random sample (1:1) of controls were selected among 8,254 singleton births occurring at the El Salvador Hospital in the Eastern area of Santiago. These deliveries represented 50% of institutional deliveries in the area. Home deliveries (2%) and private hospital deliveries were not included in the study. Information was obtained from hospital medical records by six trained medical students. Some information could not be obtained from the hospital medical records. Thus, the second step in data collection was the tracking of all the selected subjects to their referring neighborhood health centers.

2.3   Mavalankar DV, Trivedi CC, Gray RH. Maternal weight, height and risk of poor pregnancy outcome in Ahmedabad, India. Indian Pediatr. 1994 Oct;31(10):1205-12. PMID: 7875780.

This paper explores the relationships between maternal weight, height and poor pregnancy outcome using a data set from a case-control study of low birth weight (LBW) and perinatal mortality in Ahmedabad, India. Maternal height and weights were compared between mothers of 611 perinatal deaths, 644 preterm-LBW, and 1465 normal birth weight controls as well as 617 small-for-gestational age (SGA) and 1851 appropriate-for-gestational-age (AGA) births. Weight and height were much lower in this population compared to western standards. Low weight and height were associated with increased risk of perinatal death, prematurity and SGA. After adjusting for confounders, maternal weight remained significantly associated with poor pregnancy outcomes, whereas height was only weakly associated. Attributable risk estimates show that low weight is a much more

important contributor to poor outcome than low height. Improvement in maternal nutritional status could lead to substantial improvement in birth outcome in this population

2.4   Bosetti C, Nieuwenhuijsen MJ, Gallus S, Cipriani S, La Vecchia C, Parazzini F. Ambient particulate matter and preterm birth or birth weight: a review of the literature. Arch Toxicol. 2010 Jun;84(6):447-60. doi: 10.1007/s00204-010-0514-z. Epub 2010 Feb 6. PMID: 20140425.

To review epidemiologic evidence on maternal exposure to particulate matter and adverse pregnancy outcomes, we performed a MEDLINE search of the literature up to June 2009. We considered all original studies published in English including information on total suspended particles (TSP), respirable (PM (10)) or fine (PM(2.5)) particles and the risk of preterm birth, low birth weight (LBW) or very low birth weight (VLBW) and small for gestational age (SGA). We identified a total of 30 papers, including 13 with information on preterm birth, 17 on LBW or VLBW, and 4 on SGA. Eight studies on preterm birth, 11 studies on LBW/VLBW and two studies on SGA reported some increased risk (by about 10-20%) in relation to exposure to PM; no meaningful associations were found in the remaining studies. However, even in studies reporting some excess risk, this was inconsistent across exposure levels and pregnancy periods. Epidemiologic studies on maternal exposure to PM during pregnancy thus do not provide convincing evidence of an association with the risk of preterm birth and LBW/VLBW and SGA. The excess risks, if any, are small, and it is unclear whether they are causal, due to misclassification of the exposure or some sources of bias/residual confounding.

III.   PROPOSED WORK

In proposed system, we implement supervised machine learning algorithms like XGBoost Classifier, Random Forest Classifier and Support Vector Classifier and Decision Tree Classifier for prediction of low-Birth-Weight babies.
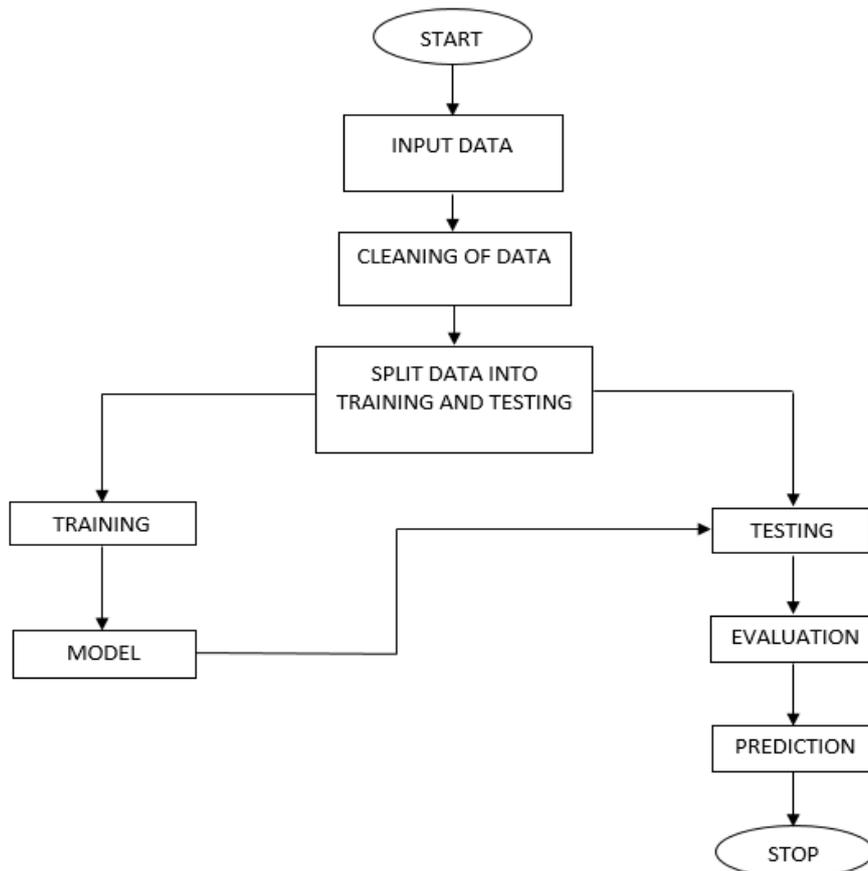
ADVANTAGES:

- High accuracy.
- Time Saving.

- Does not require highly trained staff.

- High reliability.

- Low complexities.

3.1   XGBoost

XGBoost is an algorithm that has recently been dominating applied machine learning and Kaggle competitions for structured or tabular data. XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree-based algorithms are considered best-in-class right now.

Bagging: Now imagine instead of a single interviewer, now there is an interview panel where each interviewer has a vote. Bagging or bootstrap aggregating involves combining inputs from all interviewers for the final decision through a democratic voting process.

XGBoost and Gradient Boosting Machines (GBMs) are both ensemble tree methods that apply the principle of boosting weak learners (CARTs generally) using the gradient descent architecture. However, XGBoost improves upon the base GBM framework through systems optimization and algorithmic enhancements.

3.2   Random Forest:

First, Random Forest algorithm is a supervised classification algorithm. We can see it from its name, which is to create a forest by some way and make it random. There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the result. But one thing to note is that creating the forest is not the same as constructing the decision with information gain or gain index approach.

The author gives four advantages to illustrate why we use Random Forest algorithm. The one mentioned repeatedly by the author is that it can be used for both classification and regression tasks. Overfitting is one critical problem that may make the results worse, but for Random Forest algorithm, if there are enough trees in the forest, the classifier won't overfit the model. The third advantage is the classifier of Random Forest can handle missing values, and the last advantage is that the Random Forest classifier can be modeled for categorical values.

There are two stages in Random Forest algorithm, one is random forest creation, the other is to make a prediction from the random forest classifier created in the first stage.

3.3   Decision Trees:

Trees in machine learning, like decision trees, play a significant role in both classification and regression tasks. Decision trees visually represent decisions and are commonly used in data mining. They split based on conditions and lead to outcomes, making complex data understandable. This method, known as learning decision trees, is essential for tasks like classification and regression. CART algorithms handle both types of trees, making predictions and simplifying complex data structures

3.4   Support Vector Machine:

SVM, a supervised learning model, excels in classifying two-group problems. It's effective for text classification, offering reliability with limited data. Despite terms like 'linearly separable' and 'kernel trick,' SVM's concept is straightforward, making it suitable for natural language tasks
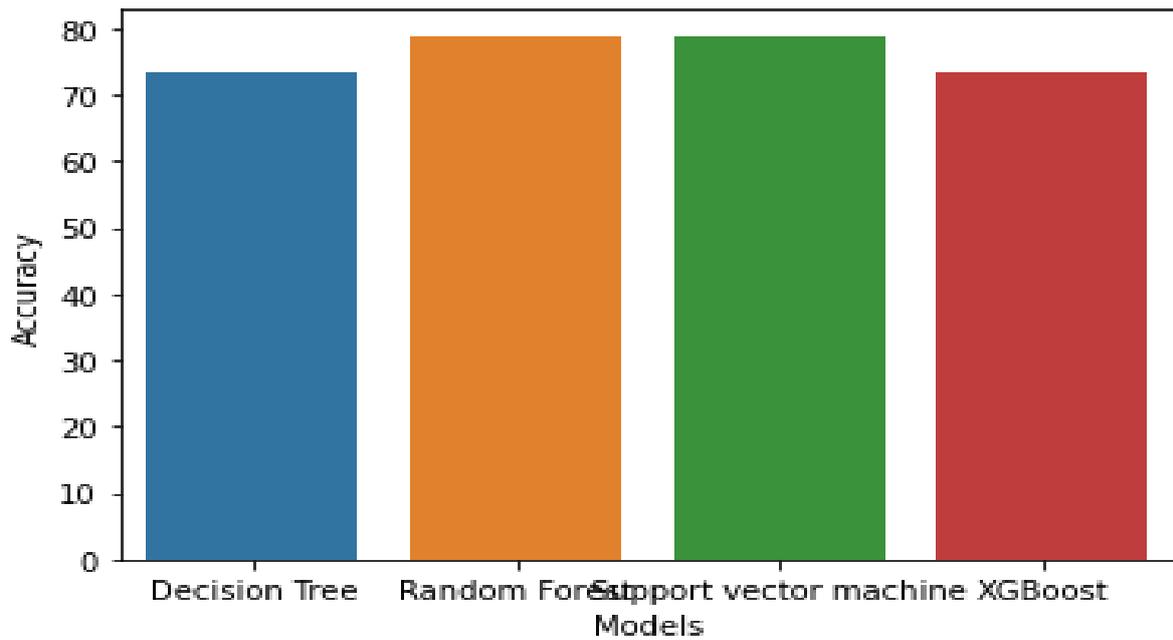
IV.   IMPLEMENTATION

- Import the dataset.

- Explore the data to figure out what they look like.

- Pre-process the data.

- Split the data into attributes and labels.

- Divide the data into training and testing sets.

- Train the Model

- Make some predictions

	Models	Accuracy
0	Decision Tree	73.684211
1	Random Forest	78.947368
2	Support Vector Machine	78.947368
3	XGBoost	73.684211

V.   CONCLUSION

We've built an ML model using Flask in Python to predict Low Birth Weight. Among XGBoost, Random Forest, Decision Tree, and Support Vector Classifiers, Decision Tree Classifier achieved the best accuracy

REFERENCES

[1] World Health Organization-1992, International statistical classification of diseases and related health problems, Tenth revision, Geneva, World Health Organization.

[2] Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. Bull World Health Organ. 1987; 65(5):663-737. PMID: 3322602; PMCID: PMC2491072.

[3] Vega J, Sáez G, Smith M, Agurto M, Morris NM. Factores de riesgo para bajo peso al nacer y retardo de crecimientointrauterino en Santiago de Chile [Risk factors for low birth weight and intrauterine growth retardation in Santiago, Chile]. Rev Med Chil. 1993 Oct; 121(10):1210-9. Spanish. PMID: 8191127.

[4] Mavalankar DV, Trivedi CC, Gray RH. Maternal weight, height and risk of poor pregnancy outcome in Ahmedabad, India. Indian Pediatr. 1994 Oct; 31(10):1205-12. PMID: 7875780.

[5] Bosetti C, Nieuwenhuijsen MJ, Gallus S, Cipriani S, La Vecchia C, Parazzini F. Ambient particulate matter and preterm birth or birth weight: a review of the literature. Arch Toxicol. 2010 Jun;84(6):447-60. Doi: 10.1007/s00204-010-0514-z. Epub 2010 Feb 6. PMID: 20140425.

[6] United nations Children's Fund and World Health Organization2004, Low Birth Weight: Country, regional and global estimates, New York, UNICEF.

[7] J.S. Deshmukh, D.D. Motghare, S.P. Zodpey and S.K. Wadhva1998, Low Birth Weight and Associated Maternal Factors in an Urban Area, Indian Pediatrics, Volume 35, Page 33-36.

[8] Aparajita Dasgupta, RivuBasu, "Determinants of low birth weight in a Blowk of Hooghly, West Bengal: A multivariate analysis," International Journal of Biological & Medical Research, 2(4),2011, pp.838-842.

[9] Bellazzi R, Zupan B., "Towards knowledge-based gene expression data mining", J Biomed Inform. 2007;40(6), pp.787-802.