# Early Prediction of Low Birth Weight (LBW) Cases Using ML Approaches

Nedambaram Sai Divya

Dept of Computer Science, Sri Venkateswara University, Tirupati

**Abstract—** *Low Birth weight (LBW) acts as an indicator of sickness in newborn babies. LBW is closely associated with infant mortality as well as various health outcomes later in life. Various studies show strong correlation between maternal health during pregnancy and the child's birth weight. This manuscript exploits machine learning techniques to gain useful information from health indicators of pregnant women for early detection of potential LBW cases. The forecasting problem has been reformulated as a classification problem between LBW and NOT-LBW classes using supervised Machine learning for LBW detection as a binary machine classification problem. Expectedly, the proposed model achieved better accuracy. Indian health care data was used to construct decision rules to be extrapolated to predictive health care in smart cities. A screening tool based on the decision model is developed to assist health care professionals in Obstetrics and Gynecology (OBG).*

## I. INTRODUCTION

World Health Organization Maternal Health and Safe Motherhood Programme-1992, Low Birth Weight. It is expected to rise at the rate of 12% every year. Nearly 39% of power is used for cooling 45% for running the Information Technology (IT), infrastructure and 13% for lights. This level of consumption costs heavily to the businesses. LBW and prematurity remain a serious public health burden worldwide. Neonatal deaths account for a major fraction of deaths of children under the age of five, globally Children with LBW are at significantly higher risks of early childhood morbidity and mortality when compared with their counterparts with normal birth weights.

Low birth weight is the term used to refer to babies born with a weight less than 2500gm Low birth weight (LBW) has been identified as a major public health problem around the world. LBW includes both pre-term babies as well as fully grown babies who are very small in size as a consequence of intra uterine growth retardation. Birth weight is closely associated with neonatal and infant mortality, mortality rates being significantly higher in LBW babies when compared to the normal birth weight (NBW) babies. This phenomenon is now of global concern in the view of serious short term and long-term problems such as development disorders, neurosensory outcomes, health outcomes including Type 2 diabetes, cerebral stroke, hypertension and various other disorders that LBW babies are prone to Studies in 2013 showed that out of the 22 million newborns about 16 percent were low birth weight cases globally. This is a major problem in developing countries, especially in India which contributes to about 30 percent of the global LBW cases.

Innumerable studies around the world indicate strong between maternal health and impact onbirth weight of babies. Popular assumptions claim that LBW can be considerably reduced, with dedicated medical care during pregnancy. In our approach, the risk factors in pregnant women that can be easily assessed with basic methods are carefully examined throughout the gestation period and form the basis for predictions. Early detection can help in preventing the chances of LBW and also to put forward some recommendations under some intervention mechanisms.

## II. LITERATURE REVIEW

**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.

**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.

**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.**

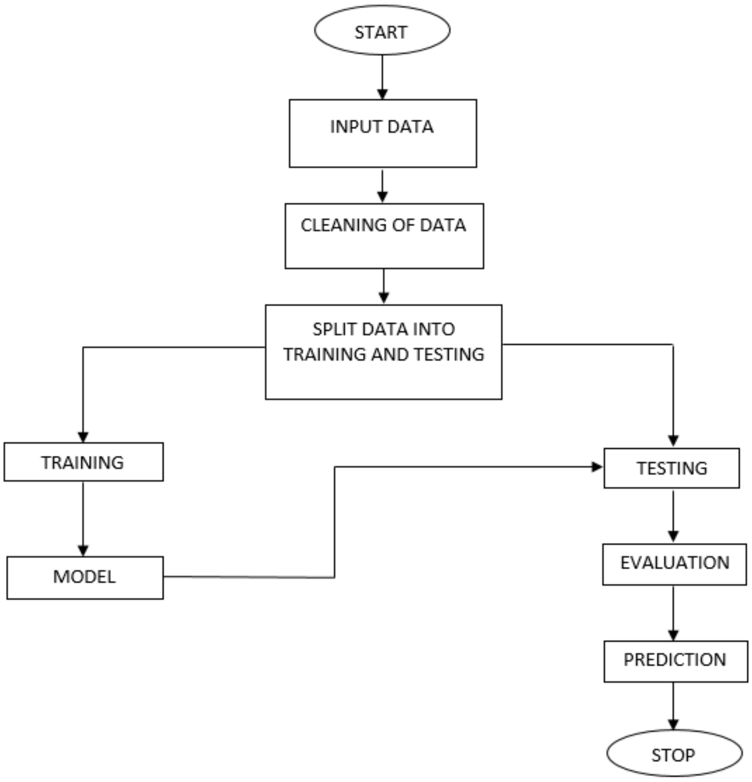
This paper explores the relationships between maternal weight, height and poor pregnancy outcome using a data set from a case-control study of low birth weight (LBW) and perinatal mortality in Ahmedabad, India. Maternal height and weights were compared between mothers of 611 perinatal deaths, 644 preterm-LBW, 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

**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:

A tree has many analogies in real life, and turns out that it has influenced a wide area of machine learning, covering both classification and regression. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in data mining for deriving a strategy to reach a particular goal. A decision tree is drawn upside down with its root at the top. In the image on the left, the bold text in black represents a condition/internal node, based on which the tree splits into branches/ edges. The end of the branch that doesn't split anymore is the decision/leaf, in this case, whether the passenger died or survived, represented as red and green text respectively. Although, a real dataset will have a lot more features and this will just be a branch in a much bigger tree, but you can't ignore the simplicity of this algorithm. The feature importance is clear and relations can be viewed easily. This methodology is more commonly known as learning decision tree from data and above tree is called Classification tree as the target is to classify passenger as survived or died. Regression trees are represented in the same manner, just they predict continuous values like price of a house. In general, Decision Tree algorithms are referred to as CART or Classification and Regression Trees. So, what is actually going on in the background? Growing a tree involves deciding on which features to choose and what conditions to use for splitting, along with knowing when to stop. As a tree generally grows arbitrarily, you will need to trim it down for it to look beautiful. Let's start with a common technique used for splitting.

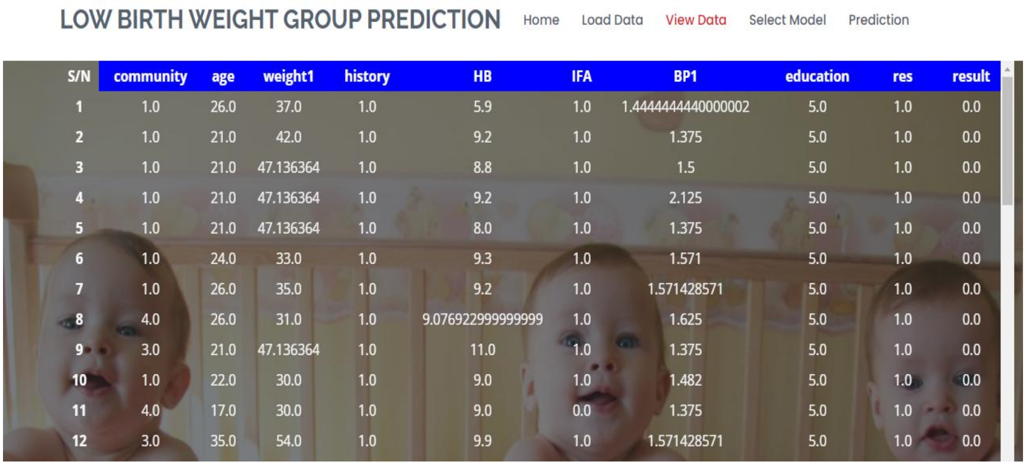
### 3.4 Support Vector Machine:

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text. So, you're working on a text classification problem. You're refining your training data, and maybe you've even tried stuff out using Naive Bayes. But now you're feeling confident in your dataset, and want to take it one step further. Enter Support Vector Machines (SVM): a fast and dependable classification algorithm that performs very well with a limited amount of data to analyze.

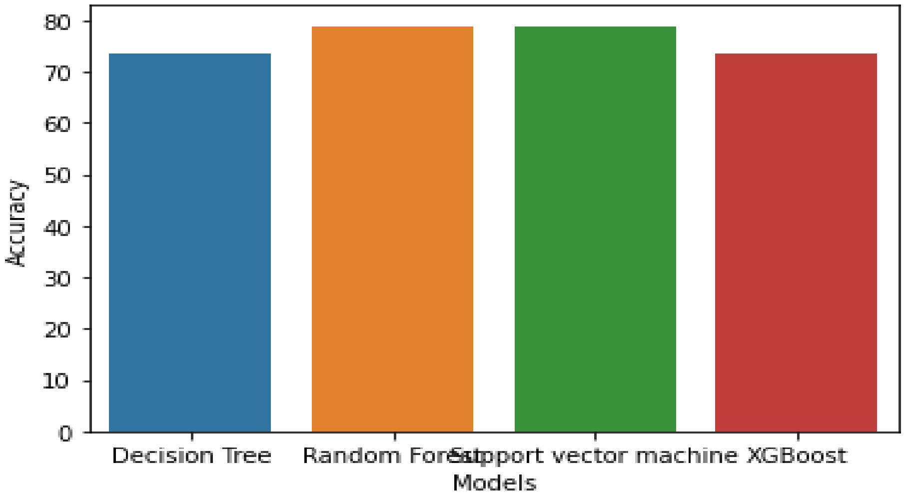
Perhaps you have dug a bit deeper, and ran into terms like linearly separable, kernel trick and kernel functions. But fear not! The idea behind the SVM algorithm is simple, and applying it to natural language classification doesn't require most of the complicated stuff.

## 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


LOW BIRTH WEIGHT GROUP PREDICTION Home Load Data View Data Select Model Prediction

| S/N | community | age | weight1 | history | HB | IFA | BP1 | education | res | result |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.0 | 26.0 | 37.0 | 1.0 | 5.9 | 1.0 | 1.4444444440000002 | 5.0 | 1.0 | 0.0 |
| 2 | 1.0 | 21.0 | 42.0 | 1.0 | 9.2 | 1.0 | 1.375 | 5.0 | 1.0 | 0.0 |
| 3 | 1.0 | 21.0 | 47.136364 | 1.0 | 8.8 | 1.0 | 1.5 | 5.0 | 1.0 | 0.0 |
| 4 | 1.0 | 21.0 | 47.136364 | 1.0 | 9.2 | 1.0 | 2.125 | 5.0 | 1.0 | 0.0 |
| 5 | 1.0 | 21.0 | 47.136364 | 1.0 | 8.0 | 1.0 | 1.375 | 5.0 | 1.0 | 0.0 |
| 6 | 1.0 | 24.0 | 33.0 | 1.0 | 9.3 | 1.0 | 1.571 | 5.0 | 1.0 | 0.0 |
| 7 | 1.0 | 26.0 | 35.0 | 1.0 | 9.2 | 1.0 | 1.571428571 | 5.0 | 1.0 | 0.0 |
| 8 | 4.0 | 26.0 | 31.0 | 1.0 | 9.076922999999999 | 1.0 | 1.625 | 5.0 | 1.0 | 0.0 |
| 9 | 3.0 | 21.0 | 47.136364 | 1.0 | 11.0 | 1.0 | 1.375 | 5.0 | 1.0 | 0.0 |
| 10 | 1.0 | 22.0 | 30.0 | 1.0 | 9.0 | 1.0 | 1.482 | 5.0 | 1.0 | 0.0 |
| 11 | 4.0 | 17.0 | 30.0 | 1.0 | 9.0 | 0.0 | 1.375 | 5.0 | 1.0 | 0.0 |
| 12 | 3.0 | 35.0 | 54.0 | 1.0 | 9.9 | 1.0 | 1.571428571 | 5.0 | 1.0 | 0.0 |




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

## V. CONCLUSION

In this application, we have successfully created a ML model to estimate whether the baby belongs to the Low Birth Weight or not belongs to the Low Birth. This is developed in a user-friendly environment using Flask via Python programming. We noticed that out of XGBoost Classifier, Random Forest Classifier, Decision Tree Classifier and Support Vector Classifier Decision Tree Classifier performs well with better 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.