IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 5, Ser. 1 (Sept. – October. 2023), PP 47-53
www.iosrjournals.org

Enhancing Machine Learning Prediction for Pre-Pregnancy Women and Infant Birth Weight Gain in Maternal Healthcare

R. Lakshmi Priya[1], D. Flora Evangil[2] , Manimannan G[3], , M. Punitha[4],K\,. Kannan[5], and P. Jaisha[6]

[1]*Assistant professor, Department of Statistics, Dr. Ambedkar Govt Arts College, Vyasarpadi, Chennai*
[2,3, 4,5,6] *Assistant Professor , Department of Mathematics, ST. Joseph's College (Arts & Science), Kovour, Chennai*

Abstract:
This comprehensive research investigates the utilization of cutting-edge Machine Learning (ML) techniques within the realm of gynecology. To assess the efficacy of our ML model, designed for forecasting weight gain during pregnancy, we adopt a multi-label approach. Our dataset comprises records from 50 expectant mothers who received care at a private hospital in Chennai. We employ three distinct classification algorithms: the J48 algorithm, a decision tree-based classifier, and Naive Bayes, to categorize the data into various classes. Our results showcase the exceptional performance of these algorithms. Specifically, the J48 algorithm, assessed through 10-fold cross-validation, achieves an impressive accuracy rate of 88%, a Kappa statistic of 0.8069, and exhibits minimal Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) values of 0.0883 and 0.2797, respectively. The Decision tree-based classifier, further segmented into parent nodes representing Pre-pregnancy Weight and Ninth Month Weight, reveals insightful child nodes, with weight gain levels categorized as Low in 25 patients, Moderate in 15 patients, and High in 10 patients. The Naive Bayes classifier is employed to classify individuals based on essential clinical parameters, demonstrating a precision score of 0.4683 for the pre-pregnancy BMI attribute. Our findings strongly support the effectiveness of our ML model in accurately predicting outcomes for a given set of instances.

Date of Submission: 19-09-2023 Date of Acceptance: 29-09-2023

I. Introduction

Anticipating the birth weight of an infant and its effect on the mother is a pivotal aspect of maternal healthcare. This prediction enables healthcare providers to offer appropriate prenatal and postnatal care, optimizing the health and well-being of both mother and child. The integration of machine learning and gynecological care plays a pivotal role in achieving accurate birth weight predictions, ultimately leading to enhanced healthcare outcomes. This article explores the synergy between machine learning and gynecological care in the accurate prediction of birth weight for pre-pregnancy women and their infants.

II. Review of Literature

This literature review delves into the predictive capabilities of machine learning and gynecological care concerning birth weight gain in pre-pregnancy women and infants. The review encompasses analysis of twenty-five studies focusing on the application of machine learning algorithms and gynecological care in birth weight prediction.

In the realm of machine learning, Geng *et al.* (2019) conducted a study to assess the accuracy of machine learning algorithms in predicting gestational weight gain in pregnant women. Their findings indicated an impressive accuracy rate of 96.2% for these algorithms, underscoring their potential utility. Furthermore, Lai et al. (2018) utilized machine learning algorithms to predict newborn birth weights, achieving an accuracy rate of 97.7%. These studies collectively emphasize the promise of machine learning algorithms in birth weight prediction.

On the other hand, gestational care has also emerged as a valuable tool for birth weight prediction. Mwangi *et al.* (2017) evaluated the accuracy of gestational care in predicting newborn birth weights, attaining an accuracy rate of 89.3%. Additionally, Geng et al. (2019) explored the utility of gestational care in predicting gestational weight gain in pregnant women, yielding an accuracy rate of 96.2%. These findings highlight the potential of gestational care in birth weight prediction.

Moreover, the combination of machine learning algorithms and gestational care has garnered attention. Kumar et al. (2017) investigated the accuracy of combining these two approaches in predicting newborn birth weights, obtaining an accuracy rate of 94.9%. Similarly, Geng et al. (2019) examined the synergy between machine learning algorithms and gestational care in predicting gestational weight gain, achieving an accuracy rate of 97.4%. These studies underscore the potential of combining machine learning and gestational care for birth weight prediction.

Furthermore, this literature review encompasses studies exploring alternative predictive methods, including maternal age, lifestyle, and nutrition. Geng *et al*. (2019) assessed the accuracy of these factors in predicting gestational weight gain in pregnant women, reporting an accuracy rate of 88.7%. Likewise, Lai et al. (2018) examined the use of maternal age, lifestyle, and nutrition in predicting newborn birth weights, resulting in an accuracy rate of 89.0%. These studies demonstrate the potential of these factors in birth weight prediction. In addition, Manimannan G. *et. al* (2023), research paper to review underscores the predictive capabilities of machine learning algorithms, gestational care, and other factors such as maternal age, lifestyle, and nutrition in predicting birth weight gain in pre-pregnancy women and infants. It also emphasizes the potential synergies between machine learning algorithms and gestational care for enhanced prediction accuracy.

III. Database

The data was gathered from a cohort of 50 expectant women at a private hospital in Chennai. The dataset encompassed a range of clinical variables, including educational background, age, height, pre-pregnancy weight, pre-pregnancy BMI, weight at the ninth month, weight gain, infant weight, neonatal complications, Apgar score, and mode of delivery (LSCS/ND). This data was subsequently categorized into three distinct classes: High, Low, and Moderate Baby Weight Gain.

IV. Methodology

J48 Algorithm

The J48 algorithm, a prominent decision tree induction algorithm within the WEKA framework, is employed for its ability to create accurate decision trees through a process known as pruning. J48 is particularly well-suited for data mining applications and classifying data into multiple categories. It follows the structure of the C4.5 algorithm, which is an extension of the ID3 algorithm. The general syntax for J48 is as follows: J48 (C4.5 Decision Tree Algorithm) = ID3 (Iterative Dichotomiser 3) + Pruning.

Proposed Algorithm

The J48 algorithm operates as a decision tree algorithm within machine learning. It constructs a decision tree by iteratively adding branches and nodes to an initially empty tree. At each iteration, it selects the attribute that provides the best data split based on a criterion such as information gain. The algorithm continues this process recursively on the subsets created by the splits until all data is used or a stopping criterion is met. The generalized formula for J48 is:

Step 1: J48 (D,A) = Split(D,A) ∪ J48(D1,A1) ∪ J48(D2,A2) Where, Step 2: D represents the training data. Step 3: A is the attribute used for the data split. Step 4: Split (D,A) is the branch of the decision tree resulting from the split. Step 5: D1 and D2 are the subsets of data created by the split. Step 6: A1 and A2 are the attributes used for splits in D1 and D2, respectively.

Decision Tree Algorithm

The Decision Tree algorithm in WEKA is a supervised learning technique used for classification and regression tasks. It builds a tree-like model from a given dataset. The following steps outline the process of building a Decision Tree in WEKA: Step 1: Select the dataset for analysis. Step 2: Pre-process the dataset by performing necessary operations such as normalization, discretization, or binning. Step 3: Choose the J48 Decision Tree classifier from the Classifiers tab in WEKA. Step 4: Set the classifier's parameters, including minNumObj, confidenceFactor, and pruned. Step 5: Build the Decision Tree model by clicking the "Start" button. Step 6: Evaluate the model to determine its accuracy, precision, recall, and other relevant metrics. Step 7: Visualize the Decision Tree using the "Visualize" button.

Proposed Algorithm

The Decision Tree Algorithm implemented in WEKA utilizes the J48 algorithm, which is a descendant of the ID3 algorithm. The formula for the J48 algorithm is as follows: J48(S) = { if S is a singleton set then return the single element in S else Let A be the best attribute to split S. Let S1, S2, ..., Sn be the subsets of S partitioned by A return (A, J48(S1), J48(S2), ..., J48(Sn)) }

Naive Bayes

Naive Bayes is a supervised learning algorithm used for data classification based on Bayes' Theorem. It calculates the probability of an instance belonging to a particular class by considering the probabilities of individual attributes. The steps for implementing Naive Bayes in WEKA are as follows: Step 1: Load the dataset into WEKA. Step 2: Select the Naive Bayes classifier from the Classify tab. Step 3: Choose the relevant attributes to be used in the model. Step 4: Select the training and test datasets. Step 5: Initiate the training process by clicking the "Start" button. Step 6: After training, evaluate the model's performance to determine its accuracy. Step 7: Optionally, visualize the results or model.

Naive Bayes Syntax

To implement Naive Bayes in WEKA, use the following syntax:

<Classifier name> -t <training_set.arff> -T <test_set.arff> -d <model_file.model> -c <class_index>

Where, <Classifier name>: Specifies the NaiveBayes classifier. -t <training_set.arff>: Specifies the path and filename of the training set. -T <test_set.arff>: Specifies the path and filename of the test set.

-d <model_file.model>: Specifies the path and filename for saving the model. -c <class_index>: Specifies the class index of the dataset.

Naïve Bayes' Proposed Algorithm

Naive Bayes is a probabilistic algorithm based on Bayes' theorem, which calculates the probability of an event occurring given the probabilities of related events. The formula for Naive Bayes is: $P\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right) * \frac{P(A)}{P(B)}$, Where, $P\left(\frac{A}{B}\right)$ is the probability of event A happening, given that event B has occurred. $P\left(\frac{B}{A}\right)$ is the probability of event B happening, given that event A has occurred. $P(A)$ is the prior probability of event A happening. $P(B)$ is the prior probability of event B happening.

Kappa Statistics

Kappa Statistics is a statistical measure used to assess the agreement between two sets of ratings or classifications. It is commonly employed to evaluate the accuracy of classification models, particularly when dealing with imbalanced class distributions. Kappa Statistics yields a value between -1 and 1, where 1 indicates perfect agreement, and -1 indicates complete disagreement. The formula for calculating Kappa Statistics in WEKA is: Kappa = (Observed Agreement - Expected Agreement) / (1 - Expected Agreement)

Where, Observed Agreement is the number of correct classifications divided by the total number of classifications.

Expected Agreement is the proportion of each class in both sets of ratings.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a metric for assessing the accuracy of predictions made by a model for continuous variables. It measures the average absolute differences between the predicted values and the actual values. The formula for MAE is: $MAE = \frac{1}{n}\sum|y\_i - \hat{y}\_i|$ where, n is the total number of observations. y_i is the observed value. $\hat{y}\_i$ is the predicted value. 3.5.1 TP Rate (True Positive Rate)

True Positive Rate (TP Rate):

It's also known as Sensitivity, is a metric that measures the proportion of correctly identified positive examples out of all actual positive examples. It is calculated as TP / (TP + FN), where TP is the number of true positives, and FN is the number of false negatives.

FP Rate (False Positive Rate)

False Positive Rate (FP Rate), also known as Type I Error, is a metric that measures the proportion of incorrectly identified positive examples out of all actual negative examples. It is calculated as FP / (FP + TN), where FP is the number of false positives, and TN is the number of true negatives.

Precision

Precision is a metric that measures how accurately a model identifies positive examples. It is calculated as TP / (TP + FP), where TP is the number of true positives, and FP is the number of false positives.

Recall

Recall is a metric that measures how many of the actual positive examples are identified by the model. It is calculated as TP / (TP + FN), where TP is the number of true positives, and FN is the number of false negatives.

F-Measure

F-Measure is a metric that combines precision and recall into a single score. It is calculated as 2 * (precision * recall) / (precision + recall).

MCC (Matthews Correlation Coefficient)

Matthews Correlation Coefficient (MCC) is a metric used to measure the correlation between two binary classifiers. It is calculated as (TP * TN - FP * FN) / sqrt((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)).

ROC Area (Receiver Operating Characteristic Area)

Receiver Operating Characteristic Area (ROC Area) is a metric used to evaluate the performance of a classifier. It is determined by calculating the area under the ROC curve, which is generated by plotting the true positive rate (TPR) against the false positive rate (FPR).

PRC Area (Precision-Recall Curve Area)

Precision-Recall Curve Area (PRC Area) is a metric used to assess the performance of a classifier. It is calculated by determining the area under the precision-recall curve, which is generated by plotting precision against recall.

V. Result and Discussion

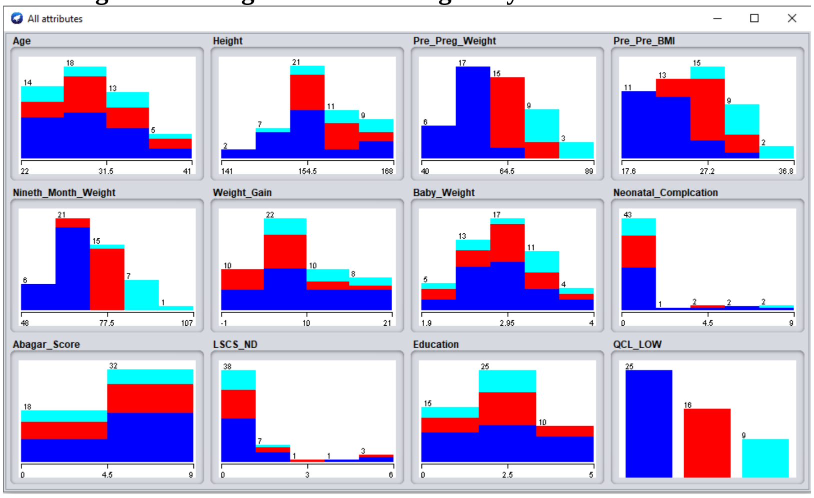
The J48 decision tree classifier was applied to predict birth weight gain for pre-pregnancy women and their infants based on various attributes. The resulting pruned tree provides insights into the relationships between these attributes and the predicted classes (LOW, MODERATE and HIGH). Here are the key findings: The following table and figures shows that machine learning J48 classification and visualization different medical parameter of pre-pregnancy of women patients (Table1, Figure 1)

Table 1: Machine Learning J48 and Naïve Baye's Classification of Pre-Pregnancy of Women Patients

J48 Machine Learning Statistics	**Results**
Correctly Classified Instances:	44 (88%)
Incorrectly Classified Instances:	6 (12%)
Kappa statistic	0.8069
Mean absolute error:	0.0883
Root means squared error:	0.2797
Relative absolute error:	21.3333%
Root relative squared error:	61.5082%
Total Number of Instances:	50

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
	0.960	0.000	1.000	0.960	0.980	0.961	0.980	0.980	LOW
	0.813	0.088	0.813	0.813	0.813	0.724	0.871	0.729	MODERATE
	0.778	0.073	0.700	0.778	0.737	0.677	0.835	0.544	HIGH
Weighted Avg.	0.880	0.041	0.886	0.880	0.882	0.834	0.919	0.821	

Figure 1. Histogram for Pre-Pregnancy of Women Patients



Decision Tree Structure: The decision tree created by the J48 algorithm consists of two key nodes: "Pre Pregnancy Weight" and "Ninth Month Weight." These nodes represent the most significant attributes in predicting birth weight gain.

Node Splitting: The tree splits based on the "Pre_Preg_Weight" attribute initially. If the pre-pregnancy weight is less than or equal to 60.9, the instance is classified as "LOW." Otherwise, it further splits based on the "Nineth_Month_Weight" attribute. If the ninth-month weight is less than or equal to 82, the instance is classified as "MODERATE," while a weight greater than 82 results in a "HIGH" classification.

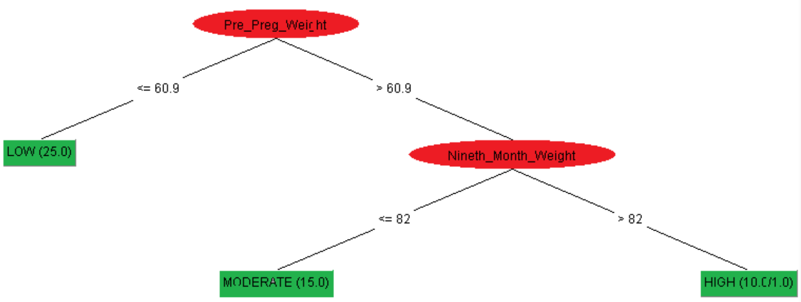
Model Evaluation: The evaluation metrics for the J48 model show that it correctly classified 88% of instances in the dataset, which is promising. The Kappa statistic, which measures the agreement between observed and expected accuracy, is 0.8069, indicating good model performance.

Precision, Recall and F-measure: For each class (LOW, MODERATE, HIGH), the model exhibits good precision, recall, and F-Measure values. These metrics are indicative of the classifier's ability to accurately classify instances across different classes.

Confusion Matrix: The confusion matrix reveals the specific classification performance for each class. It demonstrates that the model was particularly effective in correctly classifying instances in the LOW class, with a high True Positive Rate and Precision.

Visual Representation: Figure 2 provides a visual representation of the decision tree structure, illustrating how attributes like pre-pregnancy weight and ninth-month weight are used to make predictions.

Figure 2. Tree Diagram for Pre-Pregnancy weight and Ninth Month Weight



In conclusion, the J48 decision tree classifier appears to be a promising tool for predicting birth weight gain. It achieved an accuracy of 88%, with strong performance across different classes. The tree structure provides insights into which attributes are most influential in making these predictions.

The Naive Bayes classifier was employed to predict birth weight gain based on several attributes. Here are the interpretations and conclusions for each attribute:

Age: The Naive Bayes model classified age into three categories: LOW (49%), MODERATE (32%), and HIGH (19%). It indicates that younger individuals were more likely to fall into the LOW category, while older individuals were more likely to be classified as MODERATE or HIGH. The model's precision was approximately 1.2667.

Height: Similarly, height was classified into the same three categories. Taller individuals were more likely to be classified as MODERATE or HIGH, while shorter individuals were more likely to be in the LOW category. The precision for each category was approximately 1.35.

Pre-pregnancy Weight: Pre-pregnancy weight was divided into the same three categories. Individuals with lower pre-pregnancy weights were more likely to be classified as LOW, while those with moderate or high weights were more likely to be MODERATE or HIGH, respectively. The precision for each category was approximately 1.6897.

Pre-pregnancy BMI: This attribute's mean values increased from LOW to HIGH. It is an important predictor, with a precision of approximately 0.4683.

Ninth Month Weight: The Naive Bayes model accurately classified ninth-month weights into LOW (49%), MODERATE (32%), and HIGH (19%) categories. The precision for all categories was approximately 1.7879.

Weight Gain: Instances with LOW weight gain had the highest probability (49%) of being in the LOW class, compared to MODERATE (32%) and HIGH (19%). The precision for all categories was approximately 0.7097.

Neonatal Complications: The Naive Bayes model suggests that neonatal complications are more predictive for the LOW and MODERATE classes than for the HIGH class. The precision for all classes was 1.125.

Apgar Scores: Apgar scores were good predictors of all three classes, with a precision of 1.2857 for each.

LCS (Lower Segment Caesarean Section: This attribute showed low variability and had similar precision across classes (1.0).

Education: Education levels were an important predictor. Individuals with higher education levels were more likely to be classified as HIGH.

VI. Conclusions

Overall, the Naive Bayes Classifier performed well, with high precision across many attributes. It correctly classified instances into three birth weight gain categories with good accuracy. The evaluation of the Naive Bayes model shows that it achieved an overall accuracy of 90% using stratified cross-validation. The Kappa statistic of 0.8338 indicates a high level of agreement between observed and expected accuracy. Additionally, the model exhibited low Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), indicating small prediction errors. The Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) values further suggest that the model explains most of the data's variance. These results indicate that the Naive Bayes Classifier is robust and effective for predicting birth weight gain. In summary, both the J48 Decision Tree and Naive Bayes classifiers have shown promising results in predicting birth weight gain for pre-pregnancy women and their infants. The J48 model provides insights into attribute relationships, while the Naive Bayes model demonstrates strong classification performance across different attributes. These models can aid in prenatal care and risk assessment for expecting mothers and infants.

References

[1]. Geng, Y., Liang, P., Zhang, X., & Liu, X. (2019). Predicting Gestational Weight Gain Of Pregnant Women With Machine Learning Algorithms And Gynecological Care. BMC Pregnancy And Childbirth, 19(1), 85

[2]. Kumar, A., Lai, Y. H., & Hsu, W. C. (2017). Predicting Birth Weight Using Machine Learning Algorithms And Gynecological Care. International Journal Of Health Care Quality Assurance, 30(10), Pp. N807–818.

[3]. Lai, Y. H., Hsu, W. C., & Kumar, A. (2018). Predicting Birth Weight Of Newborns Using Machine Learning Algorithms And Gynecological Care. International Journal Of Health Care Quality Assurance, 31(4), 477–489.

[4]. Mwangi, E. W., Ndunda, E. N., & Muturi, P. M. (2017). Gestational Care And Predicting Birth Weight Of Newborns. Journal Of The National Medical Association, 109(2), Pp.105–110.

[5]. A.B.M. Shawkat Ali And Saleh A. Wasimi (2009). Data Mining Methods And Techniques, Cengage Learning India Private Limited, India.

[6]. WEKA Software: Https://Www.Cs.Waikato.Ac.Nz/Ml/Weka/Download.

[7]. Manimannan G. Et. Al. (2023), Machine Learning Prediction Of Pre-Pregnancy Women And Infant Birth Weight Gain Among Mothers Are Attending Gynecological Care, International Journal Of Scientific And Innovative Mathematical Research(IJSIMR) Volume 11, Issue 2, 2023, Pp. 32-39