

Performance Analysis of Supervised Learning Classifiers for the Prediction of Child Birth Weight

M.A.C. Akmal Jahan¹, A.M. Razmy¹, R.G. Ragel² and S. Deegalla²

¹*Faculty of Applied Sciences, South Eastern University of Sri Lanka*

²*Faculty of Engineering, University of Peradeniya*

Corresponding Author: akmaljahan@fas.seu.ac.lk

Even though the technology advancement has helped improving medical and health sectors, high infant mortality rate is still considered as a serious problem in developing countries. Low birth weight (LBW) plays a major role in infant mortality. There are several reasons for LBW of a child. There is a wide variation of LBW within Sri Lanka in different geographical areas. Particularly the districts belong to plantation workers, Monaragala, Ampara and Polonnaruwa districts show the highest percentages of newborns belonging to the LBW category. From the statistics Ampara district shows 17% of LBW. The objective of this work is to find a suitable way to predict the child weight using the existing pattern of low birth weight in Ampara region. There is a need to find an algorithm with good performance among several existing supervised learning classifiers to construct a decision model. This work was carried out among 2700 pregnant mothers throughout the MOH offices in Ampara district. Initially the existing data were manually classified into three classes such as Normal Birth Weight (NB), Low Birth Weight (LB) and High Birth Weight (HB). Several important parameters were captured from the data set. C4.5, CART and ID3 supervised learning classifiers in Weka data mining and machine learning tool were used in this experiment.

The data were handled through three major processes namely as pre-processing, attribute selection and construction of decision trees using classifiers. The missing values in the huge data set were handled in pre-processing. The useful and most significant parameters were selected and ranked using feature selection process. Three decision tree algorithms were used to construct decision tree. The accuracy and time complexity for the tree construction were measured using the experimental tool Weka by applying 10-fold cross validation from the experimental results on accuracy and time complexity of the decision tree classifiers, C4.5 produces higher accuracy as 86.15% and the time complexity is less than 1 minute. Considering the time complexity and accuracy C4.5 works effectively compared to the others. Therefore C4.5 was selected as a best classifier to construct decision tree model for the prediction of child weight in Ampara district.

Keywords: Classifier, C4.5, Machine learning, Weka