Classification Approach based on Association Rules mining for Unbalanced data: Application to In-hospital Maternal Mortality in Senegal and Mali

Cheikh Ndour[1,2,3] & **Simplice Dossou Gbété**[2] & **Aliou Diop**[1] & **Alexandre Dumont**[3]

[1] *University Gaston Berger, LERSTAD, Saint-Louis, Senegal*

[2] *University Pau et des pays de l'Ardour, UMR CNRS 5142, Pau, France*

[3] *Research Institute for Development, UMR IRD 216, Paris, France*

Abstract

This paper deals with the supervised classification when the response variable is binary and its class distribution is unbalanced. In such situation, it is not possible to build a powerful classifier by using standard methods such as logistic regression, classification tree, discriminant analysis, etc. To overcome this short-coming of these methods that provide classifiers with low sensibility, we tackled the classification problem here through an approach based on the association rules learning because this approach has the advantage of allowing the identification of the patterns that are well correlated with the target class. Association rules learning is a well known method in the area of data-mining. It is used when dealing with large database for unsupervised discovery of local patterns that expresses hidden relationships between variables. In considering association rules from a supervised learning point of view, a relevant set of weak classifiers is obtained from which one derives a classification rule that performs well.

Keywords. Classification, Class Association Rules, Unbalanced dataset, Maternal Mortality . . .