Prediction of Diabetes Using Data Mining Technique: A Review

MWNL de Silva# and DU Vidannagama

Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka
#mwndesilva001@gmail.com

Data mining plays an efficient role in the prediction of diseases in health care industry. Diabetes has become one of the major global health problems at present. According to the WHO 2014 report, around 422 million people worldwide are suffering from diabetes. Diabetes is a metabolic disease where the improper management of blood glucose levels led to risk of many diseases like heart attack, kidney disease and eye problems. Many algorithms have been developed for the prediction of diabetes and accuracy estimation, and a lot of automated systems have been implemented for the prediction of diabetes using those algorithms. In order to automate the overall process of diabetes prediction and severity estimation, diabetic databases are needed. This repository of diabetic database helps in the identification of impact of diabetes on various human organs. The more accuracy of prediction is kept, the more chances of accurate severity estimation. Therefore, this study presents different prediction methods of diabetes using different algorithms and their evaluation, and it gives a detailed evaluation of existing data mining methods used for prediction of diabetes. Further, it gives an idea about the tools which can be used in data mining. The study objective is to present a review on the existing automated diabetes prediction systems and the algorithms used. It gives information about the systems' accuracy and the tools which were used in accuracy measuring. Finally, when referring selected research articles, it was identified that most researchers have preferred to use decision tree related algorithms in their research work. Moreover, the proposed algorithm has given the best performance in most cases. Accordingly, it can be stated that to have the best accuracy and performance the decision tree algorithm and related algorithms can be used.

Keywords: Data Mining, Diabetes, Algorithms