COMPARISON OF DATA MINING ALGORITHMS FOR DIAGNOSIS OF DIABETES MELLITUS

Ahmed Sami Jaddoa[1] & Ziyad Tariq Mustafa Al-Ta'i[2]

[1]*Research Scholar, Business Informatics College, University of Information Technology and Communications, Iraq*

[2]*Professor, Department of Computer Science, Science College, Diyala University, Iraq*

ABSTRACT

Diabetes is specified as the most chronic and deadliest disease that results in increasing blood sugar. The medical data mining approaches were utilized for detectingun observed patterns in the medical field sof sets of data for medical diagnosis and treatment. Data classification for diabetes mellitus is quite significant. Where utilizing two types of data sets, the first is local, collected from consulting laboratories at Baqubah General Hospital, and the second is global, which is the Pima India Diabetes Database. The experiment on the Local dataset shows that the accuracy if K-NN is 90 %, the accuracy of the SVM has been 98 %, the accuracy of the NB is 98 % and the accuracy of RF is 98 %. The experiment on the Pima dataset shows that the accuracy of K-NN is 81 %, the accuracy of SVM has been82 %, the accuracy of NB is 84 % and the accuracy of RF is 82 %.

KEYWORDS: *Diabetes Mellitus, Data Mining, Diagnosis, K Nearest Neighbors, Classification, Support Vector Machine, Naive Bayes, Random Forest*