Applying Data Mining Techniques to Analysis and Prediction of Chronic Kidney Disease

P.N.D. Panawenna[1,*] and H.M.B.P. Ranaweera[2]

[1,2]*Department of Information Systems, Faculty of Management Studies, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*

*Corresponding author: nisankaugc@gmail.com

Abstract

Early prediction provides opportunities to perform possible treatments and strategies for reducing the probability of kidney failure and prevents the progress of CKD. For the early prediction, can use data mining. Major data mining techniques such as classification, clustering, association rules, and regression can be used to predict CKD. This study has two-part. The first one is finding out the best accuracy algorithm to predict CKD and buildup a model. For that, the researcher used 11 classification, 3 clustering, and 5 ensemble algorithms while using the feature selection method. All the results of the classification and ensemble algorithm were obtained from 4 methods such as full training set, supplied test set, cross-validation, and percentage-split methods. All the results of the clustering algorithm obtained from 4 methods such as full training set, supplied test set, percentage-split, and classes to clusters evaluation. Then researcher analyzed and compared the results of all the classifiers. After analyzed and compared the results of all the classifiers, the researcher ranks the classifiers according to the evaluation criteria that were used to evaluate the classifiers. The most important evaluation criteria are accuracy. Thus, the classifiers were ranked again according to the accuracy. Accordingly, the Vote ensemble algorithm was selected which is the best algorithm in the both table to build-up model to conduct the prediction. The second part is the prediction of the stage of Chronic Kidney Disease and the variation and severity of the disease being affected. To predict this, it was estimated Glomerular Filtration Rate (eGFR) using the Modification of Diet in Renal Disease (MDRD) and Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. It calculates the eGFR using the 2 equations separately. Then comparing the results with the standard values of Stages of Chronic Kidney Disease, it finds out the stage of Chronic Kidney Disease and the variation and severity of the disease being affected. So finally, using the vote algorithm able to predict the CKD, then using the 2 equation, able to predict the stage of CKD, and the variation and severity of the disease being affected earlier. So, it helps to perform possible treatments and strategies earlier, for reducing the probability of kidney failure and also helps to prevent the progress of CKD, increasing morbidity and mortality related to CKD, and high health care costs.

Keywords: Chronic kidney disease, classification algorithm, clusteringalgorithm, data mining techniques, ensemble algorithms