Abstract

Diabetes Mellitus (DM) is the third leading cause of death in Indonesia, and type II DM is more dangerous because of the combination of genetic and lifestyle factors. The high rate of patients with type II DM is caused by delay in diagnosis, therefore, early detection of disease is necessary to classify the detected patients with type II diabetes mellitus, and undetected patients. Moreover, see analysis of determinant and major variables are highly recommended. In this research is implemented the combined Classification and Regression Tree method (CART) and Random Forest (RF) to build the classification model used in the early detection of diabetes mellitus type II disease. It is based on the characteristics of the dataset used medical record that have complex variables consisting of several categorical variables and continuous variables, and the advantages of the CART models are easy to implement, and can explore the structure of complex medical records, while the RF method can handle the problem of accuracy. In this research tested the different numbers of trees and the number of candidate variables splitter. Based on the test results show that the addition of trees and variables splitter can improve the accuracy and reduce the error rate, with the optimal inputs are 50 number of trees and 3 number of variables splitter with 84% average accuracy. The important variable of early detection of diabetes mellitus type II is heredity, age, and body mass index.

Keywords : Diabetes Mellitus type II, CART, *Random Forest*, classification, the early detection of disease, the important variable.