Abstract

Diabetes is a disease caused by high blood sugar in the body or beyond normal limits. Diabetics in Indonesia have experienced a significant increase, Basic Health Research states that diabetics in Indonesia were 6.9% to 8.5% increased from 2013 to 2018 with an estimated number of sufferers more than 16 million people. Therefore, it is necessary to have a technology that can detect diabetes with good performance, accurate level of analysis, so that diabetes can be treated early to reduce the number of sufferers, disabilities, and deaths. The different scale values for each attribute in Gula Karya Medika's data can complicate the classification process, for this reason the researcher uses two data normalization methods, namely min-max normalization, z-score normalization, and a method without data normalization with Random Forest (RF) as a classification method. Random Forest (RF) as a classification method has been tested in several previous studies. Moreover, this method is able to produce good performance with high accuracy. Based on the research results, the best accuracy is model 1 (Min-max normalization-RF) of 83.19%, followed by model 2 (Z-score normalization-RF) of 82.64%, and model 3 (without data normalization-RF) of 81.63%. From these results, it can be concluded that model 1 (Min-max normalization-RF) is better than the other two data normalization models and is able to increase the performance of classification Random Forest by 83.19%.

Keywords: diabetes, classification, min-max normalization, z-score normalization, random forest.