ABSTRACT

Diabetes is a diseases that disturb the metabolic system and non-contagious. Diabetes are caused by two factors, namely modifiable factors including diet, smoking, hypertension, etc. and unmodifiable factors including age, gender, and family background. Based on these factors, it is necessary to analyze with classification algorithms because the disease suffered by patients can be classified, making it easier for doctors or medical personnel to provide appropriate treatment solutions. Classification algorithms that can be applied is CART. The CART algorithm has several weaknesses, one of which is less efficient against high-dimensional data, so it is necessary to optimize the CART algorithm with PSO to overcome the problem of feature selection. The research used data on type II diabetes patients as much as 952 data with 14 independent variables and 1 dependent variable. The application of PSO optimization can select 14 variables into 12 variables that meet. The best particle position value in each variable that fulfills produce a value greater than the threshold of 0.5. Variables that fulfill the feature selection process include age, gender, family background, high blood pressure, physical activity, smoking history, history of drinking alcohol, length of sleep, length of deep sleep, regular medication consumption, fast food consumption, and stress conditions. The best classification modeling results is CART algorithm based on PSO with accuracy, sensitivity, and specificity values of 95.79%, 99.50%, and 86.91% respectively and the ROC curve results on the CART algorithm based on PSO using training and test data of 0.999 and 0.953 respectively. This value is close to the perfect value.

Keywords: Diabetes, CART, PSO, Accuracy, Sensitivity, Specificity