ABSTRACT

Diabetes is a chronic metabolic disorder in the human body characterized by high blood sugar levels and accompanied by impaired carbohydrate, lipid and protein metabolism as a result of the inability of insulin to function adequately. Based on data from the International Diabetes Federation (IDF), Indonesia is in a diabetes alert status by ranking 7th out of 10 countries with the highest number of diabetes patients in the world. To identify diabetes, one of the things that can be done is by classifying diabetes. One way to do the classification is by doing classification in Machine Learning. Machine Learning itself can make it easier for us to get predictive results for diabetes, the dataset that will be used in this study is pima-indian-diabetes with a dataset of 768 data, from that data there are eight data features including Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI (Body Mass Index), Diabetes Pedigree Function, Age and outcome. In this study, the author will compare the accuracy results of the Random Forest algorithm and the Artificial Neural Network algorithm in classifying the pima-indian-diabetes dataset. Before comparing the results of the accuracy of the two algorithms, the author performs the preprocessing stage of the dataset by cleaning the dataset, creating a matrix of features, splitting the data, and performing feature scaling. After the data passes through the preprocessing stage, the next step is to find the best accuracy obtained by making a Confusion Matrix to find the results of the ROC AUC and the F1-Score results of each algorithm used. The results of the analysis show that the Random Forest algorithm has an accuracy value of 90.62%. The accuracy value obtained by the Random Forest algorithm has a better value when compared to the Artificial Neural Network algorithm which has an accuracy value of 82.29%

*Keywords-**Random Forest, Artificial Neural Network, Diabetes, Data Mining***