

ABSTRACT

Diabetes is known as an epidemic disease that affects almost every country, age group, and economic factor around the world. In Indonesia, based on data from the IDF in 2020, of the total adult population of 172 million people, 10.5 million people suffer from diabetes. When designing software or medical diagnosis software, disease prediction becomes one of the most time-consuming tasks because there are many aspects, each of which must be examined based on the available data. Machine learning techniques are also used as a tool to make medical diagnoses. In this study, the authors compare the accuracy of the AdaBoost algorithm with the LightGBM algorithm in classifying diabetes to find out which algorithm has the highest level of accuracy and to know the model evaluation of each model. For the dataset used is the Pima Indians Diabetes Database dataset which has a total of 768 data. From the results of this study, it was found that the LightGBM algorithm has a higher accuracy value than the AdaBoost algorithm with an accuracy value of 91.67% LightGBM and an AUC value of 0.9704 compared to the AdaBoost algorithm which has an accuracy value of 91.14% and an AUC value of 0.9693.

Keywords—Data Mining, AdaBoost, LightGBM, Classification, Pima Indians Diabetes Database