ABSTRACT

Obesity is a disease that is caused by the accumulation of excess fat due to poor diet and unhealthy lifestyle. Based on a study conducted by the NCD Risk Factor Collaboration (NCD-RisC), it was found that from 1990 to 2022, weight problems increased twice in adults and four times in children which shows that an accurate and fast detection of obesity is necessary. In general, detecting obesity risk can be done using machine learning algorithms, especially ensemble learning. However, the use of ensemble learning algorithms generally requires a long time and high computational complexity because it creates such complex trees and computations. To overcome this problem, an optimized ensemble learning algorithm can be used so the time and complexity is reduced. This research classifies obesity by applying the Light Gradient Boosting Machine Classifier (LGBM) algorithm to the Obesity or CVD Risk dataset which has gone through several preprocessing processes. To ensure that the research is able to predict accurately, several tests were carried out to determine the combination of hyperparameters and train-test data proportion with maximum accuracy. The hyperparameter combination that is used in this research is learning_rate of 0.01, n_estimator with a value of 200, and max_depth of 7 with the train-test data proportion of 80 to 20. The test results in this research show that the Light Gradient Boosting Machine Classifier algorithm is able to achieve an accuracy value of 100.00% with a precision value of 100.00%, recall of 100.00%, and F1-Score of 100.00% to predict obesity levels.

Keywords: Light Gradient Boosting Machine (LGBM), Classification, Obesity