ABSTRACT

Hypertension is one of the most prevalent non-communicable diseases and a leading cause of global mortality, making early detection and accurate classification of hypertension status crucial. This study aims to compare the performance of two classification methods Binary Logistic Regression and the Naïve Bayes algorithm—in classifying hypertension status among patients at Padangsari Community Health Center, Semarang City. Both methods are supervised learning approaches commonly used in data mining, where Binary Logistic Regression identifies significant independent variables affecting hypertension status, while Naïve Bayes classifies based on probability assuming feature independence. The study involved 600 patient records from 2024, using various proportions of training and testing data, and performance evaluation was conducted using confusion matrix, sensitivity, and AUC values. The results show that the 70%:30% proportion yields the best performance for Binary Logistic Regression compared to other proportions, with a testing data accuracy of 67,39% and an AUC of 72,4%, as well as a training data accuracy of 67,57% and an AUC of 70,8%. The small sensitivity gap between training and testing data (0,18%) indicates model stability and reduced risk of overfitting. Therefore, Binary Logistic Regression is considered more optimal in classifying patient hypertension status compared to the Naïve Bayes algorithm.

Keywords: Hypertension, Binary Logistic Regression, Naïve Bayes, Classification, Accuracy, AUC.