

ABSTRACT

Diabetes mellitus is a chronic disease whose prevalence continues to increase, especially in type 2 which is often accompanied by complications. In Indonesia, diabetes has a serious impact, thus strengthening the urgency to understand and overcome the serious impact caused by diabetes. Early warning against this disease is very necessary because it allows for faster prevention and treatment. This study aims to identify the results of the application of machine learning algorithms in diabetes prediction and to find out the factors that most affect diabetes mellitus. This study uses the Gradient Boosting Machine algorithm with medical record data from Al Ihsan Hospital, West Java Province which consists of two classes, namely DM Type 2 and DM Type 2 + Participant. In its application, it uses data sharing ratios of 90:10, 80:20, and 70:30, which are applied to the grid search technique to get optimal results. The model evaluation was carried out using the confusion matrix and AUC value on the ROC curve. The best performance of the Gradient Boosting Machine algorithm is achieved with a data sharing ratio of 80:20, with the best parameters being learning rate = 0.2, number estimators = 50, and maximum depth = 3. The results showed an accuracy of 82% and an AUC value on the ROC curve of 0.90. The precision values of label 1 = 84% and label 0 = 80%, recall label 1 = 74% and label 0 = 88%, and f1-score label 1 = 79% and label 0 = 84%. Then, the results of visualization related to factors that affect the increase in diabetes mellitus in patients at Al Ihsan Hospital, West Java Province, namely HbA1c, Random Blood Glucose (RBG), and leucocyte count. This research is useful in supporting the early diagnosis and management of diabetes mellitus and as an early warning to the general public about the importance of periodic health checks. Thus, this research not only makes an academic contribution but also has a positive impact on medical practice to improve diabetes management.

Keywords—diabetes mellitus, Gradient Boosting Machine, machine learning, medical record data