

ABSTRACT

Type 2 Diabetes Mellitus (T2DM) is a metabolic disease often accompanied by comorbidities, such as hypertension and heart disease. This study aims to develop a random forest -based classification system to distinguish T2DM patients with and without comorbidities using medical record data from Al Ihsan General Hospital in Bandung. This system assists medical staff in accelerating diagnosis and improving the accuracy of identifying at-risk patients, as well as serving as an early warning system for the community.

The method used in this study is the Cross-Industry Standard Process for Data Mining (CRISP-DM), which includes the following stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. During the modeling stage, hyperparameter tuning was performed using GridSearchCV. Evaluation results showed that the random forest model achieved the highest accuracy of 89% in an 80:20 data split scenario, with precision of 100%, recall of 76%, and an F1 score of 86% for patients with comorbidities. These results indicate a good balance between accuracy and the model's ability to detect patients with comorbidities.

Through feature importance, this study identified key clinical parameters such as HbA1C and Fasting Blood Glucose (FBG) that play a significant role in distinguishing T2DM patients with and without comorbidities. This system was also implemented in a Streamlit-based application, which visualizes classification results and risk factors, helping medical personnel make faster and more accurate decisions.

Keywords— CRISP-DM, Comorbidity, Feature Importance, Random Forest, Type 2 Diabetes Mellitus.