Implementing SMOTE Techniques to Improve the Performance of Diabetes Prediction Models Using the SVM Algorithm

Fariddudin Ar-razi AB
Data Science Study Program, Faculty of Science and Technology
University of Technology Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail: fariddudinarraziab@gmail.com

Abstract

The rapid advancement of information technology has significantly contributed to the healthcare sector, particularly in supporting fast and accurate disease diagnosis. One such application is the development of predictive models for detecting chronic diseases such as diabetes. This study aims to build a diabetes risk prediction model by applying a resampling technique, specifically the Synthetic Minority Over-sampling Technique (SMOTE) variant known as SVMSMOTE, to address class imbalance in the dataset. The research utilizes the Pima Indians Diabetes dataset, a widely used benchmark comprising patient data from the Pima Indian ethnic group. The study also evaluates the effectiveness of integrating SVMSMOTE into the Support Vector Machine (SVM) algorithm to enhance model performance. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results demonstrate that applying SVMSMOTE significantly improves model performance, achieving an accuracy of 74.68%, precision of 60.26%, recall of 85.45%, and an F1-score of 70.61%. These findings suggest that SVMSMOTE is a viable solution for addressing data imbalance in predictive modelling for diabetes diagnosis.

Keywords: Diabetes, Pima Indians, Resampling, Preprocessing, SVMSMOTE.