

ABSTRAK

Klasifikasi adalah teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan pola yang telah dipelajari. Salah satu metode yang sering diterapkan dalam klasifikasi adalah Support Vector Machine (SVM), Support Vector Machine adalah pembelajaran terarah (supervised learning) yang bekerja dengan mencari hyperplane terbaik untuk memisahkan kelas data. Support Vector Machine memiliki kelemahan dalam hal sensitivitas terhadap pemilihan parameter kernel dalam menentukan kombinasi parameter yang optimal. Penelitian ini bertujuan mengklasifikasi penyakit jantung dengan menerapkan metode Support Vector Machine (SVM) yang dioptimalkan melalui grid search serta menerapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) dalam menangani data tidak seimbang. Hasil evaluasi penelitian ini dengan model Support Vector Machine dalam berbagai kondisi pada data penyakit jantung. Model Support Vector Machine tanpa SMOTE pada pembagian data 80:20 dan 70:30 menghasilkan akurasi tertinggi 85%. Penggunaan SMOTE memberikan peningkatan pada nilai AUC yaitu sebesar 0,908 dibandingkan dengan model tanpa SMOTE yang hanya mencapai AUC 0,643. Setelah dilakukan penerapan SMOTE hasil menjadi lebih optimal yang ditunjukkan dengan peningkatan nilai AUC dibandingkan dengan model tanpa SMOTE.

Kata kunci: Klasifikasi, Parameter, Support Vector Machine, SMOTE, Grid Search

ABSTRACT

Classification is a data mining technique used to group data based on learned patterns. One method frequently applied in classification is the Support Vector Machine (SVM). Support Vector Machines are supervised learning methods that work by finding the best hyperplane to separate data classes. Support Vector Machines have the disadvantage of being sensitive to kernel parameter selection when determining the optimal parameter combination. This study aims to classify heart disease by applying the Support Vector Machine (SVM) method optimized through grid search and applying the Synthetic Minority Oversampling Technique (SMOTE) to handle imbalanced data. The results of this study evaluated the Support Vector Machine model under various conditions in the heart disease data. The Support Vector Machine model without SMOTE, with data splits of 80:20 and 70:30, yielded the highest accuracy of 85%. The use of SMOTE resulted in an increase in the AUC value of 0.908 compared to the model without SMOTE, which only achieved an AUC of 0.643. After applying SMOTE, the results were more optimal, as indicated by the increase in the AUC value compared to the model without SMOTE.

Keywords: *Classification, Parameters, Support Vector Machine, SMOTE, Grid Search*