

ABSTRAK

Salah satu penyebab kematian paling umum di dunia adalah penyakit jantung. Deteksi dini penyakit jantung sangat penting untuk mengurangi risiko kematian dan meningkatkan kualitas hidup pasien. Tujuan penelitian ini adalah untuk membuat model klasifikasi penyakit jantung menggunakan metode *Support Vector Machine* (SVM) dengan teknik *Synthetic Minority Over-Sampling Technique* (SMOTE) dan menggunakan metode *Information Gain* untuk meningkatkan kinerja model. Dataset yang digunakan dalam penelitian ini terdiri dari beberapa fitur yang berisi gejala klinis yang terkait dengan kondisi jantung. Sebelum diklasifikasikan, data diproses melalui proses preprocessing seperti penanganan *missing value*, normalisasi, dan pembagian menjadi data latih dan data uji. *Information Gain* digunakan untuk memilih fitur yang paling relevan untuk meningkatkan akurasi model. Kemudian permodelan *Support Vector Machine* digunakan untuk menentukan kelas pasien yang beresiko dan tidak beresiko penyakit jantung. Hasil penelitian menunjukkan bahwa model terbaik diperoleh dari kombinasi antara metode *Support Vector Machine* (SVM) dengan Teknik SMOTE dan *Information Gain* menggunakan kernel RBF dengan parameter $C = 100$ dan $\gamma = 0.5$ pada pembagian data 90:10 dengan nilai *accuracy* 68%, *precision* 25%, *recall* 58%, dan F1-score 35%. Hasil tersebut menunjukkan bahwa SMOTE dan *Information Gain* dapat meningkatkan kemampuan model untuk mengidentifikasi kelas minoritas sehingga performa proses klasifikasi lebih seimbang.

Kata Kunci : Penyakit Jantung, *Support Vector Machine*, SMOTE, *Information Gain*.

ABSTRACT

One of the most common causes of death worldwide is heart disease. Early detection of heart disease is crucial to reduce the risk of death and improve patients' quality of life. The purpose of this study is to create a heart disease classification model using the Support Vector Machine (SVM) method with the Synthetic Minority Over-Sampling Technique (SMOTE) and the Information Gain method to improve model performance. The dataset used in this study consists of several features containing clinical symptoms associated with heart conditions. Before classification, the data was preprocessed, including handling missing values, normalization, and dividing into training and test data. Information Gain was used to select the most relevant features to improve model accuracy. Then, Support Vector Machine modeling was used to classify patients at and without risk for heart disease. The results showed that the best model was obtained from a combination of the Support Vector Machine (SVM) method with the SMOTE and Information Gain techniques using the RBF kernel with parameters $C = 100$ and $\gamma = 0.5$, with a 90:10 data split, resulting in 68% accuracy, 25% precision, 58% recall, and 35% F1-score. These results indicate that SMOTE and Information Gain can improve the model's ability to identify minority classes, resulting in a more balanced classification process performance.

Keywords: Heart Disease, Support Vector Machine, SMOTE, Information Gain.