

ABSTRACT

HYPERPARAMETER TUNING OPTIMIZATION USING GRIDSEARCH AND RANDOMSEARCH IN SUPPORT VECTOR MACHINE (SVM) METHOD FOR THE CLASSIFICATION OF HEART DISEASE PATIENT DATA

By

DINDA MEILANI ADITYA WATI

Heart disease is one of the most deadly Non-Communicable Diseases (NCD) in the world and is the leading cause of death in Indonesia. Efforts to early detection of heart disease are very important to increase the effectiveness of treatment and reduce the risk of death. In this study, the Support Vector Machine (SVM) method is used to classify data of heart disease patients. This study aims to optimize the performance of SVM models through hyperparameter tuning techniques using GridSearch and RandomSearch, as well as the application of data balancing methods, namely Random Oversampling. The data used is secondary data obtained from the Kaggle site, consisting of 918 patient data with 11 independent variables and 1 target variable. The research process includes data preprocessing stages (cleaning, feature selection, balancing, scaling), modeling with SVM, and model performance evaluation using accuracy, precision, recall, and f1-score metrics. The results show that hyperparameter tuning optimization using RandomSearch with Random Oversampling technique produces the best performance with an accuracy of 89.21%, compared to GridSearch and the default model which has an accuracy of 85.29%. Thus, RandomSearch is proven to be more effective in improving the performance of SVM classification models to detect heart disease.

Keywords: Support Vector Machine, Hyperparameter Tuning, GridSearch, RandomSearch, Heart Disease, Random Oversampling.

ABSTRAK

OPTIMASI HYPERPARAMETER TUNING MENGGUNAKAN GRIDSEARCH DAN RANDOMSEARCH PADA METODE SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI DATA PASIEN PENYAKIT JANTUNG

Oleh

DINDA MEILANI ADITYA WATI

Penyakit jantung merupakan salah satu penyakit tidak menular yang paling mematikan di dunia dan menjadi penyebab utama kematian di Indonesia. Upaya deteksi dini terhadap penyakit jantung sangat penting untuk meningkatkan efektivitas penanganan dan mengurangi risiko kematian. Dalam penelitian ini, digunakan metode Support Vector Machine (SVM) untuk melakukan klasifikasi data pasien penyakit jantung. Penelitian ini bertujuan untuk mengoptimalkan kinerja model SVM melalui teknik hyperparameter tuning menggunakan GridSearch dan RandomSearch, serta penerapan metode balancing data yaitu Random Oversampling. Data yang digunakan merupakan data sekunder yang diperoleh dari situs Kaggle, terdiri dari 918 data pasien dengan 11 variabel independen dan 1 variabel target. Proses penelitian meliputi tahap preprocessing data (cleaning, feature selection, balancing, scaling), pemodelan dengan SVM, serta evaluasi performa model menggunakan metrik akurasi, presisi, recall, dan f1-score. Hasil penelitian menunjukkan bahwa optimasi hyperparameter tuning menggunakan RandomSearch dengan teknik Random Oversampling menghasilkan performa terbaik dengan akurasi sebesar 89,21%, dibandingkan dengan GridSearch maupun model default yang memiliki akurasi sebesar 85,29%. Dengan demikian, RandomSearch terbukti lebih efektif dalam meningkatkan performa model klasifikasi SVM untuk mendeteksi penyakit jantung.

Kata kunci: Support Vector Machine, Hyperparameter Tuning, GridSearch, RandomSearch, Penyakit Jantung, Random Oversampling