

ABSTRACT

IMPLEMENTATION OF SUPPORT VECTOR MACHINE WITH RANDOM OVERSAMPLING TO OVERCOME IMBALANCE DATA IN CLASSIFYING PATIENTS WITH CARDIOVASCULAR DISEASE

BY

CANTIKA MERITA

Support Vector Machine (SVM) is one of the machine learning methods used for classification by dividing data into two different classes. The working principle of SVM is to find the best separating function (hyperplane). If the data cannot be separated linearly, then the data is nonlinear. One method to overcome this is by using a kernel function. The objective of this study is to apply the SVM method to determine the best kernel function based on the highest accuracy value for classifying patients with cardiovascular disease. However, the dataset used in this study has an imbalance data problem. Therefore, Random Oversampling (ROS) was used to address this issue. The results of the study indicate that the best kernel function for balanced data is the Radial Basis Function (RBF) function, with a gamma parameter of 0,1 and a cost of 1 in a 90% training and 10% testing data scheme, yielding an accuracy value of 73,86%.

Keywords : Cardiovascular; Kernel Radial Basis Function; Imbalance Data;
Random Oversampling; Support Vector Machine.

ABSTRAK

IMPLEMENTASI *SUPPORT VECTOR MACHINE* DENGAN *RANDOM OVERSAMPLING* UNTUK MENGATASI DATA TAK SEIMBANG PADA KLASIFIKASI PENDERITA PENYAKIT *CARDIOVASCULAR*

OLEH

CANTIKA MERITA

Support Vector Machine (SVM) adalah salah satu metode *machine learning* yang digunakan untuk pengklasifikasian dengan membagi data menjadi dua kelas yang berbeda. Prinsip kerja SVM adalah mencari fungsi pemisah (*hyperplane*) yang terbaik. Apabila data tidak dapat dipisahkan secara linear maka data tersebut merupakan data nonlinear. Salah satu metode untuk mengatasi hal tersebut adalah dengan menggunakan fungsi *kernel*. Tujuan penelitian ini adalah menerapkan metode SVM untuk mengetahui kinerja fungsi *kernel* terbaik berdasarkan nilai akurasi tertinggi terhadap klasifikasi penderita penyakit *cardiovascular*. Namun, pada dataset yang digunakan dalam penelitian memiliki masalah ketidakseimbangan data (*imbalance data*). Oleh karena itu, digunakan *Random Oversampling* (ROS) untuk mengatasi masalah tersebut. Hasil penelitian menunjukkan bahwa fungsi *kernel* terbaik pada data yang seimbang adalah fungsi *Radial Basis Function* (RBF), dengan parameter *gamma* 0,1 dan *cost* 1 pada skema data *training* 90% dan *testing* 10% didapat nilai akurasi sebesar 73,86%.

Kata Kunci : *Cardiovascular; Kernel Radial Basis Function; Ketidakseimbangan Data; Random Oversampling; Support Vector Machine.*