

ABSTRAK

PERBANDINGAN EMPIRIS FUNGSI *KERNEL* PADA METODE KLASIFIKASI *SUPPORT VECTOR MACHINE* DAN PENERAPANNYA PADA DATA PENDERITA PENYAKIT JANTUNG

Oleh

Dira Dini Dian Kemala

Support Vector Machine (SVM) adalah salah satu metode *machine learning* yang digunakan untuk mengklasifikasikan dengan membagi data menjadi dua kelas yang berbeda. Prinsip metode 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 fungsi *kernel*. Tujuan penelitian ini adalah mengetahui fungsi *kernel* terbaik yang menghasilkan akurasi tertinggi pada SVM melalui studi simulasi dan menerapkan metode SVM untuk klasifikasi penderita penyakit jantung. Hasil yang diperoleh, melalui studi simulasi bahwa fungsi *kernel* terbaik adalah fungsi *kernel Radial Basis Function* (RBF), proporsi *split* dataset yang terbaik adalah 80 *training* dan 20 *testing*, dan parameter terbaik adalah $cost = 1$, $degree = 1$, dan $gamma = 0,5$. Selain itu, penerapan metode SVM pada data penyakit jantung diperoleh hasil bahwa fungsi *kernel* RBF dengan *split* dataset 80 diperoleh nilai akurasi = 1, artinya dalam mengklasifikasikan seseorang apakah termasuk penderita atau tidak penderita dapat diklasifikasikan secara tepat. Dengan menggunakan fungsi *kernel* dan proporsi *split* dataset tersebut diperoleh parameter w dan b , yakni sebagai berikut: $w_{Age} = -9,43$, $w_{Gender(1)} = 4,22$, $w_{Gender(2)} = -4,22$, $w_{Chest\ Pain\ Type(1)} = 6,73$, $w_{Chest\ Pain\ Type(2)} = -0,92$, $w_{Chest\ Pain\ Type(3)} = -5,44$, $w_{Resting\ Blood\ Pressure} = -12,49$, $w_{Cholesterol} = 11,264$, $w_{Fasting\ Blood\ Sugar1} = -4,639$, $w_{Resting\ ECG1} = -3,769$, $w_{Resting\ ECG2} = -1,456$, $w_{MaxHR} = 19,413$, $w_{Exercise\ Angina1} = -9,583$, $w_{Oldpeak} = -25,042$, $w_{STSlope1} = -13,259$, $w_{STSlope2} = -1,762$, $b = 0,0734$.

Kata Kunci : *Support Vector Machine*; *Kernel Radial Basis Function*; Klasifikasi; Penyakit Jantung

ABSTRACT

Empirical Comparison of Kernel Function in Support Vector Machine Classification Method and its Application to Data of Heart Disease Patients

By

Dira Dini Dian Kemala

Support Vector Machine (SVM) is a machine learning method that is used to classify data by dividing data into two different classes. The principle of the SVM method is to find the best separator / hyperplane function. If the data cannot be separated linearly, then the data is nonlinear. One method to overcome this is by kernel function approach. This research aims to find the best kernel function that produces the highest accuracy in SVM through a simulation study and to apply the SVM method to classify people with heart disease. The results of simulation study showed that the best kernel function is the Radial Basis Function (RBF) kernel function, the best proportion of split datasets is 80 training and 20 testings, and the best parameters are cost = 1, degree = 1, and gamma = 0, 5. In addition, the application of the SVM method to heart disease data yields the result that the RBF kernel function with a split dataset of 80 obtains an accuracy value of = 1, meaning that in classifying a person, whether a person is a sufferer or not, a patient can be classified correctly. By using the kernel function and the split proportion of the dataset, the parameters w and b are obtained, namely as follows:

$w_{Age} = -9,43, w_{Gender(1)} = 4,22, w_{Gender(2)} = -4,22, w_{Chest\ Pain\ Type(1)} = 6,73,$
 $w_{Chest\ Pain\ Type(2)} = -0,92, w_{Chest\ Pain\ Type(3)} = -5,44, w_{Resting\ Blood\ Pressure} = -12,$
 $49, w_{Cholesterol} = 11,26, w_{Fasting\ Blood\ Sugar1} = -4,64, w_{RestingECG1} = -3,76,$
 $w_{RestingECG2} = -1,456, w_{MaxHR} = 19,413, w_{ExerciseAngina1} = -9.583, w_{Oldpeak} = -25,0$
 $42, w_{STslope1} = -13,259, w_{STslope2} = -1,762, b = 0,0734.$

Keywords: Support Vector Machine; Kernel Radial Basic Functions; Classification; Heart disease