

ABSTRACT

IMPLEMENTATION OF SUPPORT VECTOR MACHINE FOR CLASSIFICATION OF DIAGNOSIS OF BREAST CANCER PATIENTS

By

Elizabeth Ceasarina Sitompul

Support Vector Machine (SVM) is one of the machine learning methods used for classification problems by dividing data into two different classes through the formation of an optimal separating function called a hyperplane. The purpose of this study is to apply the SVM method and determine the best kernel function based on the highest accuracy value in the classification process. However, the dataset used in this study has a problem of data imbalance, so a special handling method is required. Therefore, the Random Oversampling (ROS) method is used to balance the data distribution before the classification process is carried out. In the classification process, several kernel functions were applied, namely linear, polynomial, sigmoid, and Radial Basis Function (RBF) kernels. The results of the study showed that the RBF kernel function provided the best performance compared to other kernels. In the data division scheme of 90% as training data and 10% as test data, with parameters $C = 10, 50, \text{ and } 100$ and $\gamma = 0.001$, the SVM model produced the highest accuracy value of 97.22%. These results indicate that the use of the RBF kernel is able to handle nonlinear data patterns more optimally. Using this kernel configuration, the SVM model parameters obtained, namely the \mathbf{w} and b , are as follows:

$\mathbf{w}_{\text{radius_mean}} = 377.7713$, $\mathbf{w}_{\text{texture_mean}} = 340.6203$, $\mathbf{w}_{\text{perimeter_mean}} = 2851.3717$,
 $\mathbf{w}_{\text{area_mean}} = 37057.1680$, \dots , $\mathbf{w}_{\text{fractal_dimension_worst}} = 2.8828$, $b = 0.5153$.

Keywords: Support Vector Machine, Radial Basis Function Kernel, Data Imbalance, Random Oversampling, Breast Cancer.

ABSTRAK

IMPLEMENTASI *SUPPORT VECTOR MACHINE* TERHADAP KLASIFIKASI DIAGNOSIS PENDEKITA KANKER PAYUDARA

Oleh

Elizabeth Ceasarina Sitompul

Support Vector Machine (SVM) merupakan salah satu metode *machine learning* yang digunakan untuk permasalahan klasifikasi dengan membagi data ke dalam dua kelas yang berbeda melalui pembentukan fungsi pemisah optimal yang disebut *hyperplane*. Tujuan penelitian ini adalah menerapkan metode SVM serta menentukan fungsi *kernel* terbaik berdasarkan nilai akurasi tertinggi pada proses klasifikasi. Namun, dataset yang digunakan dalam penelitian kali ini memiliki permasalahan ketidakseimbangan data (*imbalance data*), sehingga diperlukan metode penanganan khusus. Dengan demikian, digunakan metode *Random Oversampling* (ROS) untuk menyeimbangkan distribusi data sebelum proses klasifikasi dilakukan. Dalam Proses klasifikasi selanjutnya diterapkan menggunakan beberapa fungsi *kernel*, yaitu *kernel* linear, polinomial, sigmoid, dan *Radial Basis Function* (RBF). Hasil penelitian menunjukkan bahwa fungsi *kernel* RBF memberikan kinerja terbaik dibandingkan *kernel* lainnya. Pada skema pembagian data 90% sebagai data latih dan 10% sebagai data uji, dengan parameter $C = 10, 50, \text{ dan } 100$ serta $\gamma = 0.001$, model SVM menghasilkan nilai akurasi tertinggi sebesar 97,22%. Hasil ini menunjukkan bahwa penggunaan *kernel* RBF mampu menangani pola data nonlinear secara lebih optimal. Dengan menggunakan konfigurasi *kernel* tersebut, diperoleh parameter w dan b , yakni sebagai berikut:

$w_{\text{radius_mean}} = 377,7713$, $w_{\text{texture_mean}} = 340,6203$, $w_{\text{perimeter_mean}} = 2851,3717$,
 $w_{\text{area_mean}} = 37057,168$, \dots , $w_{\text{fractal_dimension_worst}} = 2,8828$, $b = 0,5153$.

Kata Kunci: *Support Vector Machine*, *Kernel Radial Basis Function*, Ketidakseimbangan Data, *Random Oversampling*, Kanker Payudara.