

ABSTRAK

Partial Least Square (PLS) merupakan metode pemodelan regresi, diskriminasi dan klasifikasi yang populer untuk analisis data kemometrik, industri dan produksi. PLS dapat digunakan ketika jumlah variabel yang diamati jauh lebih besar daripada jumlah pengamatan dan terdapat multikolinearitas antar variabel. Namun, PLS adalah metode linier yang tidak sesuai untuk menggambarkan karakteristik nonlinear. Untuk mengatasi masalah nonlinear, penelitian ini menggunakan *Least Square Support Vector Machine* (LSSVM) untuk memetakan hubungan kompleks dan nonlinier antara input dan output variabel. PLS dan LSSVM merupakan salah satu metode nonlinear yang digunakan untuk menyelesaikan permasalahan klasifikasi. Proses estimasi parameter yang ada pada metode PLS dan LSSVM menggunakan algoritma NIPALS. Metode ini akan diterapkan pada data penderita kanker limfoma atau *Diffuse Large B-Cell Lymphoma* (DLBCL) dengan jumlah 7399 ekspresi gen. Hasil penelitian memperlihatkan bahwa metode PLS dan LSSVM untuk pembagian data latih dan data uji sebesar 70% dan 30%, variabel tereduksi menjadi 70 komponen dan memiliki tingkat akurasi sebesar 97,22% dan sensitivitas sebesar 95,7%. Untuk pembagian data latih dan data uji sebesar 80% dan 20%, variabel tereduksi menjadi 40 komponen dan memiliki tingkat akurasi sebesar 89,584% dan sensitivitas sebesar 87,1%

Kata kunci: Algoritma *NIPALS*, Klasifikasi, *Least Square Support Vector Machine*, Nonlinier, *Partial Least Square*.

ABSTRACT

Partial Least Square (PLS) is a method of modeling, regression, differences and differences that are popular in its domain. PLS can be used when the number of variables observed is much larger and the number of multicollinearity between variables. However, PLS is a linear method that is not suitable for nonlinear features. To solve the nonlinear problem, this research uses Least Square Support Vector Machine (LSSVM) to map complex and nonlinear relationships between input and variable outputs. PLS and LSSVM is one of the nonlinear methods used to solve the classification problem. The process to estimate the parameter in the PLS and LSSVM methods is using NIPALS algorithm. This method will be applied to data of patients with lymphoma cancer or Diffuse Large B-Cell Lymphoma (DLBCL) with the total of 7399 gene expression. The results show that the PLS and LSSVM method for 70% training data and 30% test data, that variables reduce to 70 components and has an accuracy of 97,22% and the sensitivity of 95,7%. For training data 80% and 20% data test, the variables are reduce to 40 components and have an accuracy of 89,584% and sensitivity of 87,1%.

Keywords: *NIPALS algorithm, Classification, Least Square Support Vector Machine, Nonlinear, Partial Least Square.*