

INTISARI

Kernel Logistik Partial Least Squares (KL-PLS) untuk Reduksi Dimensi Nonlinear dan Klasifikasi Biner (Studi Kasus : Data Kanker Ovarium SELDI-TOF *High Resolution*)

Oleh :

Titik Buntariningsih

10/305419/PA/13511

Kanker ovarium merupakan kanker yang berkembang dengan sangat cepat. Perkembangan kanker dari stadium awal hingga stadium lanjut dapat terjadi dalam kurun waktu 1 tahun. Kanker ovarium merupakan penyebab kematian terbesar kelima di Indonesia. Pertumbuhan jumlah penderita kanker ini terus bertambah setiap tahunnya. Untuk itulah, diperlukan diagnosa dini kanker dibutuhkan untuk mengurangi resiko kematian. Sebuah studi yang dilakukan oleh Tenenhaus dkk. (2007) mengusulkan sebuah model yang memanfaatkan Kernel Logistik Partial Least Square (KLPLS) untuk memprediksi adanya kanker.

Skripsi ini bertujuan untuk memanfaatkan metode KLPLS untuk memprediksi penyakit kanker ovarium menggunakan data spektrometri massa ovarium SELDI-TOF *High Resolution* seperti yang telah diusulkan oleh (Tang dkk (2010). KLPLS sendiri merupakan pengembangan dari Partial Least Square dengan melibatkan fungsi kernel di dalam prosesnya. Fungsi kernel yang digunakan dalam implementasi metode ini adalah fungsi kernel polinomial. KLPLS terbagi ke dalam dua tahap. Tahap pertama adalah tahap latih untuk mencari koefisien regresi. Tahap kedua adalah tahap uji untuk memprediksi label data uji berdasarkan koefisien regresi dari tahap latih.

Uji coba dilakukan dengan memprediksi kanker pada empat dataset yang berbeda. Berdasarkan uji coba KLPLS menghasilkan rata-rata akurasi terbaik sebesar 93,53%, sensitivitas sebesar 89,81% dan spesifisitas sebesar 98,89% pada data Set A. Berdasarkan performa uji coba klasifikasi data, dapat ditarik kesimpulan bahwa model klasifikasi yang memanfaatkan KLPLS mampu melakukan klasifikasi kanker ovarium pada data ekspresi protein dalam format SELDI-TOF.

Kata Kunci : kanker ovarium, kernel, klasifikasi, KLPLS, penggalian data, reduksi dimensi, SELDI-TOF

ABSTRACT

Kernel Logistic Partial Least Squares (KL-PLS) for Nonlinear Dimensionality Reduction and Binary Classification (Case Study : Ovarian Cancer SELDI-TOF High Resolution Data)

By :

Titik Buntariningsih

10/305419/PA/13511

Ovarian cancer is a cancer that is growing very rapidly. Development of early-stage cancer to an advanced stage can occur within a period of one year. Ovarian cancer is the fifth leading cause of death in Indonesia. Growth in the number of cancer patients is growing every year. For that, we need early diagnosis of cancer are needed to reduce the risk of death. A study conducted by Tenenhaus et al. (2007) proposed a model that utilizes Logistics Kernel Partial Least Squares (KLPLS) to predict the presence of cancer.

This thesis aims to utilize KLPLS method for predicting ovarian cancer using mass spectrometry data is SELDI-TOF ovarian High Resolution as proposed by Tang et al. (2010). KLPLS it self a development of the Partial Least Square involving the kernel function in the process. Kernel function used in the implementation of this method is a polynomial kernel function. KLPLS divided into two stages. The first stage is the stage trained to look for the regression coefficients. The second stage is the stage of the test to predict the test data labels based on regression coefficient of practice stage.

The test is done to predict cancer in four different datasets. Based on trials KLPLS average yield of 93.53% best accuracy, sensitivity of 89.81% and a specificity of 98.89% in Data Set A. Based on the performance test data classification, it can be deduced that the classification model that utilizes KLPLS able to perform classification of ovarian cancer at the protein expression data SELDI-TOF format.

Keywords : ovarian cancer, kernel, classification, KLPLS, data mining, dimension reduction, SELDI-TOF