**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**USULAN TUGAS AKHIR**

# **IDENTITAS PENGUSUL**

Nama : **LISA YULI KURNIAWATI**

NRP : **5108 100 089**

Dosen Wali : **UMI LAILI YUHANA, S.KOM, M.Sc**

1. **JUDUL TUGAS AKHIR**

***“Implementasi Metode Kernel Partial Least Square untuk Klasifikasi Kanker Ovarium”***

1. **LATAR BELAKANG**

Ovarium adalah organ reproduksi wanita yang bentuknya menyerupai kacang dan tersimpan di dalam rongga perut sebelah bawah. Ovarium ini berjumlah dua buah (kiri dan kanan) dengan peranan sebagai tempat diproduksinya/dihasilkannya sel-sel telur(sel kelamin wanita). Ovarium merupakan organ terpenting dalam hal reproduksi wanita. Karenanya, wanita dapat mempunyai keturunan/anak. Sehingga ketika ovarium seorang wanita terganggu fungsinya maka secara otomatis semua proses reproduksi dan hidup wanita tersebut akan terganggu.

Kanker ovarium adalah suatu bentuk kanker yang menyerang ovarium. Kanker ini dapat berkembang sangat cepat. Bahkan dari stadium awal hingga stadium lanjut dapat dimungkinkan terjadi dalam kurun waktu 1 tahun. Kanker ovarium merupakan suatu proses lebih lanjut dari suatu tumor malignan di ovarium. Tumor malignan itu sendiri merupakan suatu bentuk perkembangan sel-sel yang tidak terkontrol sehingga berpotensi menjadi kanker.

Dari seluruh penjuru dunia, penyakit kanker ovarium merupakan penyakit di urutan kedelapan yang paling banyak menyerang wanita. Pertumbuhan jumlah penderita penyakit ini pun terus bertambah setiap tahunnya. Kanker Ovarium juga disebut sebagai *silent killer* atau pembunuh bisu. Hal ini karena kanker ini menjangkiti wanita tanpa disadari ditahap awal dan ketika terdeteksi, kanker sudah berada di stadium lanjut yang mematikan. Penyebab kematian terbanyak dari semua kanker ginekologi adalah kanker ovarium. Fakta ini sangat mengkhawatirkan karena banyak sekali kasus kanker ovarium baru terdeteksi setelah stadium lanjut. Tindakan yang biasa dilakukan bagi penderita kanker ovarium jika kanker belum menyebar ke luar ovarium, maka dilakukan pengangkatan ovarium yang terkena dan mungkin dengan *tuba falopii*nya (saluran indung telur). Namun jika kanker telah menyebar ke luar ovarium, maka dilakukan pengangkatan kedua ovarium dan rahim, serta kelenjar getah bening dan struktur di sekitarnya. Kemudian setelah pembedahan bisa dilakukan terapi penyinaran dan kemoterapi untuk menghancurkan sisa-sisa sel kanker. Disinilah pentingnya pendeteksian dini kanker ovarium, sehingga dapat mengurangi resiko yang timbul setelahnya.

Kemajuan terbaru dalam teknologi proteomik memungkinkan dilakukannya analisis skala besar pada pola ekspresi protein kompleks, protein-protein interaksi dan modifikasi posttranslational. Salah satu teknologinya adalah Surface-Enhanced Laser Desorption/Ionization Time Of Flight (SELDI-TOF) yakni teknologi untuk mendeteksi protein (peptida) dari sampel jaringan atau darah. Sebuah array protein berupa matriks dari suatu 'tempat', dimana ‘tempat’ yang dimaksud adalah analog dengan data yang diperoleh dari hibridisasi sampel yang unik pada seluruh DNA *microarray*. DNA *microarray* adalah teknologi yang digunakan untuk melihat urutan sekuens asam nukleat yang berada pada lokasi tertentu dan dapat digunakan untuk menganalisa beribu-ribu sampel pada waktu yang bersamaan. Dengan adanya DNA *microarray* ini memudahkan pakar medis untuk melakukan klasifikasi kanker.DNA *microarray* ini kemudian direpresentasikan menjadi vektor. Vektor hasil representasi inilah yang kemudian diolah dan diklasifikasi dengan metode-metode yang sudah ada.

Banyak metode klasifikasi data yang sudah berkembang, diantaranya CART(classification and regression tree)[1] , *genetic algorithm*/*KNN*[2], *recursive support vector machine* (R-SVM)[3] dan lainnya. Salah satu metode yang unggul dalam penanganan dimensi data yang tinggi adalah Partial Least Square (PLS). PLS adalah metode regresi untuk memprediksi nilai Y dari nilai X yang diketahui.

Perkembangan dari metode PLS adalah KPLS. KPLS (*Kernel Partial Least Square*)[4] mampu melakukan diskriminasi secara handal antara kanker dan sampel sehat. Untuk meningkatkan performa klasifikasi kanker ovarium dapat menggunakan metode berdasar pada momen statistik untuk melakukan transformasi data. Empat buah momen statistik (*mean*, *variance*, *skewness* dan *kurtosis*) digunakan untuk mendeskripsikan data dan mengurangi dimensi fitur dan ekstraksi karakteristik data pada waktu yang bersamaan. Pengurangan dimensi data dan klasifikasi dapat dilakukan secara bersamaan dengan menggunakan KPLS.

Tugas akhir ini bertujuan mengimplementasikan metode KPLS untuk klasifikasi kanker ovarium. Implementasi ini kemudian diharapkan dapat bermanfaat di bidang kedokteran dalam diagnosa awal penyakit kanker ovarium.

1. **TUJUAN**

Tugas akhir ini bertujuan mengimplementasikan metode KPLS (*Kernel Partial Least Square*) yang kemudian dapat digunakan dalam kasus klasifikasi penyakit kanker ovarium.

1. **MANFAAT TUGAS AKHIR**

Tugas akhir ini dikerjakan dengan harapan dapat memberikan manfaat yang besar

pada bidang kedokteran dalam mendiagnosa penyakit kanker ovarium lebih dini.

1. **RUMUSAN MASALAH**

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat beberapa permasalahan yang dapat dirumuskan sebagai berikut:

1. Bagaimana melakukan representasi data *DNA Microarray* menjadi sebuah vektor sehingga dapat diolah dengan lebih mudah.
2. Bagaimana memodelkan KPLS pada data *DNA Microarray* dalam kasus klasifikasi penyakit kanker ovarium.
3. Bagaimana melakukan evaluasi model KPLS untuk memprediksi penyakit kanker ovarium.
4. **BATASAN MASALAH**

Permasalahan yang dibahas dalam Tugas Akhir ini memiliki beberapa batasan, antara lain sebagai berikut :

1. Sistem perangkat lunak klasifikasi kanker ovarium ini dibangun dengan menggunakan perangkat lunak MATLAB 7.6.
2. Dataset yang digunakan pada Tugas Akhir ini yaitu data microarray ovarian SELDI-TOF[5].
3. **RINGKASAN TUGAS AKHIR**

Pembuatan Tugas Akhir ini dikerjakan dalam beberapa modul. Modul-modul itu antara lain:

* ***Partial Least Square* (PLS)**

Partial least square (PLS)[6] adalah suatu tekhnik [statistik](http://statistikakomputasi.wordpress.com/) multivariat yang bisa menangani banyak variabel respon dan variabel eksplanatori sekaligus. PLS merupakan alternatif yang baik untuk metode analisis regresi berganda dan regresi komponen utama karena metode PLS bersifat lebih robust, artinya parameter model tidak banyak berubah ketika sampel baru diambil dari total populasi.

PLS pertama kali dikembangkan pada tahun 1960-an oleh Herman O. A. Wold dalam bidang ekonometrik. PLS merupakan suatu tekhnik prediktif yang bisa menangani banyak variabel independen, bahkan sekalipun terjadi multikolinieritas diantara variabel-variabel tersebut.

Adapun langkah-langkah metode PLS antara lain:

Langkah 1

Merancang inner model. Inner Model atau Model Struktural menggambarkan hubungan antar variabel laten berdasarkan pada substantive theory. Perancangan Model Struktural hubungan antar variabel laten didasarkan pada rumusan masalah atau hipotesis penelitian.

Langkah 2

Merancang outer model. Outer Model atau Model Pengukuran mendefinisikan bagaimana setiap blok indikator berhubungan dengan variabel latennya. Perancangan Model Pengukuran menentukan sifat indikator dari masing-masing variabel laten, apakah refleksif atau formatif, berdasarkan definisi operasional variabel.

Langkah 3

Konversi diagram jalur ke persamaan

* + - * 1. Model persamaan dasar dari Inner Model dapat ditulis sebagai berikut:

Ŋ = β0 + βŋ + Гξ + ζ

Ŋj = Σi βji ŋi + Σi үjb ξb + ζj

b. Model persamaan dasar Outer Model dapat ditulis sebagai berikut:

X = Λx ξ + εx Y = Λy ŋ + εy

Langkah 4

Estimasi : Weight, Koefisien Jalur, dan Loading

Metode pendugaan parameter (estimasi) di dalam PLS adalah metode kuadrat terkecil (least square methods). Proses perhitungan dilakukan dengan cara iterasi, dimana iterasi akan berhenti jika telah tercapai kondisi kenvergen. Pendugaan parameter di dalam PLS meliputi 3 hal, yaitu:

* Weight estimate yang digunakan untuk menghitung data variabel laten.
* Path estimate yang menghubungkan antar variabel laten dan estimasi loading antara variabel laten dengan indikatornya.
* Means dan parameter lokasi (nilai konstanta regresi, intersep) untuk indikator dan variabel laten.

Langkah 5

Goodness of Fit Model diukur menggunakan R2 variabel laten dependen dengan interpretasi yang sama dengan regresi. Q2 predictive relevance untuk model struktural mengukur seberapa baik nilai observasi dihasilkan oleh model dan juga estimasi parameternya.   
Q2 = 1 – ( 1 - R12 ) ( 1 – R22 ) … (1 – Rp2)

Langkah 6

Pengujian Hipotesis (β, ү, dan λ) dilakukan dengan metode resampling Bootstrap yang dikembangkan oleh Geisser & Stone. Statistik uji yang digunakan adalah statistik t atau uji t. Penerapan metode resampling, memungkinkan berlakunya data terdistribusi bebas (distribution free) tidak memerlukan asumsi distribusi normal, serta tidak memerlukan sampel yang besar (direkomendasikan sampel minimum 30). Pengujian dilakukan dengan t-test, bilamana diperoleh p-value.

* ***Kernel Partial Least Square* (KPLS)**

KPLS, sama seperti halnya SVM, adalah sebuah metode pembelajaran berdasarkan pada fungsi kernel. Rosipal[7] memperkenalkan KPLS (*Kernel Partial Least Square*) sebagai sebuah pengembangan metode PLS. Metode KPLS dapat digunakan untuk klasifikasi dan regresi.

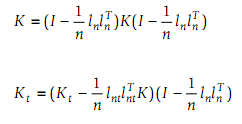
KPLS terdiri dari kontruksi matriks kernel dan PLS. Perbedaan antara KPLS dan PLS adalah XXT digantikan oleh sebuah matriks kernel. Hubungan antara variabel yang ada digantikan oleh matriks kernel.

Prinsip dasar dari sebuah kernel adalah fitur ruang  dimana item-item data asli diletakkan pada ruang tersebut. Dalam ruang ini relasi linier dieksplorasi antara *image* dari item data. Namun, fitur *image*  tidak perlu dan juga tidak mungkin untuk didapatkan. Dengan demikian, inner produk berpasangan dihitung dari item data asli menggunakan fungsi kernel untuk menunjukkan hubungan yang implisit.

Diberikan sebuah kumpulan vector X = (x1, . . . n) dimana n adalah jumlah sampel dan K menunjuk pada fungsi kernel. K didefinisikan sebagai matriks kernel yang mana entri dari kernel dihasilkan oleh pola seperti:



Matriks kernel untuk data training adalah simetrik dan persegi. Untuk sentralisasi data dipetakan dalam ruang fitur F dengan menerapkan prosedur berikut:



(Matriks kernel untuk data training)

(Matriks kernel untuk data testing)

Dimana masing-masing I adalah sebuah n-dimensi matrik identitas dan ln mewakili vektor yang elemen-elemennya adalah 1 panjang n dan nt. Kt adalah matriks uji berukuran (nt x n) yang elementnya berupa:



dimana adalah point data testing dan training.

Sebelum mengunakan KPLS, identifikasi fungsi kernel yang tepat dan parameter kernel yang optimal adalah hal yang sangat penting. Fungsi kernel yang umum digunakan termasuk linier, polinomial, dasar radial dan fungsi sigmoid. Fungsi kernel yang berbeda dipilih untuk masalah yang berbeda. Saat ini, tidak ada metodologi yang sistematis untuk pemilihan fungsi kernel. Dalam Tugas Akhir ini, digunakan fungsi polynomial sebagai beriku:.



Parameter p menunjukkan derajat dari polinomial dan didefinisikan sebagai p = 3 dalam percobaan. Parameter r menunjukkan bobot relatif yang mononomials tingkat yang berbeda. Berikut r adalah tetap sebagai 1 untuk kesederhanaan di bawah asumsi bahwa semua himpunan bagian sesuai dengan mononomial masing-masing membuat kontribusi yang sama dengan kernel.

1. **METODOLOGI**

Metodologi yang akan dilakukan dalam Tugas Akhir ini memiliki beberapa tahapan, diantaranya sebagai berikut:

1. Penyusunan Proposal Tugas Akhir

Tahap awal untuk memulai pengerjaan Tugas Akhir adalah penyusunan Proposal Tugas Akhir. Pada proposal ini, penulis mengajukan gagasan metode klasifikasi yang dapat diimplementasikan pada bidang bioinformatika (dalam hal ini deteksi penyakit kanker ovarium melalui dataset *microarray*).

1. Studi Literatur

Pada tahap ini dilakukan pencarian, pengumpulan, penyaringan, pembelajaran dan pemahaman literatur yang berhubungan dengan metode klasifikasi KPLS (*Kernel Partial Least Square*) dan pengembangannya, dataset *DNA Microarray* dan pengolahannya, serta pengaplikasian metode klasifikasi pada dataset. Literatur yang digunakan dalam pengerjaan Tugas Akhir ini sebagian besar berasal dari internet berupa makalah ilmiah, tesis, artikel, materi kuliah, serta beberapa buku referensi.

1. Implementasi

Implementasi merupakan tahap untuk membangun sistem tersebut.

1. Pengujian dan Evaluasi

Pada tahap ini dilakukan uji coba terhadap sistem yang telah dibuat, mengamati kinerja sistem yang baru dibuat, serta mengidentifikasi kendala yang mungkin timbul.

5. Penyusunan Buku Tugas Akhir

Tahap terakhir merupakan penyusunan laporan yang memuat dokumentasi mengenai pembuatan serta hasil dari implementasi perancangan dan *prototype* yang telah dibuat.

1. **JADWAL PEMBUATAN TUGAS AKHIR**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No. | Kegiatan | 2011 | | | | | | | | | | | | | | | | 2012 | | | |
| September | | | | Oktober | | | | Nopember | | | | Desember | | | | Januari | | | |
| 1 | Penyusunan Proposal Tugas Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Perancangan Perangkat Lunak |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Implementasi Perangkat Lunak |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Pengujian dan Evaluasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Penyusunan Buku Tugas Akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. **DAFTAR PUSTAKA**

[1]Vlahou A, Schorge JO, Gregory BW, Coleman RL: **Diagnosis of ovarian cancer using decision tree classification of mass spectral data**. *Journal of Biomedicine and Biotechnology*  2003, 2003:308-314.

[2]Li L, Umbach DM, Terry P, Taylor JA: **Application of the GA/KNN method to SELDI proteomics data**. *Bioinformatics* 2004, 20(10):1638-1640.

[3]Zhang X, Lu X, Shi Q, Xu XQ, Leung HC, Harris LN, Iglehart JD, Miron A, Liu JS, Wong WH: **Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data**. *BMC Bioinformatics* 2006, 7:197.

[4]Tang Kai-Lin, Li Tong-Hua, Xiong Wen-Wei, Chen Kai: **Ovarian cancer classification based on dimensionality reduction for SELDI-TOF data***. BMC Bioinformatics*: 2010, 11:109.

[5]SELDI-TOF, **http://home.ccr.cancer.gov/ncifdaproteomics/ppatterns.asp** (National Cancer Institute)

[6]Muliati Niswah: **Analisis Data Menggunakan Partial Least Square(PLS)***. http://annisty.blogspot.com/2010/01/analisis-data-menggunakan-partial-least.html*: diakses pada 18 Oktober 2011.

[7]Rosipal R, Trejo LJ: **Kernel partial least** **squares regression in Reproducing Kernel Hilbert Space**. J Mach Learn Res 2002, 2(2):97-123.

**LEMBAR PENGESAHAN**

###### Surabaya, 18 Oktober 2011

Menyetujui,

|  |  |  |
| --- | --- | --- |
| Dosen Pembimbing I,  **(Prof. Ir. Handayani Tjandrasa, M.Sc, Ph.D)**  **( NIP. 19490823 197603 2001 )** | Dosen Pembimbing II,  **(Isye Arieshanti, S.Kom. M. Phil.)**   |  | | --- | | **(NIP. 19780412 200604 2001**) | |