

IMPLEMENTASI METODE MULTIPLE KERNEL SUPPORT VECTOR MACHINE UNTUK SELEKSI FITUR DARI DATA EKSPRESI GEN DENGAN STUDI KASUS LEUKIMIA DAN TUMOR USUS BESAR

Ariana Yunita¹
Chastine Fatichah²
Umi Laily Yuhana³

Jurusan Teknik Informatika, Fakultas Sains dan Teknologi
Universitas Islam Maulana Malik Ibrahim Malang¹
Jurusan Teknik Informatika Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember Surabaya^{2,3}

Abstrak - Pada penelitian ini mengimplementasikan metode *multiple kernel support vector machine* untuk seleksi fitur. *Multiple kernel* merupakan metode modifikasi fungsi kernel yang mengalikan tiap elemen dari data. Metode ini melakukan seleksi fitur terhadap fitur yang kurang penting dengan tingkat akurasi lebih baik daripada metode dasar *support vector machine*. Uji coba dilakukan dengan menggunakan dataset ekspresi gen leukimia dan tumor usus besar. Hasil uji coba dibandingkan dengan tingkat akurasi metode *support vector machine* tanpa seleksi fitur. Tingkat akurasi metode *multiple kernel support vector machine* yang dihasilkan untuk data ekspresi gen leukimia yaitu 85% dan untuk data tumor usus besar sebesar 69%. Sedangkan tingkat akurasi dengan metode dasar *support vector machine* yaitu sebesar 82% untuk data leukimia dan 59% untuk data tumor usus besar. Seleksi fitur dapat mempersingkat waktu komputasi sehingga dapat dikembangkan untuk banyak aplikasi pengenalan pola.

Kata Kunci: *Multiple kernel, support vector machine*, seleksi fitur, data ekspresi gen

1. PENDAHULUAN

Kanker merupakan penyakit yang mematikan di dunia dan jumlah penderitanya terus bertambah dari tahun ke tahun. Klasifikasi kanker terus berkembang selama 30 tahun terakhir ini, tetapi belum ada pendekatan umum untuk mengidentifikasi jenis kanker. Para ilmuwan meneliti berbagai metode untuk menemukan pendekatan terbaik dalam hal klasifikasi kanker. Salah satu diagnosa kanker melalui yaitu melalui ekspresi gen.

Ekspresi gen merupakan proses dari informasi yang diwariskan dari gen, seperti urutan DNA, dibuat untuk produk

gen seperti protein atau RNA. Data ekspresi gen dapat dikatakan sebagai representasi urutan gen. Kanker pada dasarnya merupakan kesalahan pada gen, sehingga dengan menggunakan data ekspresi gen dapat menjadi cara untuk mendiagnosa kanker. Permasalahan utama dalam hal klasifikasi kanker adalah jumlah data ekspresi gen yang sedikit sedangkan jumlah fitur sangat banyak. Sehingga perlu dilakukan seleksi fitur.

Algoritma SVM (*Support Vector Machine*) selama ini dipercaya sebagai metode ampuh untuk menganalisis data biologi. Metode *Multiple Kernel Support Vector Machine* merupakan perbaikan

dari metode *Support Vector Machine*. Kedua metode ini akan dibandingkan tingkat akurasinya.

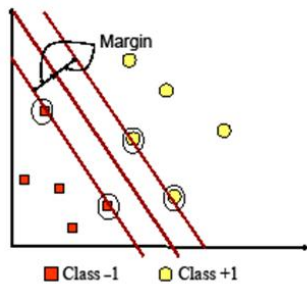
Pada bagian 2 akan *dibahas* mengenai pengantar SVM. Bagian 3 dijelaskan tentang metode *multiple kernel support vector machine*. Bagian 4 dijelaskan tentang algoritma dan arsitektur sistem. Bagian 5 dijelaskan tentang hasil uji coba dan analisis. Kemudian pada bagian 6 dijelaskan mengenai kesimpulan.

2. DASAR TEORI

2.1 Sekilas tentang SVM

Pada bagian ini akan dijelaskan secara singkat konsep dasar dari SVM untuk permasalahan klasifikasi 2 kelas. Konsep dasar SVM yaitu bertujuan untuk menemukan *hyperplane* terbaik sebagai pemisah dua buah kelas. Seperti gambar 1 berikut ini yang menunjukkan pemisahan 2 kelas yaitu kelas -1 dan kelas +1 dengan metode SVM.

Usaha untuk mencari lokasi *hyperplane* merupakan inti dari proses pembelajaran pada SVM. Data yang tersedia dinotasikan sebagai $x_i \in \mathbb{R}^D$ sedangkan label masing-masing dinotasikan $y_i \in \{-1,+1\}$ untuk $i = 1,2,\dots,l$, yang mana l adalah banyaknya data.



Gambar 1. Klasifikasi 2 kelas dengan metode SVM

Diasumsikan kedua kelas -1 dan +1 dapat terpisah secara sempurna oleh *hyperplane* berdimensi d , yang didefinisikan

$$\bar{w} \cdot \bar{x} + b = 0 \quad (1)$$

Pattern \bar{x}_i yang termasuk kelas -1 (sampel negatif) dapat dirumuskan sebagai *pattern* yang memenuhi pertidaksamaan

$$\bar{w} \cdot \bar{x}_i + b \leq -1 \quad (2)$$

sedangkan *pattern* \bar{x}_i yang termasuk kelas +1 (sampel positif)

$$\bar{w} \cdot \bar{x}_i + b \geq 1 \quad (3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara *hyperplane* dan titik terdekatnya, yaitu $\frac{1}{\|\bar{w}\|}$

Hal ini dapat dirumuskan sebagai Permasalahan Pemrograman Kuadratik (QP), yaitu mencari titik minimal persamaan 4, dengan memperhatikan batasan tersebut

$$\min_{\bar{w}} \tau(\bar{w}) = \frac{1}{2} \|\bar{w}\|^2$$

$$y_i (\bar{w} \cdot \bar{x}_i + b) - 1 \geq 0, \forall i \quad (4)$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya Lagrange Multiplier.

$$L(\bar{w}, b, \alpha) = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i (\bar{w} \cdot \bar{x}_i + b) - 1)$$

$$(i = 1, 2, \dots, l) \quad (5)$$

α_i adalah Lagrange multipliers, yang bernilai nol atau positif ($\alpha_i \geq 0$). Nilai optimal dari persamaan 5 dapat dihitung dengan meminimalkan L terhadap \bar{w} dan b , dan memaksimalkan L terhadap α_i . Dengan memperhatikan sifat bahwa pada titik optimal gradient $L = 0$, persamaan 5 dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung α_i saja, sebagaimana persamaan 6 di bawah ini.

$$\sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j$$

$$\alpha_i \geq 0 \quad (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (6)$$

Dari hasil dari perhitungan ini diperoleh α_i yang kebanyakan bernilai positif. Data yang berkorelasi dengan α_i yang positif disebut dengan *support vector*.

Penjelasan di atas berdasarkan asumsi bahwa kedua kelas dapat terpisah secara sempurna oleh *hyperplane*. Akan tetapi, umumnya dua buah kelas pada

ruang input tidak dapat terpisah secara sempurna. Hal ini menyebabkan *constraint* pada persamaan 3 tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan. Untuk mengatasi masalah ini, SVM dirumuskan ulang dengan memperkenalkan teknik *soft margin*. Dalam *soft margin*, persamaan 3 dimodifikasi dengan memasukkan *slack variabel*

$$\xi_i (\xi_i > 0) \text{ sebagai berikut}$$

$$y_i (\bar{w} \cdot \bar{x}_i + b) \geq 1 - \xi_i, \forall i \quad (7)$$

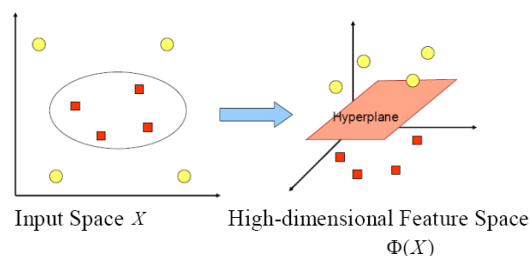
Dengan demikian persamaan menjadi sebagai berikut

$$\min_{\bar{w}} \tau(\bar{w}, \xi) = \frac{1}{2} \|\bar{w}\|^2 + C \sum \xi_i \quad (8)$$

Paramater C dipilih untuk mengontrol *tradeoff* antara margin dan error klasifikasi ξ . Nilai C yang besar berarti akan memberikan penalti yang lebih besar terhadap *error* klasifikasi tersebut.

2.2 Fungsi Kernel

Pada umumnya masalah dalam domain dunia nyata jarang yang bersifat *linear separable*. Kebanyakan bersifat non linear. Untuk menyelesaikan problem non linear, SVM dimodifikasi dengan memasukkan fungsi kernel. Dalam non linear SVM, data \bar{x} dipetakan oleh fungsi $\phi(\bar{x})$ ke ruang vektor yang berdimensi lebih tinggi. Pada ruang vektor yang baru ini, *hyperplane* yang memisahkan kedua kelas tersebut dapat dikonstruksikan. Ilustrasi dari konsep ini dapat dilihat pada gambar 2.



Gambar 2. Fungsi memetakan Φ memetakan data ke ruang vektor yang berdimensi lebih tinggi

Beberapa fungsi kernel yang umum dipakai yaitu fungsi kernel linear, polynomial dan gaussian. Ringkasan beberapa fungsi kernel tersebut terdapat pada tabel 2.1

Tabel 2.1 Kernel yang Umum Digunakan di SVM

Jenis Kernel	Definisi
Linear	$K(\bar{x}_i, \bar{x}_j) = (\bar{x}_i, \bar{x}_j)$
Polynomial	$K(\bar{x}_i, \bar{x}_j) = (\bar{x}_i, \bar{x}_j + 1)^p$
Gaussian	$K(\bar{x}_i, \bar{x}_j) = \exp(-\frac{\ \bar{x}_i - \bar{x}_j\ ^2}{2\sigma^2})$

Dengan memasukkan fungsi kernel pada persamaan optimasi SVM, maka persamaan menjadi

$$\max \{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(\bar{x}_i, \bar{x}_j) \}$$

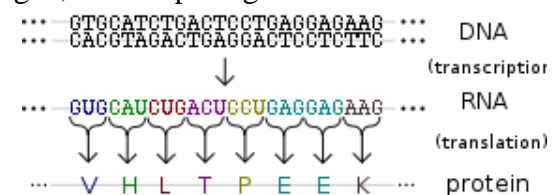
$$\text{s.t.} \quad \sum_{i=1}^n \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C \quad i = 1, \dots, n$$

$$\beta_d \geq 0 \quad i = 1, \dots, m \quad (9)$$

2.3 Data Ekspresi Gen

Kanker pada dasarnya merupakan kesalahan pada gen sehingga dengan menggunakan data ekspresi gen dapat menjadi cara untuk mendiagnosa. Yang dimaksud dengan ekspresi gen merupakan proses dari informasi yang diwariskan dari gen, seperti urutan DNA, dibuat untuk produk gen seperti protein atau RNA. Secara umum, proses pembentukan protein yaitu berasal dari DNA. Dari DNA lalu di transkripsi menjadi RNA. Dari RNA mengalami proses translasi menjadi protein. Berikut ini merupakan gambaran dari ekspresi gen, terlihat pada gambar 2.3



Gambar 3. Ekspresi Gen dari DNA menjadi Protein

2.3.1 Human Acute Leukemia

Dataset

Data pada eksperimen ini berasal dari studi yang dilakukan oleh Golub, dan tersedia online di internet. Data diambil dari 72 pasien penderita *myeloid leukemia* (AML) dan *acute lymphoblastic leukemia* (ALL). Data ini dibagi dalam dua kelompok: *training set* (27 ALL dan 11 AML), dan *test set* (20 ALL dan 14 AML). Tiap sampel terdiri dari vektor berdimensi 7129 yang berasal dari ekspresi gen si pasien sebagai hasil analisis *Affymetrix high-density oligonucleotide microarray*.

2.4.1 Colon Tumor Dataset

Data pada eksperimen ini berasal dari studi yang dilakukan oleh Alon, yang terdiri dari 40 *tissue tumor* dan 22 *tissue normal colon*. Tiap sampel berasal dari hasil analisis *Affymetrix oligonucleotide arrays* terhadap 6500 gen manusia. Dari 6500 gen ini, sebanyak 2000 diantaranya yang diseleksi terlebih dahulu berdasarkan kriteria tertentu, dipergunakan untuk keperluan klasifikasi.

3. METODE

3.1 MK SVM I untuk Seleksi Fitur

Metode ini berusaha mengoptimalkan klasifikasi dengan cara mengubah fungsi kernel. Fungsi kernel yang diajukan pada Multiple Kernel SVM yaitu

$$k(\bar{x}_i, \bar{x}_i) = \sum_{d=1}^m \beta_d k(x_{i,d}, x_{i,d}) \quad (10)$$

dimana β_d merupakan bobot dari single feature kernel. Dengan menggunakan parameter β_d kita dapat melihat fitur yang relevan untuk training. Dalam hal ini, β_d yang bernilai 0 akan dianggap sebagai fitur yang tidak relevan, sehingga tidak dimasukkan sebagai data training.

Pada saat *single feature kernel* dimasukkan ke perhitungan kernel, maka persamaan optimasi berubah menjadi

$$\max \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \sum_{d=1}^m \beta_d k(x_{i,d}, x_{j,d}) \right\}$$

$$\begin{aligned} \text{s.t} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n \\ & \beta_d \geq 0 \quad i = 1, \dots, m \end{aligned} \quad (11)$$

Hal ini berbeda dengan permasalahan SVM biasa, pada kasus multiple kernel SVM, α dan β butuh dioptimasi. Oleh karena itu, permasalahan tersebut didekomposisi ke dalam dua permasalahan optimasi *convex* sederhana. α dapat dihitung dengan nilai β yang sudah pasti. Sedangkan nilai β dapat dihitung setelah didapatkan nilai α . Sehingga langkah pertama yaitu dihitung nilai α , lalu selanjutnya dihitung nilai β dari persamaan optimasi *dual linear programming*.

3.2 MK II SVM untuk Pemodelan Prediksi

Pada MK SVM II digunakan kernel seperti pada persamaan 10, dengan kondisi

$$\beta_d > 0 \text{ dan } \sum_{d=1}^m \beta_d = 1$$

Selain itu diasumsikan bahwa $\gamma_{j,d} = \alpha_j \beta_d$

Persamaan optimasi pada MK SVM II menjadi

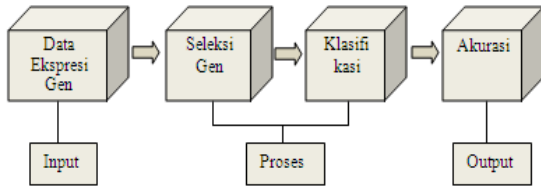
$$\min J(\gamma^-, \xi^-) = \sum_{i,d} \gamma_{i,d} + \lambda \sum_{i=1}^n \xi_i$$

$$\begin{aligned} \text{st} \quad & \gamma_i (\sum_{j,d} \gamma_{j,d} y_j k(x_{i,d}, x_{j,d}) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, \dots, n \\ & \gamma_{j,d} \geq 0, d = 1, \dots, m \end{aligned} \quad (12)$$

$\gamma_{j,d}$ disebut sebagai *mixture coefficient* dan dapat menunjukkan jumlah *support vector* yang ada. Menghitung *support vector* dengan syarat apabila dalam tiap baris γ terdapat nilai nonzero.

4. PERANCANGAN SISTEM

Secara garis besar, sistem tergambar dalam blok diagram di bawah ini.



Gambar 4. Blok Diagram Sistem Multiple Kernel SVM

Berikut ini akan dijelaskan mengenai algoritma dan perancangan sistem dari Multiple Kernel SVM.

4.1 Seleksi Fitur

Seleksi gen bertujuan untuk membuang gen-gen yang tidak informatif. Gen yang terpilih ada gen yang memiliki nilai β tidak 0. Pada fase ini menggunakan algoritma MK-SVM I. Algoritma MK-SVM I yaitu sebagai berikut :

Input : data vektor $\bar{x}_i \in \mathbb{R}^m$ dan $y_i \in \{+1, -1\}$

1. Inisialisasi : parameter $C = 1$ dan koefisien fitur $\beta_d = 1$.

2. While $\left| 1 - \sum_{d=1}^m \frac{\beta_d^{(t)}}{\beta_d^{(t-1)}} \right| \geq \varepsilon$

a. Menghitung koefisien Lagrange α_j dengan menggunakan pemrograman kuadratik berdasarkan pada persamaan di bawah ini

$$\begin{aligned} & \max \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \alpha_j \alpha_j y_j y_i \sum_{d=1}^m \beta_d k(x_{i,d}, x_{j,d}) \right\} \\ & \text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n \\ & \beta_d \geq 0 \quad d = 1, \dots, m \quad (13) \end{aligned}$$

b. Menghitung koefisien fitur β_d dengan menggunakan *dual linear programming* berdasarkan persamaan di bawah ini

$$\begin{aligned} & \max \sum_{i=1}^n u_i^{(t)} \\ & \text{s.t.} \\ & \sum_{i=1}^n u_i^{(t)} y_i \left(\sum_{j=1}^n \alpha_j y_j k(x_{i,d}, x_{j,d}) \right) \leq 1 \\ & \sum_{i=1}^n u_i^{(t)} y_i = 0 \\ & 0 \leq u_i^{(t)} \leq \lambda, \quad i = 1, \dots, n \quad (14) \end{aligned}$$

c. Selesai

3. Evaluasi parameter dan kembali ke langkah 2 sampai dengan optimal.

Output : koefisien fitur β_d . Fitur yang terpilih adalah fitur yang memiliki nilai koefisien fitur tidak 0.

Algoritma di atas apabila digambarkan dengan diagram alir pada gambar 5. Nilai beta yang didapatkan pada MK SVM I akan digunakan untuk seleksi fitur.

4.2 Pemodelan Prediksi dan Klasifikasi

Input dari MK-SVM II merupakan gen yang terpilih pada MK-SVM I. Pada MK-SVM II ini bertujuan untuk prediksi pemodelan sehingga dapat dihitung tingkat akurasi.

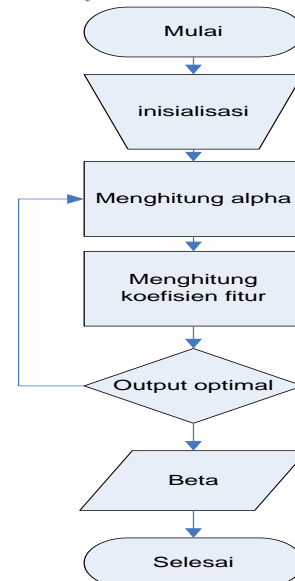
Algoritma SVM II yaitu sebagai berikut :

Input : data vektor $\bar{x}_i \in \mathbb{R}^m$ dan $y_i \in \{+1, -1\}$

1. Inisialisasi : parameter $\lambda = 1$.

2. Menghitung koefisien $\gamma_{j,d}$ dengan menggunakan *dual linear programming* berdasarkan persamaan di bawah ini

$$\begin{aligned} & \max \sum_{i=1}^n u_i^{(t)} \\ & \text{s.t.} \begin{cases} \sum_{i=1}^n u_i y_i \gamma_{j,d} y_j k(x_{i,d}, x_{j,d}) \leq 1 \\ \sum_{i=1}^n u_i y_i = 0 \\ 0 \leq u_i^{(t)} \leq \lambda, \quad i = 1, \dots, n \end{cases} \quad (15) \end{aligned}$$



Gambar 5. Flowchart MK-SVM I

3. Menghitung support vector dengan syarat apabila dalam tiap baris γ terdapat nilai nonzero.
4. Memodelkan prediksi dengan menggunakan

$$f(x^{test}) = \text{sign}(\sum_{d=1}^m \gamma_{j,d} \times (\sum_{i=1}^n \alpha_j y_j k(x_d^{test}, x_{j,d})) + b) \quad (16)$$
5. Evaluasi parameter dan kembali ke langkah 2 sampai dengan optimal.
Output : koefisien $\gamma_{j,d}$, support vector dan hasil klasifikasi.

Algoritma di atas digambarkan dengan diagram alir pada gambar 6.

5. UJI COBA DAN ANALISIS

Uji coba dilakukan pada sebuah PC dengan prosesor Intel Pentium 4 2.26GHz, memori 1 GB. Sistem operasi yang digunakan adalah Windows XP Professional Version 2002 SP2.

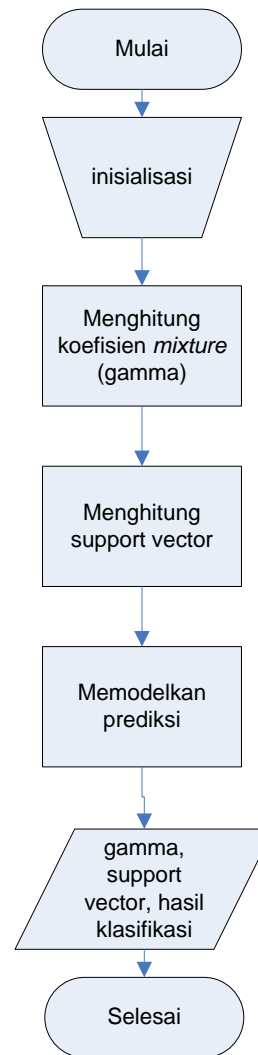
Bahasa komputasi yang digunakan untuk implementasi metode adalah Matlab 7.0.

Uji coba dilakukan dengan beberapa skenario. Uji coba dilakukan dengan 3 Skenario antara lain :

- Skenario 1 : Uji Coba metode standar SVM untuk data ekspresi gen leukimia dan tumor usus besar
- Skenario 2 : Uji Coba MK SVM I untuk data ekspresi gen leukimia dan tumor usus besar serta data alternatif
- Skenario 3 : Uji Coba MK SVM II untuk data ekspresi gen leukimia dan tumor usus besar

5.1 Hasil Uji Coba dan Analisis Skenario I

Pada skenario pertama dilakukan uji coba untuk data ekspresi gen dengan menggunakan fungsi kernel linear, polynomial dan gaussian. Hasil uji coba pada skenario pertama dirangkum dalam tabel 1.



Gambar 6. Flowchart MK-SVM II

Tabel 1. Nilai akurasi data ekspresi gen dengan metode dasar SVM

Data Ekspresi Gen	Jumlah Fitur	SVM		
		Linear	Polynomial	Gaussian
Leukemia Dataset	7129	79,41%	82,35%	0%
Colon Tumor Dataset	2000	59%	50%	0%

Di antara 3 kernel yang diujicoba, untuk data ekspresi gen leukimia didapatkan yang terbaik yaitu dengan menggunakan kernel polyomial. Selanjutnya untuk data ekspresi gen

tumor usus besar, didapatkan akurasi tertinggi pada fungsi kernel linear.

5.2 Hasil Uji Coba dan Analisis Skenario II

5.2.1. Data Ekspresi Gen

Pada metode Multiple Kernel SVM I, dilakukan uji coba pada data ekspresi gen. Uji coba untuk data ekspresi gen digunakan multiple kernel dengan polynomial kernel. Dari rumus fungsi kernel polynomial yang terdapat pada sub bab 2.4, dilakukan uji coba untuk data ekspresi gen leukimia dan tumor usus besar. Nilai beta yang didapatkan yaitu hampir mendekati 0, sehingga dibutuhkan *threshold*.

Nilai *threshold* yang digunakan untuk data ekspresi gen leukimia yaitu sebesar $1e-2$ sehingga didapatkan jumlah beta yang bernilai tidak 0 sebanyak 56 fitur. Sedangkan untuk data ekspresi gen tumor usus besar, nilai *threshold* yang digunakan yaitu sebesar $1e-6$, sehingga didapatkan jumlah fitur sebanyak 13 fitur.

Hasil uji coba MK SVM I untuk data ekspresi gen terangkum dalam tabel 2. Pada tabel 2 dapat dilihat gen mana saja yang terpilih untuk masing-masing dataset.

Tabel 2. Jumlah fitur yang terseleksi untuk Data Ekspresi Gen

No	Data Ekspresi Gen	Jumlah Fitur Awal	Fitur Terpilih
1	Leukimia	7129	56
2	Tumor Usus Besar	2000	13

Tabel 3. Deskripsi gen/fitur yang terseleksi untuk Data Ekspresi Gen

No	Data Ekspresi Gen	Gen Terpilih
1	Leukimia	hum_alu, AFFX-HSAC07/X00351_5, AFFX-HSAC07/X00351_M, D78586, HG2873-HT3017, L06505, L11566, L38941, M19283, M19507, M24194, M31520, M60854, M91036_rna1, U09953, U12404, U12465, U13706, U14971, U14973, U49869_rna1, V01512_rna1, X04500, X51466, X52966, X56997_rna1, X60489, X62691, X68277, X69150, X69391, X73460, X79234, Z48950, Z84721_cds2, D86974, M23178, D49824, HG1428-HT1428, M14199, M10277, V00599, M21119, J04130, X03689, L04483, M13560, M26708, M28130_rna1, Y00787, Z19554, M21142_cds2, M63438, X04347, U00947, HG2887-HT3031
2	Tumor Usus Besar	Hsa.2357, Hsa.20836, Hsa.1447, Hsa.2555, Hsa.467, Hsa.2357, Hsa.20836, Hsa.1648, Hsa.1836, Hsa.7877, Hsa.8068, Hsa.1737, Hsa.31

5.2.2. Data Alternatif

Analisis pengaruh λ dengan σ^2 diperlukan data dengan atribut kecil. Oleh karena itu, dipilih data alternatif yang mempunyai fitur sedikit. Data kredit Jerman yang memiliki atribut sebanyak 24. Data yang digunakan yaitu jumlah data training 100 dan jumlah data testing 30. Percobaan dilakukan dengan

Created with

memasukkan nilai input parameter λ dan σ^2 yang berbeda.

Uji coba pertama yaitu nilai λ 30 dengan σ^2 0,05. Dari MK SVM I didapatkan nilai lagrange multiplier α dengan menggunakan pemrograman kuadratik. Nilai α untuk uji coba pertama berubah sebanyak maksimal 3 kali karena looping yang dihasilkan sebanyak 3 kali. Begitu pula dengan variabel β yang terus diperbaharui sehingga didapatkan nilai β yang optimal. Nilai α dan β terdapat pada lampiran. Jumlah fitur yang terseleksi dari jumlah fitur awal sebanyak 24 menjadi 13 fitur.

Uji coba kedua yaitu nilai λ 40 dengan σ^2 0,05. Dari hasil yang didapatkan terpilih 13 fitur pula. Untuk uji coba ketiga sampai dengan ke-8, didapatkan nilai fitur yang sama yaitu 13.

Selanjutnya dilakukan uji coba dengan mengubah nilai σ^2 0,1. Lalu dilakukan uji coba untuk nilai λ 30 sampai dengan 100. Dari uji coba tersebut didapatkan jumlah fitur yang sama yaitu 12.

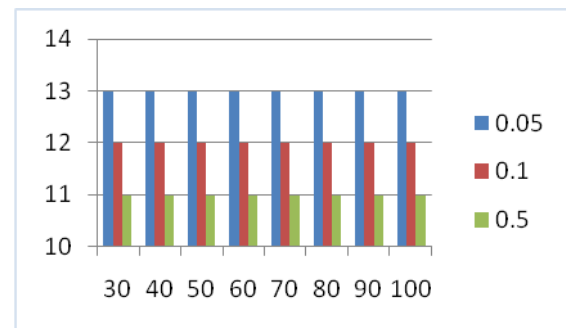
Uji coba untuk interval λ 30 sampai dengan 100 dengan nilai σ^2 0,5 dihasilkan jumlah fitur yang sama yaitu sebanyak 11 fitur.

Pada hasil uji coba dapat dilihat bahwa nilai β pada uji coba dengan nilai σ^2 yang sama menghasilkan nilai β yang sama. Sehingga hal ini yang menyebabkan jumlah fitur yang terseleksi sama.

Uji coba metode MK-SVM I untuk seleksi gen pada data kredit Jerman dirangkum dalam tabel 3. Apabila digambarkan dalam grafik, uji coba data kredit dengan MK SVM I terdapat pada gambar 7. Dari grafik tersebut, dapat diambil kesimpulan bahwa untuk nilai σ^2 yang sama dihasilkan fitur dengan jumlah yang sama. Untuk nilai σ^2 yang semakin besar, maka dihasilkan jumlah fitur lebih sedikit.

Tabel 4. Jumlah Fitur Terseleksi Dengan Input Parameter untuk Data Kredit German

$\lambda \backslash \sigma^2$	0.05	0.1	0.5
30	13	12	11
40	13	12	11
50	13	12	11
60	13	12	11
70	13	12	11
80	13	12	11
90	13	12	11
100	13	12	11



Gambar 7. Hasil Uji Coba Data Kredit dengan MK SVM I

5.3 Hasil Uji Coba dan Analisis Skenario III

Uji coba ini menggunakan input dari MK SVM I. Setelah dilakukan seleksi fitur dari MK SVM I maka dilakukan uji coba untuk pemodelan prediksi sehingga bisa dihitung nilai akurasi. Dari hasil penghitungan akurasi oleh MK-SVM II didapatkan nilai akurasi sebesar 85 % untuk data ekspresi gen leukimia dan 69% untuk data ekspresi gen tumor usus besar.

Apabila dibandingkan dengan skenario I, maka nilai akurasi pada skenario III yaitu dengan Multiple Kernel SVM menunjukkan tingkat akurasi lebih tinggi daripada metode dasar SVM.

Tabel 5 Jumlah Fitur dan Akurasi dengan MK SVM I dan MK SVM II untuk Data Ekspresi Gen

No	Data Ekspresi Gen	Jumlah Fitur	Akurasi dari MK SVM II
1	Leukimia	56	85,20%
2	Tumor Usus Besar	13	69%

5. KESIMPULAN

Setelah dilakukan uji coba dan analisis terhadap perangkat lunak yang dibuat, maka dapat diambil kesimpulan sebagai berikut:

1. Seleksi fitur pada Multiple Kernel SVM menggunakan parameter beta yang dipengaruhi oleh parameter λ dan σ^2
2. Nilai parameter σ^2 apabila nilai σ^2 semakin besar, maka akan membuat fitur yang terseleksi semakin sedikit.
3. Nilai akurasi pada classifier MK SVM II tidak dipengaruhi oleh variabel λ .

DAFTAR PUSTAKA

- [ERA01] Era Muslim. "Meski Angka Kematian Akibat Kanker Masih Tinggi, Penderita Masih Punya Harapan Sembuh" Situs diakses pada tanggal 12 September 2008. <http://www.eramuslim.com/berita/nas/6321144714-meski-angka-kematian-akibat-kanker-masih-tinggi-penderita-masih-punya-harapan-sembuah.htm>
- [CHE07] Zhenyu Chen, Jianping Li dan Liwei Wei, "A Multiple Kernel Support Vector Scheme for Feature Selection and Rule Extraction from Gene Expression Data of Cancer Tissue", ElSevier Journal, Juli 2007.
- [AND06] Ruszczyński, Andrzej. "Nonlinear Optimization". Princeton University Press. 2006.
- [GOL99] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasengeek M, Mesirov JP, et al. "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring". Science Magazine, Vol.286, Oktober 1999.
- [ALO99] Alon U, et al., "Broad Pattern of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays", Proc. Natl. Acad. Sci. USA, No. 96, 1999.
- [ANT03] Nugroho Anto S, Witarto Arief B, Handoko Dwi, "Support Vector Machine, Teori dan Aplikasinya dalam ilmukomputer.com, Desember 2003.
- [STE04] Boyd, Stephen and Vandenberghe, Lieven. "Convex Optimization". Cambridge University Press. 2004. <http://www.stanford.edu/~boyd/cvxbook/>
- [TAH03] Taha, Hamdy. "Operation Research: An Introduction". Prentice Hall. 2003
- [WIK01] Wikipedia. "Gene Expression". Situs diakses pada 12 September 2008. http://en.wikipedia.org/wiki/Gene_expression
- [WIK02] Wikipedia. "VC Dimension". Situs ini diakses pada 10 Juli 2009. http://en.wikipedia.org/wiki/VC_dimension
- [WIK03] Wikipedia. "Infimum". Situs diakses pada 1 Juni 2009. <http://en.wikipedia.org/wiki/Infimum>

[WIK04]Wikipedia."Supremum". Situs diakses pada 1 Juni 2009.
<http://en.wikipedia.org/wiki/Supremum>

[DAT01]Dataset Ekspresi Gen Leukimia.
http://www.broad.mit.edu/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=43
Situs diakses pada tanggal 10 Oktober 2008

[DAT02]Dataset Ekspresi Gen Tumor Usus Besar
<http://microarray.princeton.edu/oncology/affydata/index.html>
Situs diakses pada tanggal 10 Oktober 2008.