

Klasifikasi *Microarray* “*Prostate Cancer*” Menggunakan Metode *Fuzzy Support Vector Machine (FSVM)-Genetic Algorithm*

Cicilia Ajeng Pratiwi dan Irhamah

Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data (FMKSD)

Institut Teknologi Sepuluh Nopember (ITS)

e-mail: irhamah@statistika.its.ac.id

Abstrak—Salah satu jenis kanker yang menjadi penyebab terbanyak kematian pada populasi pria adalah kanker prostat. Penyakit ini hanya terdapat pada pria karena pada wanita tidak memiliki kelenjar prostat. Secara global, kanker prostat menduduki urutan keempat, kanker yang paling sering ditemukan pada manusia setelah kanker payudara, paru dan kolorektum. Sedangkan angka kejadian kanker pada pria, kanker prostat menduduki urutan ke-2. Pada umumnya penderita baru mengetahui penyakit tersebut sudah memasuki stadium lanjut. Terlambatnya penanganan pada penderita *prostate* bisa berakibat fatal bahkan dapat menyebabkan kematian. Oleh karena itu, penyakit kanker prostat sangat penting untuk didiagnosis sedini mungkin sebelum penyebaran sel kanker ke organ internal. Pada perkembangan saat ini, terdapat teknologi *microarray* yang memiliki pengaruh besar dalam menentukan gen informatif menyebabkan kanker. Penelitian ini menggunakan data *microarray* “*prostate cancer*”. Ekspresi gen yang terdapat pada data *microarray* “*prostate*” dapat digunakan untuk mengklasifikasikan pasien yang mengalami tumor prostat dan normal. Penelitian ini diperoleh hasil klasifikasi *Fuzzy Support Vector Machine (FSVM)* dengan menggunakan seleksi *Fast Correlation Based Filter (FCBF)* tanpa optimasi *genetic algorithm* menghasilkan nilai akurasi lebih tinggi dibandingkan tanpa seleksi. Selain itu, diperoleh juga nilai akurasi klasifikasi *FSVM* dengan menggunakan seleksi dan optimasi *genetic algorithm* lebih tinggi dibandingkan tanpa seleksi.

Kata Kunci—*Fast Correlation Based Filter, Fuzzy Support Vector Machine, Genetic algorithm, Microarray.*

I. PENDAHULUAN

PENYAKIT kanker merupakan salah satu penyebab kematian utama di seluruh dunia. Pada tahun 2012, penyakit kanker menjadi penyebab kematian sekitar 8,2 juta orang [1]. Salah satu jenis kanker yang menjadi penyebab terbanyak kematian pada populasi pria adalah kanker prostat. Penyakit ini hanya terdapat pada pria karena pada wanita tidak memiliki kelenjar prostat. Kanker prostat menduduki urutan ke-2 yaitu sekitar 14,8% setelah kanker paru 16,8%. Pada tahun 2012 kejadian kanker prostat menempati urutan ke-3 kanker pada pria setelah kanker paru dan kanker kolorektum, sedangkan angka kematian menempati urutan ke-4 di Indonesia [2].

Data di USA menunjukkan bahwa lebih dari 90% kanker prostat ditemukan pada stadium dini, sedangkan di Indonesia banyak ditemukan pada stadium lanjut karena terjadi keterlambatan diagnosis [2]. Oleh karena itu, penyakit kanker pros-

tat sangat penting untuk didiagnosis sedini mungkin sebelum penyebaran sel kanker ke organ internal. Pada perkembangan saat ini, terdapat teknologi *microarray* yang memiliki pengaruh besar dalam menentukan gen informatif yang menyebabkan kanker.

Microarray mampu menentukan ekspresi ribuan gen dan secara simultan memantau proses biologis yang sedang berlangsung [3]. *Microarray* merupakan bagian dari *high dimensional data* karena memiliki ratusan sampai dengan ribuan fitur [4]. Namun permasalahan yang terjadi adalah data *microarray* mempunyai jumlah variabel yang lebih besar dibandingkan dengan jumlah observasinya. Oleh karena itu, perlu dilakukan seleksi variabel dengan menggunakan metode *fast correlation based filter* untuk menyelesaikan permasalahan klasifikasi.

Penelitian sebelumnya telah melakukan analisis dan implementasi *feature selection* menggunakan algoritma *Fuzzy Support Vector Machine (FSVM)*. Hasil dari penelitian tersebut adalah *feature selection* tipe *wrapper* dengan menggunakan metode *Fuzzy Support Vector Machine* dimana *feature selection* dilakukan bersamaan dengan pemodelan dan evaluasi *feature* berdasarkan *classification rate* yang dihasilkan oleh *classifier* yaitu *Fuzzy Support Vector Machine*. Semakin baik parameter *FSVM* maka hasil *feature selection* juga akan semakin baik berbanding lurus dengan hasil klasifikasi [5]. Berdasarkan penelitian mengenai *fuzzy support vector machine (FSVM)* pada data *microarray* diperoleh hasil tingkat akurasi yang cukup tinggi dibandingkan metode klasifikasi lainnya seperti *SVM*, *ANN*, *CART*, maupun analisis diskriminan untuk klasifikasi data “*prostate cancer*”. Selain itu, diperoleh kesimpulan performansi model *FSVM* dengan atau tanpa *feature selection* yaitu tingkat akurasi model *FSVM* dengan *feature selection* *SNR* lebih tinggi dibandingkan model *FSVM* dengan *feature selection* *SVM-RFE* dan tanpa *feature selection* [6].

Metode klasifikasi yang digunakan dalam penelitian ini adalah *Fuzzy Support Vector Machine (FSVM)* dengan atau tanpa seleksi variabel menggunakan *fast correlation based filter*. Dalam penelitian ini juga akan dibandingkan kinerja Algoritma Genetika dibandingkan *grid search* dalam memperoleh parameter *SVM* yang optimum berdasarkan tingkat ketepatan klasifikasi. Algoritma Genetika menunjukkan kinerja yang baik dalam memperoleh solusi yang global optimum dan mudah dihibridisasi dengan metode

lain [7]. Tujuan klasifikasi ini antara lain adalah apabila ingin memprediksi pasien terkena kanker atau tidak bisa menggunakan gen yang sudah terseleksi saja tanpa mengambil 6033 gen. Hal tersebut dilakukan untuk me-minimalisir biaya yang dikeluarkan.

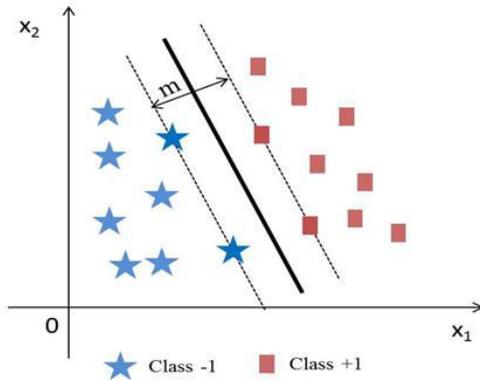
II. TINJAUAN PUSTAKA

A. Support Vector Machine (SVM)

SVM merupakan suatu metode *machine learning* yang bekerja atas dasar prinsip *Structural Risk Minimization* dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* [8]. Inti dari proses pembelajaran pada SVM adalah mendapatkan lokasi *hyperplane* [9]. Terdapat tiga penjelasan mengenai SVM yaitu dapat dijelaskan sebagai berikut.

1. SVM pada *Linear Separable Data*

SVM *Linearly separable data* merupakan data yang dapat dipisahkan secara linier. Misalkan $x_i = \{x_i, x_{i+1}, \dots, x_n\}$ adalah *dataset* dan $y_i = \{+1, -1\}$ adalah label kate-gori untuk *dataset*. Pada klasifikasi *linear*, SVM dapat dibedakan menjadi dua yaitu *linearly separable* dan *linearly non-separable*. Pada Gambar 2.1 dapat dilihat ilustrasi li-nier *separable case*.



Gambar 1. Konsep *Hyperplane* pada SVM.

Gambar 1 menunjukkan garis putus-putus merupakan bidang membatasi kelas -1 dan kelas +1. Garis tegas lurus diantara garis putus-putus merupakan bidang pemisah antara kelas +1 dan kelas -1. Persamaan *hyperplane* dapat ditulis seperti persamaan (1)

$$w^T x + b = 0 \tag{1}$$

w : vektor bobot (*weight vector*)

b : konstanta.

Nilai margin antara bidang pembatas adalah $\frac{2}{\|w\|}$

2. SVM pada *Linear Non-Separable Data*

Pada beberapa kenyataan, besar kemungkinan terjadi *misclassification*. Mengatasi hal tersebut maka dilakukan klasifikasi *linear non separable* dengan beberapa pengembangan. Masalah optimasi baik pada fungsi obyektif maupun kendala dimodifikasi dengan variabel *slack* $\xi > 0$ merupakan sebuah ukuran kesalahan klasifikasi [10]. Penyele-saian untuk

kasus linear SVM *non separable* adalah diperoleh *hyperplane* pemisah yang optimal

$$f(x) = \sum_{i,j=1}^n \alpha_i y_i x_i^T x_j + b \tag{2}$$

3. SVM pada *Non-Linearly Separable Data* dengan Menggunakan Metode Kernel

SVM pada *Non-linearly separable data*, merupakan data nonlinear yang dapat dipisahkan tetapi memerlukan metode tersendiri dalam penyelesaiannya. Penyelesaian data jenis ini dapat dilakukan dengan menggunakan SVM yang di kombinasikan dengan metode *kernel*. Metode kernel mentransforma-sikan data ke dalam dimensi ruang fitur sehingga dapat dipisahkan secara linier pada *fitur space*.

Kernel RBF direkomendasikan untuk diuji pertama kali. Hal ini dikarenakan fungsi kernel RBF memiliki performansi yang sama dengan kernel linier pada parameter tertentu [11].

Tabel 1. Fungsi Kernel pada SVM Non-linier

Fungsi Kernel	Fungsi pembentuk matriks kernel
Linier	$K(x_i, x_j) = x_i^T x_j$
RBF	$K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2), \gamma > 0$
Polinomial	$K(x_i, x_j) = (x_i^T x_j + 1)^p, p = 1, \dots, z$

B. Fuzzy Support Vector Machines (FSVM)

Titik kunci dalam metode *Fuzzy Support Vector Machines* (FSVM) adalah menentukan fungsi keanggotaan *fuzzy*. Pada metode *Fuzzy Support Vector Machines* (FSVM) masih memiliki kelemahan dalam menangani data *imbalanced*. Terdapat beberapa penerapan yang hanya fokus pada akurasi untuk klasifikasi suatu kelas. Misalkan diberikan rangkaian *training* sebagai berikut [12].

$$(y_1, x_1, s_1), \dots, (y_n, x_n, s_n) \tag{3}$$

Keanggotaan *fuzzy* s_i menjadi fungsi pada kelas y_i yaitu $s_i = 1$ jika $y_i = 1$ dan $s_i = 0,1$ jika $y_i = -1$. Fungsi keputusan untuk FSVM dituliskan dalam persamaan (4)

$$f(x) = \sum_{i,j=1}^n \alpha_i y_i K(x_i, x_j) + b \tag{4}$$

dimana $0 \leq \alpha_i \leq s_i C, i = 1, 2, \dots, n$

C. Algoritma Genetika

Konsep dasar yang mengilhami timbulnya algoritma genetika adalah teori evolusi alam. Dikarenakan algoritma genetika diilhami oleh ilmu genetika maka istilah yang dipergunakan dalam algoritma genetika banyak diadaptasi dari ilmu tersebut [13]. Terdapat beberapa langkah dalam melakukan analisis menggunakan algoritma genetika: [14].

1. Menentukan pengaturan dari operator algoritma genetika yang cocok dengan masa-lah yang akan dianalisis.
2. Melakukan proses inialisasi.
3. Mendapatkan nilai *fitness* $f(u)$ pada tiap kromosom v_u dalam populasi.

4. Mengevaluasi nilai *fitness* $f(u)$ pada tiap kromosom v_u dalam populasi.
5. Menerapkan seleksi *roulettewheel*. Tahapan seleksi *roulettewheel* adalah menghitung nilai *fitness* masing-masing kromosom v_u , menghitung total nilai *fitness*, menghitung proporsi masing-masing kromosom, menghitung nilai kumulatif proporsi untuk tiap kromosom, membangkitkan sebuah r_n angka dengan range $[0,1]$. Jika $r_n \leq S_1$, maka kromosom v_1 yang dipilih, lalu selainnya v_u yang dipilih, sehingga $S_{u-1} < r_n \leq S_u$.
6. Melakukan pindah silang (*Crossover*).
7. Melakukan proses mutasi dimana mutasi diterapkan dengan probabilitas P_m .
8. Melakukan proses elitisme
9. Melakukan pergantian antara nilai *fitness* dengan keturunan baru.
10. Berhenti ketika nilai *fitness* sudah konvergen dan kembali ke solusi terbaik dalam populasi saat ini jika kriteria telah terpenuhi, jika belum kembali ke Langkah 3.

D. Fast Corelation Based Filter (FCBF)

Fast Corelation Based Filter merupakan salah satu algoritma *feature selection* yang bersifat multivariat dan mengukur kelas fitur dan korelasi antara fitur-fitur [15]. Terdapat dua pendekatan dengan mengukur korelasi antara dua variabel acak yaitu berdasar pada *classical linear correlation/linear correlation coefficient* dan berdasarakan pada *information-theoretical concept of entropy*.

E. K-Fold Cross Validation

Metode validasi dengan *k-folds* sangat cocok digunakan untuk kasus data yang jumlah sampelnya terbatas. Dalam *k-folds cross validation*, data (D) dibagi ke dalam *k subsets* data D_1, D_2, \dots, D_k dengan jumlah yang sama. Data yang digunakan untuk *training* adalah *subsets* data *k-1* yang dikombinasikan secara bersama-sama dan kemudian diaplikasikan untuk sisa satu *subsets* data sebagai hasil *testing*. Proses ini diulangi sebanyak *k subsets* dan hasil akurasi klasifikasi yaitu hasil rata-rata dari setiap data *training* dan *testing* [16].

F. Evaluasi Performasi Klasifikasi

Evaluasi performasi klasifikasi yaitu dilakukan perhitungan akurasi klasifikasi dengan *confusion matrix* digunakan untuk komposisi data yang *balance*. Tabel klasifikasi untuk pengukuran performa klasifikasi ditunjukkan dalam Tabel 2.

Tabel 2.
Tabel Klasifikasi

Riil	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan :

TP : *True Positive* (jumlah prediksi benar pada kelas positif)

FP : *False Positive* (jumlah prediksi salah pada kelas positif)

FN : *False Negative* (jumlah prediksi salah pada kelas negatif)

TN : *True Negative* (jumlah prediksi benar pada kelas negatif)

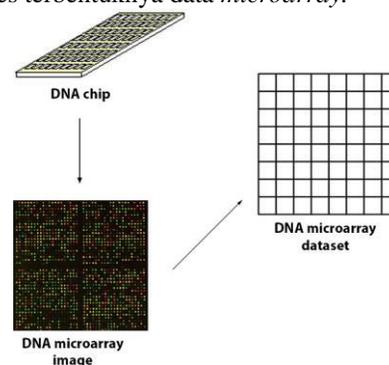
Berdasarkan Tabel 2, perhitungan akurasi dapat dilakukan dengan rumus sebagai berikut.

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP} \tag{5}$$

Sensitivitas merupakan akurasi kelas positif sedangkan spesifitas merupakan akurasi pada kelas negatif [8].

G. Microarray Data

Micorarray mampu menentukan ekspresi ribuan gen dan secara simultan memantau proses biologis yang sedang berlangsung [3]. Selanjutnya ekspresi dari ribuan gen yang merepresentasikan suatu jaringan pada manusia, akan diklasifikasikan sebagai jaringan kanker atau bukan khususnya pada penyakit *prostate*. Karakteristik *microarray data* adalah jumlah data sedikit dan jumlah variabel yang sangat banyak sedangkan jumlah data sedikit karena harga untuk mendapatkan data sangat mahal. Data ini berisi informasi gen karena itu jumlah variabel-nya sangat banyak. Berikut ini gambaran mengenai proses terbentuknya data *microarray*.



Gambar 2. Ekspresi Gen *Microarray* [15].

H. Prostat

Prostat merupakan kelenjar seukuran buah kenari yang terdapat di dalam sistem reproduksi pria, yang terletak di antara leher kandung kemih dan saluran kemih (uretra). Prostat mengeluarkan cairan berwarna putih yang memberi nutrisi dan mengangkut sperma, yang disebut sebagai semen. Hormon pria yang disekresi oleh testis secara langsung memengaruhi pertumbuhan dan fungsi prostat. Kasus prostat yang bengkak umum terjadi di kalangan pria paruh baya dan lanjut usia, namun sebagian besar kasus yang terjadi merupakan hiperplasia jinak (peningkatan jumlah sel yang tidak normal) [17].

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam penelitian ini adalah data jenis *microarray "prostate cancer"* dari penelitian yang dilakukan di Florida pada Maret 2002 [18].

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian disajikan dalam Tabel 3.

Tabel 3.
Variabel Penelitian

Dataset	Banyak Variabel	Banyak Class	Banyak Sampel	
			Tumor prostate (1)	Normal (0)
<i>Prostate</i>	6033	2	52	50

Struktur data *prostate* dapat ditunjukkan melalui Tabel 4.

Tabel 4.
Struktur Data *Prostate*

Pengamatan ke-	X ₁	X ₂	...	X ₆₀₃₃	Y
1	X ₁₍₁₎	X ₁₍₂₎	...	X ₁₍₆₀₃₃₎	1
2	X ₂₍₁₎	X ₂₍₂₎	...	X ₂₍₆₀₃₃₎	1
3	X ₃₍₁₎	X ₃₍₂₎	...	X ₃₍₆₀₃₃₎	1
.
.
102	X ₁₀₂₍₁₎	X ₁₀₂₍₂₎	...	X ₁₀₂₍₆₀₃₃₎	0

Keterangan:

X : Jenis Gen

Y : Pasien

C. Langkah Analisis

Langkahanalisisyang disusun untuk melakukan penelitianiniadalahsebagiaiberikut.

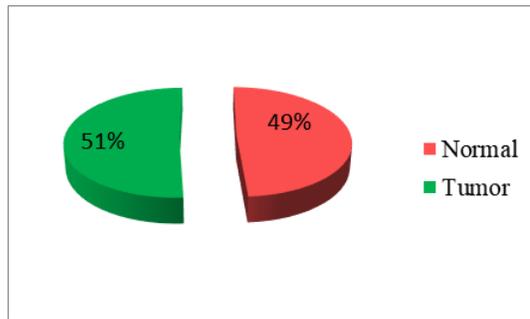
1. Mendeskripsikan data *microarray* “*Prostate*”.
2. Melakukan *feature selection* menggunakan metode FCBF.
3. Menentukan fungsi kernel yang digunakan yaitu kernel RBF
4. Membagi data *training* dan *testing* menggunakan 10-fold *cross validation*.
5. Analisis klasifikasi menggunakan metode *Grid Search* FSVM pada data *microarray* “*Prostatecancer*”.
 - a. Mendapatkan parameter yang optimal dengan nilai *cost* (C) berada diantara $2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^{13}, 2^{14}, 2^{15}, 100, 1000$ dan nilai *gamma* (γ) diantara $2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^1, 2^2, 2^3$
 - b. Melakukan klasifikasi FSVM dengan kombi-nasi nilai parameter C dan γ .
 - c. Menghitung akurasi klasifikasi
 - d. Menentukan nilai parameter optimal C dan γ dari seluruh kombinasi parameter mengguna-kan data *training*.
 - e. Menghitung performa klasifikasi menggunakan data *testing*.
6. Analisis klasifikasi menggunakan FSVM-GA pada data *microarray* “*Prostatecancer*”.
7. Melakukan perbandingan hasil klasifikasi metode *Grid Search* FSVM dengan FSVM-GA dengan seleksi atau tanpa seleksi.
8. Menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

Hasil performansi klasifikasi berupa nilai akurasi yang diperoleh pada metode FSVM dan optimasi *genetic algorithm* dengan atau tanpa *feature selection* FCBF pada data *microarray* “*prostate cancer*”.

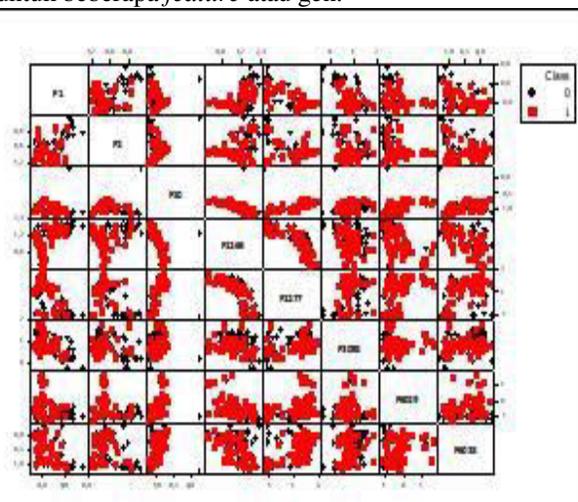
A. Karakteristik Data *Microarray Prostate Cancer*

Data *microarray prostate cancer* terdiri 6033 variabel, dua kategori kelas, dan 102 observasi yang berasal dari ekspresi gen pasien *prostate*.



Gambar 3. Piechart Proporsi Tiap Kategori

Gambar3 terlihat bahwa proporsi tiap kategori memiliki proporsi yang hampir sama (*balance*), dari 102 pasien kanker prostat sebanyak 51% pasien masuk dalam kelas tumor atau sebanyak 52 pasien dan 49% pasien masuk dalam kelas normal. Berikut akan ditampilkan penyebaran *prostate data-sets* untuk beberapa *feature* atau gen.



Gambar 4. Penyebaran Beberapa *Feature Prostate Datasets*.

Gambar 4 didapatkan bahwa data untuk tiap kelas tersebar secara merata, hal tersebut akan mempersulit dalam melakukan proses klasifikasi, sehingga diperlukan fungsi pemisah atau *hyperplane* untuk mempermudah proses klasifikasi data. Proses pemisahan data prostat tidak bisa dipisahkan secara linear, sehingga diperlukan pemisah untuk data secara tidak linier dengan meng-gunakan metode kernel.

B. Klasifikasi Data *Microarray Prostate Cancer* dengan Menggunakan Metode FSVM Tanpa Seleksi Variabel

Sebelum melakukan analisis FSVM, maka dilakukan partisi data menjadi data *training* dan data *testing* dengan proporsi 90:10. Pembagiannya menggunakan 10-fold. Selanjutnya akan dilakukan analisis klasifikasi dengan menggunakan metode FSVM dengan mencari parameter yang optimal dengan nilai *cost* berada diantara $2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^{13}, 2^{14}, 2^{15}, 100, 1000$ dan nilai *gamma* diantara $2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^1, 2^2, 2^3$. Sehingga diperoleh jumlah kombinasi *cost* dan *gamma* sebanyak 437 kombinasi atau melakukan *running* data sebanyak 437 kali. Didapatkan nilai *cost* dan *gamma* yang paling optimal dari data *training* adalah $2^9, 1000$ dan $2^{-15}, 2^{-14}$ dengan nilai rata-rata akurasi sebesar 100%. Hasil ukuran performansi yang didapat dari parameter *cost* dan *gamma*

yang optimal diperoleh akurasi sebesar 90,18% menggambarkan kemampuan dalam membedakan sampel berdasarkan kelas dan sisanya masuk dalam kesalahan klasifikasi.

C. Klasifikasi Data Microarray Prostate Cancer Menggunakan Metode FSVM dengan Genetic Algorithm (GA)

Pada subbab ini akan dibahas mengenai klasifikasi data prostate cancer dengan seleksi maupun tanpa seleksi FCBF. Hasil seleksi variabel dengan FCBF menghasilkan 29 variabel yang terseleksi. Diperoleh nilai *cost* dan *gamma* yang paling optimal terdapat pada 2^{10} dan 2^{-4} dengan nilai rata-rata akurasi sebesar 97%. Ringkasan tabel untuk nilai akurasi, spesifitas, dan sensitifitas data testing dari parameter yang optimal disajikan dalam Tabel 5.

Tabel 5.
Ukuran Performansi untuk Parameter Optimal

Fold	Akurasi	Sensitifitas	Spesifitas
1	0,9090909	1	0,8333333
2	1	1	1
3	1	1	0
4	0,9	0,8	1
5	0,9	0,8	1
6	1	1	1
7	0,9	0,8	1
8	1	1	1
9	0,9	0,8	1
10	0,9	1	0,8
Rata-rata	0,9409091	0,92	0,8633333

Tabel 5 menunjukkan ukuran performansi yang didapat dari parameter *cost* dan *gamma* yang optimal. Diperoleh nilai akurasi sebesar 94,09% menggambarkan kemampuan dalam membedakan sampel berdasarkan kelas dan sisanya masuk dalam kesalahan klasifikasi. Nilai sensitifitas sebesar 92%, ini menunjukkan bahwa sampel mampu membedakan kelas tumor dengan benar. Nilai spesifitas menunjukkan sampel mampu membedakan kelas normal dengan benar sebesar 86,3%.

Selanjutnya optimasi parameter GA yang berasal dari seleksi atau tanpa seleksi variabel menggunakan metode FCBF. Langkah pertama yaitu melakukan inisialisasi kromosom sebanyak 100. Nilai parameter optimal *C* sebesar 1024 dan nilai parameter γ sebesar 0,0625. Berikut ini adalah gambaran mengenai ilustrasi optimasi GA dalam Gambar 5.

Parameter		<i>C</i>	γ
Kromosom		1024	0,0625

Kromosom Ke-	<i>Fitness</i>	Proporsi Nilai <i>Fitness</i>	Nilai <i>Fitness</i> Kumulatif	Random Number
1	0,9878	0,011	0,011	0,08
2	0,9754	0,0108	0,0218	0,031
⋮	⋮	⋮	⋮	⋮
100	0,7090	0,0079	1,0000	0,056

Gambar 5. Ilustrasi Optimasi GA.

Setelah selesai melakukan proses seleksi, selanjutnya melakukan proses pindah silang atau *crossover* yaitu menghasilkan kromosom baru dari hasil perpaduan 2 kromosom orang tua.

Sebelum <i>Crossover</i>		
Orang tua 1	1024,679	0,01
Orang tua 2	1021,611	0,015
Setelah <i>Crossover</i>		
Anak 1	1024,2219	0,010745
Anak 2	1022,0681	0,014255

Sebelum Mutasi	0,008	0,079
Kromosom	1022,068132	0,014255
Setelah Mutasi mutasi		
Kromosom	1020,50043	0,014255

Gambar 5. Ilustrasi Optimasi GA (Lanjutan)

Apabila sudah dilakukan proses pindah silang maka dilanjutkan dengan proses mutasi. Proses mutasi dengan membandingkan antara nilai probabilitas mutasi (P_m) sebesar 0,01. Jika nilai bilangan *random* lebih kecil dibandingkan probabilitas mutasi, maka gen yang bersesuaian akan mengalami mutasi dengan cara mengganti gen tersebut dengan bilangan *random*. Dilanjutkan dengan proses elitisme dalam Gambar 6 untuk mempertahankan kromosom terbaik dalam populasi dengan melihat nilai *fitness* tertinggi untuk generasi selanjutnya. Kromosom yang dipertahankan sebesar 5% dari total kromosom dalam populasi yaitu sebanyak 5 kromosom.

Kromosom ke-	Gen		<i>Fitness</i>
	<i>C</i>	γ	
1	1024,06	0,056	0,988
2	1023,67	0,044	0,978
3	1015,70	0,034	0,959
4	1000,06	0,028	0,871
5	956,9	0,040	0,882
⋮	⋮	⋮	⋮
100	1005,98	0,062	0,715

} Digunakan pada Generasi ke-2

Gambar 6. Ilustrasi Elitisme Optimasi GA.

Kromosom yang dipertahankan sebesar 5% dari total kromosom. Nilai *fitness* tertinggi dari generasi ke-1 digunakan untuk kromosom awal generasi ke-2. Sedangkan kromosom hasil generasi ke-2 digunakan untuk kromosom awal pada generasi ke-3. Proses optimasi parameter dengan algoritma genetika dilanjutkan hingga mendapatkan nilai *fitness* yang konvergen.

Perbandingan hasil klasifikasi menggunakan metode *Grid Search* FSVM dengan GA-FSVM dengan atau tanpa seleksi FCBF.

Tabel 7.
Perbandingan Hasil Klasifikasi

Metode		Akurasi
Tanpa Seleksi	<i>Grid Search</i> FSVM	90,18%
Seleksi FCBF	GA-FSVM	76,45%
Seleksi FCBF	<i>Grid Search</i> FSVM	94,09%
Seleksi FCBF	GA-FSVM	99%

Tabel 7 menunjukkan nilai akurasi klasifikasi FSVM dengan menggunakan seleksi dan optimasi *genetic algorithm* lebih tinggi dibandingkan tanpa seleksi. Hasil klasifikasi FSVM dengan seleksi FCBF tanpa optimasi *genetic algorithm*

menghasilkan nilai akurasi lebih tinggi dibandingkan tanpa seleksi.

V. KESIMPULAN DAN SARAN

Kesimpulan yang diperoleh dari hasil analisis yang dilakukan adalah karakteristik data proporsi tiap kategori memiliki proporsi yang hampir sama (*balance*), dari 102 pasien kanker prostat sebanyak 51% pasien masuk dalam kelas tumor atau sebanyak 52 pasien dan 49% pasien masuk dalam kelas normal. Pada klasifikasi data *prostate cancer* didapatkan nilai *cost* dan *gamma* yang paling optimal terdapat pada 2^9 dan 2^{-15} dengan hasil ukuran performansi akurasinya sebesar 90,18%. Didapatkan nilai sensitifitas sebesar 86% menunjukkan bahwa sampel mampu membedakan kelas tumor dengan benar. Nilai akurasi klasifikasi FSVM dengan menggunakan seleksi dan optimasi *genetic algorithm* lebih tinggi dibandingkan tanpa seleksi.

Saran yang bisa diberikan yaitu melakukan penelitian dengan menggunakan simulasi data dengan berbagai karakteristik data, sehingga dapat diketahui kinerja algoritma genetika bukan hanya untuk optimasi tetapi seleksi variabel berdasarkan karakteristik data yang telah disimulasikan.

DAFTAR PUSTAKA

- [1] Kemenkes, "Situasi Penyakit Kanker," *Bul. Jendela Data dan Inf. Kesehatan*, pp. 1–35, 2015.
- [2] V. R. Solang, A. Monoarfa, and F. Tjandra, "Profil penderita kanker prostat di RSUP Prof. Dr. R. D. Kandou Manado periode tahun 2013–2015," *J. e-Clinic*, vol. 4, no. 2, 2016.
- [3] P. T. Ramadhani, U. N. Wisesty, and A. Aditsania, "Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Functional Link Neural Network dengan Seleksi Fitur Genetic Algorithm," *Ind. J. Comput.*, vol. 2, no. 2, pp. 11–22, 2017.
- [4] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," in *Proceedings of the Twentieth International Conference on Machine Learning (ICML)*, 2003.
- [5] R. Hermawan, A. P. Kurniati, and Suyanto, "Analisis Dan Implementasi Feature Selection Menggunakan Algoritma Fuzzy Support Vector Machine (FSVM)," *Tek. Inform.*, pp. 1–37.
- [6] M. Hajiloo, H. R. Rabiee, and M. Anoooshahpour, "Fuzzy support vector machine: an efficient rule based classification technique for microarrays," *Bioinformatics*, pp. 1–11, 2013.
- [7] Z. Ismail and Irhamah, "Genetic Algorithm and Tabu Search for Vehicle Routing Problems with Stochastic Demand," in *AIP Conference Proceedings*, 2010.
- [8] R. Feldman and J. Sanger, *The Text Mining Hand Book*. New York: Cambridge University Press, 2007.
- [9] dan D. H. Nugroho, Anto S., Arief Budi W., "Support Vector Machine "Teori dan Aplikasinya dalam Bioinformatika," *Ilmu Komputer.Com*, 2003. .
- [10] S. Gunn, "Support Vector Machines for Classification and Regression," Southampton, 1998.
- [11] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A Practical Guide to Support Vector Classification," Taiwan, 2003.
- [12] C. Lin and S. Wang, "Fuzzy Support Vector Machines," *IEEE Trans. Neural Net-work*, pp. 464–471, 2002.
- [13] K. Setiawan, *Paradigma Sistem Cerdas*. Malang: Bayumedia, 2003.
- [14] Z. Ismail and Irhamah, "Solving the Vehicle Routing Problem with Stochastic Demands via Hybrid Genetic Algorithm-Tabu Search," *J. Math. Statistics*, pp. 161–167, 2008.
- [15] A. Alonso, S. Noelia, and B. Veronica, *Feature Selection for High-Dimensional Data. Artificial Intelligence: Foundations, Theory, and Algorithms*. Switzerland: Springer International Publishing Switzerland, 2015.
- [16] V. B. Canedo, N. S. Marono, A. A. Betanzos, J. Benitez, and F. Herrera, "A Review of Microarray Datasets and Applied Feature Selection Methods," *Inf. Sci.*, pp. 111–135, 2014.
- [17] H. Cheng, "Kanker Prostat," 2017. [Online]. Available: [https://www21.ha.org.hk/smartpatient/EM/MediaLibraries/EM/Diseases/Cancer/Prostate Cancer/Cancer-Prostate-Cancer-Indonesian.pdf?ext=.pdf](https://www21.ha.org.hk/smartpatient/EM/MediaLibraries/EM/Diseases/Cancer/Prostate%20Cancer/Cancer-Prostate-Cancer-Indonesian.pdf?ext=.pdf). [Accessed: 19-Jul-2018].
- [18] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, and C. Ladd, "Gene expression correlates of clinical prostate cancer behavior," *Cancer Cell*, vol. 1, pp. 203–209, 2002.