

Optimasi Fitur Menggunakan *Backward Elimination* Dan Algoritma SVM Untuk Klasifikasi Kanker Payudara

Farizul Ma'arif¹, Toni Arifin²

¹Universitas BSI
e-mail: farizulmaarif@gmail.com

²Universitas BSI
e-mail: toni.tfn@bsi.ac.id

Abstrak

Organisasi Kesehatan Dunia (WHO) tahun 2004, menyatakan bahwa 5 besar kanker di dunia adalah kanker paru-paru, kanker payudara, kanker usus besar, kanker lambung, dan kanker hati. WHO mengestimasi bahwa 84 juta orang meninggal akibat kanker dalam rentang waktu 2005-2015. Survei yang dilakukan WHO menyatakan 8 sampai 9% wanita mengalami kanker payudara. Hal itu membuat kanker payudara sebagai jenis kanker yang paling banyak ditemui pada wanita setelah kanker leher rahim. Pada penelitian ini dilakukan klasifikasi tingkat keganasan *breast cancer* dengan menggunakan metode optimasi fitur *Backward Elimination* dan *Support Vector Machine* (SVM), yang bertujuan untuk memudahkan ahli dalam mengidentifikasi kanker payudara. Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa metode *Support Vector Machine* (SVM) adalah Algoritma yang baik diantara algoritma yang penulis telah uji untuk pengklasifikasian Kanker Payudara menggunakan Dataset WBC (*Wisconsin Breast Cancer*). Dimana nilai klasifikasi performansi Akurasi dan AUC nya adalah yang tertinggi, sedangkan untuk penggabungan algoritma seleksi fitur *Backward Elimination* dan *Support Vector Machine* (SVM) mendapatkan peningkatan Akurasi sebesar 14% sehingga nilai tingkat akurasi akhirnya sebesar 97.14% dan nilai AUC mencapai 0.995.

Keywords: Kanker Payudara, *Backward Elimination*, *Support Vector Machine*, Klasifikasi.

Abstract

The World Health Organization (WHO) in 2004, States that the 5 large world cancer is lung cancer, breast cancer, colon cancer, stomach cancer, and liver cancer. The who estimates that 84 million people died of cancer in 2005-2015 time frame. A survey conducted the WHO declares the 8 to 9% of women experiencing breast cancer. It makes breast cancer as the most common type of cancer found in women after cervical cancer. This research was conducted on the classification of the degree of malignancy of breast cancer by using *Backward Elimination* feature optimization method and *Support Vector Machine* (SVM), which aims for make it easier the expert in identifying breast cancer. Based on the results and discussion has been done, it can be concluded that the method of *Support Vector Machine* (SVM) is a good algorithm between the algorithms that the author has been testing for breast cancer classification using Dataset WBC (*Wisconsin Breast Cancer*). Where the value of classification Accuracy performance and his AUC is the highest, while for the merge feature selection algorithm *Backward Elimination* and *Support Vector Machine* (SVM) get an increased Accuracy of 14% accuracy level value so that the end of 102.72% and AUC values reach 0995.

Keyword: Breast Cancer, *Backward Elimination*, *Support Vector Machine*, Classification.

1. Pendahuluan

Organisasi Kesehatan Dunia (WHO) tahun 2004, menyatakan bahwa 5 besar kanker di dunia adalah kanker paru-paru, kanker payudara, kanker usus besar, kanker

lambung, dan kanker hati. WHO mengestimasi bahwa 84 juta orang meninggal akibat kanker dalam rentang waktu 2005-2015. Survei yang dilakukan WHO menyatakan 8 sampai 9 persen wanita

mengalami kanker payudara. Hal itu membuat kanker payudara sebagai jenis kanker yang paling banyak ditemui pada wanita setelah kanker leher rahim (Anggorowati, 2013).

Kanker merupakan salah satu penyebab kematian utama bagi manusia di seluruh penjuru dunia. Pada tahun 2012, sekitar 8,2 juta kematian disebabkan oleh kanker, dengan spesifikasi kanker paru, hati, perut, kolorektal, dan kanker payudara adalah penyebab terbesar kematian akibat kanker setiap tahunnya (Infodatin, 2014). Penyakit yang ditandai dengan kelainan siklus sel khas yang menimbulkan kemampuan sel ini dapat menyebabkan pembelahan sel melebihi batas normal. Akibat yang ditimbulkan dalam pembelahan sel adalah menyerang jaringan biologis di dekatnya, dan dapat bermigrasi ke jaringan tubuh melalui sirkulasi darah (Sairun, 2015).

Kanker payudara (*breast cancer*) adalah jenis kanker paling umum yang diderita kaum wanita. Hingga saat ini, salah satu cara pengobatan yang umum dilakukan adalah dengan pembedahan dan jika perlu dilanjutkan dengan kemoterapi maupun radiasi. Namun pengobatan tersebut tidak akan memberikan dampak yang signifikan jika kankernya sudah mencapai stadium akhir. Oleh karena itu apa bila penyakit ini dapat dideteksi lebih awal, dampak buruk yang mungkin ditimbulkan oleh kanker payudara dapat dicegah (Zamani, Amaliah, & Munif, 2012). Penelitian-penelitian klasifikasi tumor payudara telah banyak dilakukan. Namun penelitian tersebut hanya sebatas mengetahui normal atau tidaknya payudara, ganas atau tidaknya tumor di payudara dan hanya menguji variabel-variabel yang berpengaruh terhadap hasil klasifikasi tumor payudara, seperti riwayat keluarga, terapi hormon, ukuran tumor, letak tumor, dan usia *menopause* (Makhfudhoh, 2014).

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk menggali potensi-potensi informasi yang ada. Para pengambil keputusan berusaha untuk memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna membantu mengambil keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah penggalian informasi atau pola yang penting atau

menarik dari data dalam jumlah besar, yang disebut dengan data *mining*. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi didalam gudang data sehingga menjadi informasi yang berharga (Huda, 2010).

Backward Elimination adalah metode untuk mengidentifikasi fitur dalam kumpulan data yang sama pentingnya, dan membuang semua fitur lain seperti informasi yang tidak *relevan* dan berlebihan (Apriana, 2013). Menurut Maimon dan Rokach dalam Apriana (2013) *Feature selection* adalah masalah berkaitan erat dengan pengurangan dimensi. Tujuan *feature selection* adalah untuk mengidentifikasi fitur dalam kumpulan data yang sama pentingnya, dan membuang semua fitur lain seperti informasi yang tidak *relevan* dan berlebihan. Karena *feature selection* mengurangi dimensi dari data, sehingga memungkinkan operasi yang lebih efektif dan algoritma *data mining* yang cepat (yaitu data algoritma *mining* dapat dioperasikan lebih cepat dan lebih efektif dengan menggunakan *feature selection*).

Metode klasifikasi yang kini banyak dikembangkan dan diterapkan adalah *Support Vector Machine* (SVM). Metode ini berakar dari teori pembelajaran statistik yang hasilnya sangat menjanjikan untuk memberikan hasil yang lebih baik dari pada metode yang lain. SVM juga bekerja dengan baik pada set data berdimensi tinggi (Prasetyo, 2012). Berdasarkan penelitian sebelumnya tentang perbandingan klasifikasi tingkat keganasan *breast cancer* dengan menggunakan *Regresi Logistik Ordinal* dan *Support Vector Machine* (SVM) (Rachman & Purnami, 2012). Kemudian, implementasi algoritma genetika pada struktur *Backpropagation Neural Network* untuk klasifikasi kanker payudara (Zamani, Amaliah, & Munif, 2012). Serta, analisis diagnosis pasien kanker payudara menggunakan *Regresi Logistik* dan *Support Vector Machine* (SVM) berdasarkan hasil *Mamografi* (Novianti & Purnami, 2012).

Kanker Payudara

Menurut data GLOBOCAN (IARC) tahun 2012 diketahui bahwa kanker payudara merupakan penyakit kanker dengan persentase kasus baru (setelah dikontrol oleh umur) tertinggi, yaitu sebesar 43,3%, dan persentase kematian (setelah dikontrol oleh umur) akibat kanker payudara sebesar 12,9% (Infodatin, 2015:2).

Kanker payudara adalah pertumbuhan sel yang abnormal pada jaringan payudara seseorang. Payudara wanita terdiri dari *lobulus* (kelenjar susu), *duktus* (saluran susu), lemak dan jaringan ikat, pembuluh darah dan *limfe*. Sebagian besar kanker payudara bermula pada sel-sel yang melapisi *duktus* (kanker *duktal*), beberapa bermula di *lobulus* (kanker *lobular*), serta sebagian kecil bermula di jaringan lain menurut Ellis, E.O, dkk dalam Novianti dan Purnami (2012:147).

Kanker payudara adalah jenis kanker yang terdapat pada jaringan payudara. Kanker payudara merupakan penyakit yang paling banyak menyerang kaum wanita. Kaum pria juga dapat terserang kanker payudara, walaupun kemungkinannya lebih kecil. Faktor-faktor yang menyebabkan kanker payudara menurut *American Cancer Society* dalam (Makhfudhoh, 2014) adalah: usia, jenis kelamin, riwayat reproduksi, riwayat keluarga, obesitas, dan konsumsi makanan lemak tinggi.

Parameter-parameter kanker payudara terdiri dari sebagai berikut menurut Frank, A. dan Asuncion dalam (Hermawanti, 2015):

1. *Clump thickness*: sel *benign* cenderung dikelompokkan dalam *monolayers*, sementara sel-sel kanker sering dikelompokkan dalam *multilayers*.
2. *Uniformity of cell size*: sel-sel kanker mempunyai ukuran bervariasi.
3. *Uniformity of cell shape*: sel-sel kanker mempunyai bentuk bervariasi.
4. *Marginal adhesion*: sel-sel normal cenderung tetap bersama-sama.
5. *Single epithelial cell size*: sel-sel epitel yang signifikan diperbesar menjadi sel *malignant*.
6. *Bare nuclei*: adalah istilah yang digunakan untuk inti (*nuclei*) yang tidak dikelilingi oleh cytoplasm (seluruh sel). Biasanya terlihat di *benign*.
7. *Bland Chromatin*: inti "tekstur" seragam yang dilihat dalam sel *benign*. Dalam sel-sel kanker *chromatin* cenderung lebih kasar.
8. *Normal nucleoli*: *nucleoli* adalah struktur kecil yang terlihat dalam inti atom. Pada sel-sel normal *nucleolus* biasanya sangat kecil jika terlihat sama sekali. Dalam sel-sel *cancer nucleoli* menjadi lebih menonjol.
9. *Mitoses* : pembelahan satu sel menjadi dua sel.

10. *Class* : kelas.

2. Metode Penelitian

Dalam Metode Algoritma ini penulis, menggunakan Metode Algoritma sebagai berikut:

Algoritma Support Vector Machine (SVM)

Menurut Vapnik dalam Rachman dan Purnami, (2012) *Support Vektor Machine* (SVM) melakukan suatu teknik untuk menemukan fungsi pemisah yang bisa memisahkan dua set data dari dua kelas yang berbeda. Menurut Nugroho, dkk, dalam Novianti dan Purnami (2012:149) mengemukakan "*Support Vector Machine* (SVM) pertama kali dikenalkan oleh Vapnik pada tahun 1992 pada saat dipresentasikan di *Annual Workshop on Computational Learning Theory*".

$$f(x) = \text{sign} \left(\sum_{i=1}^n y_i \hat{\alpha}_i (\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)) + \hat{b} \right)$$

$$= \text{sign} \left(\sum_{i=1}^n y_i \hat{\alpha}_i (K(\mathbf{x}_i, \mathbf{x}_j)) + \hat{b} \right)$$

dengan fungsi *sign* semua nilai $f(x) < 0$ diberi label -1 dan nilai $f(x) > 0$ diberi label +1.

Metode klasifikasi yang kini banyak dikembangkan dan diterapkan adalah *Support Vector Machine* (SVM). Metode ini berakar dari teori pembelajaran statistik yang hasilnya sangat menjanjikan untuk memberikan hasil yang lebih baik dari pada metode yang lain. SVM juga bekerja dengan baik pada set data berdimensi tinggi (Prasetyo, 2012).

Optimasi Backward Elimination

Menurut Han, J. dan Kamber dalam Hermawanti, (2015:44) *Backward Elimination* menghilangkan atribut-atribut yang tidak *relevan*. Sedangkan menurut Noori, dkk, dalam Hermawanti, (2015:44) Algoritma *Backward Elimination* didasarkan pada model *regresi linear*. Langkah-langkah *Backward Elimination* adalah:

1. Mulai semua variabel pada model F-statistik parsial dihitung setiap variabel pada model.
2. Menentukan variabel dengan F-statistik parsial terkecil dan menguji *Fmin*.
3. Jika *Fmin* tidak signifikan, dalam kasus ini, variabel dihilangkan dari model.
4. Menentukan variabel dengan F-statistik parsial .
5. Pada sisi lain, variabel dengan F-statistik terkecil adalah variabel

indikator. Bagaimanapun, *p-value* diasosiasikan dengan *F_{min}* tidak cukup membenarkan model yang tidak inklusi (*noninclusion*) menurut kriteria (lebih dari bit). Maka dari itu, prosedur menghasilkan dan melaporkan model sebagai berikut:

$$y = \beta_0 + \beta_1(\text{single epithelial cell size}) + \beta_2(\text{normal nucleoli}) + \beta_3(\text{marginal adhesion}) + \epsilon$$

6. Menghitung *F-test* parsial.

Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari jumlah kelas yang tersedia. Menurut (Prasetyo, 2012) dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu:

1. Pembangunan model sebagai prototipe untuk disimpan sebagai memori

Penggunaan model tersebut untuk melakukan pengenalan, klasifikasi, atau prediksi pada suatu objek data lain agar diketahui dikelas mana objek data tersebut dalam model yang sudah disimpnanya.

3. Pembahasan

3.1 Rancangan Algoritma

Pada tahap ini ditampilkan hasil eksperimen klasifikasi dataset Kanker Payudarah menggunakan algoritma *Support Vector Machine*. Tahap awal dilakukan identifikasi dataset, diketahui bahwa jumlah dataset WBC (*Wisconsin Breast Cancer*) sebanyak 699 *record*, tahap selanjutnya dilakukan pemisahan data *training* dan *testing* menggunakan 10 *fold cross validation* pembagian data menjadi data *training* 90% dan data *testing* 10%. Data *training* untuk dijadikan model dan data *testing* untuk mendapatkan akurasi dengan sebaran data pada tabel 1.

Tabel 1. Sebaran Dataset WBC (*Wisconsin Breast Cancer*)

NO	Jenis Klasifikasi	Jumlah Record Dataset	Data Training	Data Testing
1	<i>Malignant</i>	241	217	24
2	<i>Benign</i>	458	412	46
	Jumlah	699	629	70

Pada tabel 1. terlihat bahwa dataset WBC (*Wisconsin Breast Cancer*) diklasifikasikan kedalam 2 kelas yaitu klasifikasi *Malignant* dan *Benign*, pada dataset WBC (*Wisconsin Breast Cancer*) selain memiliki dua *class*

klasifikasi memiliki atribut sebanyak 10 atribut dengan penjelasan pada 2.

Tabel 2. Ringkasan 10 Fitur dalam Database

No	Atribut
1	<i>Clump Thickness</i>
2	<i>Uniformity of Cell Size</i>
3	<i>Uniformity of Cell Shape</i>
4	<i>Marginal Adhesion</i>
5	<i>Single Epithelial Cell Size</i>
6	<i>Bare Nuclei</i>
7	<i>Bland Chromatin</i>
8	<i>Normal Nucleoli</i>
9	<i>Mitoses</i>
10	<i>Class (Benign dan Malignant)</i>

3.2 Hasil Eksperimen Support Vector Machine

Pada tahap ini dijelaskan hasil eksperimen menggunakan Algoritma klasifikasi *Support Vector Machine*. dengan menggunakan atribut sebanyak 10 atribut dan menggunakan satu nilai bias, hasil klasifikasi model dari algoritma *Support Vector Machine* dijelaskan pada table 3.

Tabel 3. Kernel Model Algoritma Support Vector Machine

No	Atribut	Bobot
1	<i>w[Clump Thickness]</i>	0.518
2	<i>w[Uniformity of Cell Size]</i>	0.081
3	<i>w[Uniformity of Cell Shape]</i>	0.373
4	<i>w[Marginal Adhesion]</i>	0.156
5	<i>w[Single Epithelial Cell Size]</i>	0.181
6	<i>w[Bare Nuclei]</i>	0.598
7	<i>w[Bland Chromatin]</i>	0.392
8	<i>w[Normal Nucleoli]</i>	0.179
9	<i>w[Mitoses]</i>	0.323
10	<i>Bias (offset)</i>	-0.320

Pada tabel 3. dijelaskan *Kernel* model dari *Support Vector Machine* yang didapat dari *tools* Rapidminer, selanjutnya dilakukan proses identifikasi akurasi hasil klasifikasi. Selain menggambarkan model algoritma *Support Vector Machine*, *tools* Rapidminer menghasilkan hasil akurasi klasifikasi algoritma *Support Vector Machine* dengan nilai akurasi sebesar **97.00%**. Pada tabel 4 dijelaskan nilai dari akurasi yang dihasilkan oleh algoritma *Support Vector Machine*. Keterangan yang diberikan pada tabel

tersebut diantaranya, pred. *Benign* dan pred. *Malignant* merupakan prediksi dari *dataset* dikategorikan menjadi 2 kelas yaitu Pred. *Benign* = Jinak, Pred. *Malignant* = Ganas. Sedangkan untuk *True Benign*, *True Malignant*, hasil komparasi antara *dataset testing* dengan *dataset training*, dimana tingkat akurasi yang diberikan dari masing-masing prediksi maupun hasil akurasi prediksi dari keseluruhan *dataset*.

Tabel 4. Model *confusion matrix* Algoritma *Support Vector Machine*

	<i>true benign</i>	<i>true malignant</i>	<i>class precision</i>
pred. <i>benign</i>	446	9	98.02%
pred. <i>malignant</i>	12	232	95.08%
<i>class recall</i>	97.38%	96.27%	

Ket :

Benign = Jinak
Malignant = Ganas.

Pada gambar 1. menjelaskan *curve ROC (Receiver Operating Characteristic)* atau bentuk grafik dari algoritma *Support Vector Machine* sebelum menggunakan seleksi fitur *Backward Elimination* dengan akurasi 97.00% dan nilai *AUC* sebesar 0.995.



Gambar 1. *Curve ROC (Receiver Operating Characteristic)* Algoritma *Support Vector Machine*

3.3 Hasil Eksperimen *Support Vector Machine* dan *Backward Elimination*

Pada bagian ini akan ditampilkan hasil eksperimen dengan menggunakan algoritma *Support Vector Machine* dan seleksi fitur menggunakan *Backward Elimination* pada *dataset WBC (Wisconsin Breast Cancer)*.

Tabel 5. Ringkasan Atribut Optimasi Fitur BE dan Algoritma SVM

No	Atribut	Bobot
1	<i>w[Clump Thickness]</i>	0.617
2	<i>w[Uniformity of Cell Shape]</i>	0.461
3	<i>w[Marginal Adhesion]</i>	0.201
4	<i>w[Single Epithelial Cell Size]</i>	0.110
5	<i>w[Bare Nuclei]</i>	0.636
6	<i>w[Bland Chromatin]</i>	0.362
7	<i>w[Normal Nucleoli]</i>	0.238
8	<i>Bias (offset)</i>	-0.443

Pada tahap ini dijelaskan hasil eksperimen menggunakan Algoritma klasifikasi SVM dan seleksi fitur menggunakan *Backward Elimination* dengan menggunakan atribut sebanyak 10 dan *Output* sebanyak 2. Setelah dilakukan perhitungan klasifikasi Kanker Payudarah menggunakan seleksi fitur *Backward Elimination* dan algoritma klasifikasi SVM diperoleh atribut optimasi yang relevan terhadap *class* sebanyak 7 atribut dengan dijelaskan pada tabel 5.

Tabel 6. Atribut terpilih Model Algoritma *Support Vector Machine* dan *Backward Elimination*

No	Atribut
1	<i>w[Clump Thickness]</i>
2	<i>w[Uniformity of Cell Shape]</i>
3	<i>w[Marginal Adhesion]</i>
4	<i>w[Single Epithelial Cell Size]</i>
5	<i>w[Bare Nuclei]</i>
6	<i>w[Bland Chromatin]</i>
7	<i>w[Normal Nucleoli]</i>
8	<i>Bias (offset)</i>

Pada tabel 6 dan 7 dijelaskan model dari *Support Vector Machine* dan *Backward Elimination* yang didapat dari *tools Rapidminer*, selanjutnya dilakukan proses identifikasi akurasi hasil klasifikasi. Selain menggambarkan model algoritma *Support Vector Machine* dan *Backward Elimination*, *tools Rapidminer* menghasilkan hasil akurasi klasifikasi algoritma *Support Vector Machine* dan *Backward Elimination* dengan nilai akurasi sebesar **97.14%**. Pada tabel 7 dijelaskan nilai dari akurasi yang dihasilkan oleh algoritma *Support Vector Machine*. Keterangan yang diberikan pada tabel tersebut diantaranya, pred. *Benign* dan pred. *Malignant* merupakan prediksi dari *dataset* dikategorikan menjadi 2 kelas yaitu Pred. *Benign* = Jinak, Pred. *Malignant* = Ganas. Sedangkan untuk *True Benign*, *True Malignant*, hasil komparasi antara *dataset testing* dengan *dataset training*, dimana tingkat akurasi yang diberikan dari masing-

masing prediksi maupun hasil akurasi prediksi dari keseluruhan dataset.

Tabel 7. Model Confaction Matrix Algoritma Support Vector Machine dan Backward Elimination

	true benign	true malignant	class precision
pred. benign	446	8	98.24%
pred. malignant	12	233	95.10%
class recall	97.38%	96.68%	

Ket:

pred. benign = Jinak

pred. malignant = Ganas

Pada gambar 2 menjelaskan curve ROC (Receiver Operating Characteristic) atau bentuk grafik dari algoritma Support Vector Machine setelah menggunakan seleksi fitur Backward Elimination dengan akurasi 97.14% dan nilai AUC sebesar 0.995



Gambar 2. Curve ROC (Receiver Operating Characteristic) Algoritma Support Vector Machine dan Backward Elimination.

3.4 Perhitungan Algoritma Support Vector Machine

Pada tahap ini menjelaskan tentang perhitungan manual klasifikasi algoritma Support Vector Machine. Tahapan melakukan perhitungan manual algoritma Support Vector Machine dan seleksi fitur menggunakan Backward Elimination dijelaskan sebagai berikut:

a. Perhitungan Normalisasi

Pada tahap awal perhitungan algoritma Support Vector Machine dilakukan normalisasi data. Teknik standarisasi atau biasa juga disebut normalisasi data. Teknik standarisasi data numerik ke dalam skala baru agar memenuhi homogenitas

(persamaan) dan sebaran data. Untuk data kategori umumnya digunakan teknik discretize yang akan dibahas pada pembahasan Data Discretization. Teknik standarisasi data numerik yang digunakan adalah standarisasi Min-Max yang diperoleh melalui perhitungan sebagai berikut:

$$x'_{ij} = \frac{x_{ij} - x_{min,j}}{x_{max,j} - x_{min,j}} (x'_{max,j} - x'_{min,j}) + x'_{min,j}$$

Dimana :

$$x_{min,j} = \min_i x_{ij}, x_{max,j} = \max_i x_{ij}$$

Rumus Normalisasi = ((data-Min)/(Max-Min))*(1-(-1))+(-1)

Secara umum rentang nilai $x_{min,j}$ dan $x_{max,j}$ adalah [-1, 1] atau [0, 1].

Tabel 8. Nilai Min dan Max setiap Atribut

Atribut	Min	Max
Clump Thickness	1	10
Uniformity of Cell Shape	1	10
Marginal Adhesion	1	10
Single Epithelial Cell Size	1	10
Bare Nuclei	0	10
Bland Chromatin	1	10
Normal Nucleoli	1	10

b. Nilai Bobot Masing-masing Atribut SVM

Pada tahap ini mencari bobot setiap atribut yang dihasilkan oleh algoritma Support Vector Machine dan seleksi fitur menggunakan Backward Elimination, pada tahap sebelumnya sudah diketahui bahwa algoritma Support Vector Machine memiliki atribut sebanyak 10, setelah dilakukan optimasi fitur menggunakan Algoritma Backward Elimination didapat atribut yang relevan terhadap label atau kelas sebanyak 8 atribut seperti pada penjelasan tabel 9. jumlah atribut yang terseleksi dan nilai masing-masing atribut. Model klasifikasi Support Vector Machine dan seleksi fitur menggunakan yang dihasilkan oleh tools rapidminer, diolah menjadi perhitungan manual, sebagai rujukan untuk melakukan perhitungan manual klasifikasi dataset Kanker Payudara.

Tabel 9. Bobot dari SVM dan BE

No	Atribut	Bobot
1	$w[\text{Clump Thickness}]$	0.617

2	<i>w[Uniformity of Cell Shape]</i>	0.461
3	<i>w[Marginal Adhesion]</i>	0.201
4	<i>w[Single Epithelial Cell Size]</i>	0.110
5	<i>w[Bare Nuclei]</i>	0.636
6	<i>w[Bland Chromatin]</i>	0.362
7	<i>w[Normal Nucleoli]</i>	0.238
8	<i>Bias (offset)</i>	-0.443

c. Perhitungan Manual Algoritma SVM

Pada tahap ini menjelaskan tentang perhitungan manual klasifikasi algoritma *Support Vector Machine*. Tahapan melakukan perhitungan manual algoritma *Support Vector Machine* dijelaskan sebagai berikut:

- 1) Perhitungan Normalisasi

$$\text{Normalisasi} = \frac{(\text{data} - \text{Min})}{(\text{Max} - \text{Min})} * (1 - (-1)) + (-1)$$

$$= \frac{(7 - 1)}{(10 - 1)} * (1 - (-1)) + (-1)$$

$$= -0,33$$

Selanjutnya dilakukan proses perhitungan sampai dengan 10 atribut yang dinormlaisasi.

- 2) Nilai Bobot masing-masing atribut Dari model algoritma SVM (*Support Vector Machine*) yang didapat dari *tools* Rapidminer didapat nilai bobot masing-masing atribut, dimana nilai masing-masing bobot atribut dijadikan acuan model untuk dilakukan perhitungan manual dari algoritma *Support Vector Machine* (SVM).
Rumus:
Nilai Bobot 1 = Atribut 1 Normalisasi * Bobot Atribut
Lakukan sampai dengan 7 kali sesuai dengan jumlah Optimasi *Feature* dengan algoritma *Backward Elimination* dan algoritma SVM.
- 3) Jumlah Nilai Bobot *Output*
Rumus: Nilai Bobot *Output* = Nilai Bobot1 + Nilai Bobot2 + Nilai Bobot3 + Nilai Bobot4 + Nilai Bobot5 + Nilai Bobot6 + Nilai Bobot7 + Bias
- 4) Bobot Atribut Hasil Klasifikasi Rumus:
=SIGN(Nilai Bobot *Output*)

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan

bahwa optimasi fitur dengan menggunakan Algoritma *Backward Elimination* dapat menyeleksi fitur dengan baik diantara metode algoritma optimasi fitur yang telah penulis uji dengan menggunakan dataset WBC (*Wisconsin Breast Cancer*).

Metode Algoritma *Support Vector Machine* (SVM) adalah Algoritma yang baik untuk klasifikasi Kanker Payudara menggunakan Dataset WBC (*Wisconsin Breast Cancer*), dimana nilai klasifikasi performansi Akurasi dan AUC nya adalah yang tertinggi diantara algoritma yang telah penulis uji, sedangkan untuk penggabungan algoritma seleksi fitur *Backward Elimination* dan *Support Vector Machine* (SVM) mendapatkan peningkatan Akurasi sebesar 14% sehingga nilai tingkat akurasi akhirnya sebesar 97.14% dan nilai AUC mencapai 0.995.

Diharapkan untuk penelitian selanjutnya dilakukan pemisahan data training dan testing tidak hanya menggunakan 10 fold cross validation saja untuk pembagian data training 90% dan data testing 10 % agar dapat mengetahui perbedaan nilai akurasi.

Referensi

- Anggorowati, L. (2013). FAKTOR RISIKO KANKER PAYUDARA WANITA. *Jurnal Kesehatan Masyarakat*, 122.
- Arifin, T. (2015). Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur. *Jurnal Informatika*.
- Arifin, T. (2014). Klasifikasi Inti Sel Pap Smear Berdasarkan Analisis Tekstur Menggunakan Correlation-Based Feature Selection Berbasis Algoritma C4.5. *Jurnal Informatika*.
- Apriana, V. (2013). Prediksi Beban Listrik Jangka Pendek Berbasis Backward Elimination. *Pilar Nusa Mandiri*, 23-28.
- Hermawanti, L. (2015). Penggabungan Algoritma Backward Elimination Dan Naive Bayes. *Momentum*, 42-45.
- Huda, N. M. (2010). *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*.

Diponegoro: Universitas
Diponegoro.

Infodatin. (2015). *Stop Kanker*. Jakarta Selatan: Pusat data dan informasi kementrian RI.

Makhfudhoh, N. U. (2014). *Klasifikasi Kanker Payudara Dari Citra Mammografi Menggunakan Model Fuzzy*. Yogyakarta: Universitas Negeri Yogyakarta.

Novianti, F. A., & Purnami, S. W. (2012). analisis diagnosis pasien kanker payudara menggunakan regresi logistik dan support vector machine (SVM) berdasarkan hasil mamografi. *JURNAL SAINS DAN SENI ITS*, 147-152.

Prasetyo, E. (2012). *DATA MINING-Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI.

Rachman, F., & Purnami, S. W. (2012). Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer Dengan Menggunakan Regresi Logistik Ordinal Dan Support Vector Machine (SVM). *JURNAL SAINS DAN SENI ITS 2012*.

Sairun, W. N. (2015, Agustus 25). *Mengapa Kanker Paru dan Payudara Tertinggi di Indonesia*. Retrieved July 12, 2016, from Kompasiana: http://www.kompasiana.com/nol3lima/mengapa-kanker-paru-dan-payudara-tertinggi-di-indonesia_55dbf6f36e7a613105c59ea3

Zamani, A. M., Amaliah, B., & Munif, A. (2012). Implementasi Algoritma Genetika pada Struktur Backpropagation Neural Network untuk Klasifikasi Kanker Payudara. *JURNAL TEKNIK ITS*, 222.