

Klasifikasi Pasien Kanker Payudara Menggunakan Metode *Support Vector Machine* dengan *Backward Elimination*

¹Rina Resmiati, ²Toni Arifin

¹Sistem Informasi, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya,
²Teknik Informatika, Fakultas Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya
Jl. Sekolah Internasional No. 1-2, Antapani, Bandung 40282, Indonesia
*e-mail: rinaresmiati25@gmail.com

(received: 16 Januari 2021, revised: 9 Maret 2021, accepted: 1 April 2021)

Abstrak

Kanker payudara merupakan tumor ganas yang tumbuh pada sel-sel payudara dan dapat menyebar di antara jaringan atau organ di sekitar payudara dan berpindah ke bagian tubuh lainnya. Jika deteksi kanker dilakukan sejak dini, memungkinkan dilakukan penanganan yang lebih baik dan timbulnya sel-sel kanker dapat diatasi dengan segera dan dihentikan penyebarannya. Untuk membantu meningkatkan kemampuan pendeteksian otomatis dapat digunakan teknik *machine learning* dengan metode klasifikasi. Salah satu metode klasifikasi yang dapat digunakan yaitu metode *Support Vector Machine*. Pada penelitian ini, metode *Support Vector Machine* diterapkan pada *Breast Cancer Coimbra Data Set*. Penerapan *Backward Elimination* bertujuan untuk mengoptimalkan performa suatu model dengan sistem kerja pemilihan mundur dan memilih atribut yang paling relevan pada proses klasifikasi. Hasil penelitian klasifikasi pasien kanker payudara menggunakan metode *Support Vector Machine* menghasilkan nilai akurasi sebesar 65,22% dan nilai AUC sebesar 0,700 yang termasuk ke dalam kategori *Fair Classification*. Sedangkan hasil penelitian klasifikasi pasien kanker payudara menggunakan metode *Support Vector Machine* dengan *Backward Elimination* menghasilkan nilai akurasi sebesar 95,65% dan nilai AUC sebesar 1,000 yang termasuk ke dalam kategori *Excellent Classification*.

Kata Kunci: *backward elimination*, kanker payudara, klasifikasi, *support vector machine*

Abstract

Breast cancer is a malignant tumor that grows on the cells of the breast and can spread between tissues or organs around the breast and move to other parts of the body. If the detection of cancer is done early, it is possible to do better treatment and the emergence of cancer cells can be treated immediately and stopped spreading. To help improve automatic detection capabilities, machine learning techniques with classification methods can be used. One of the classification methods that can be used is the Support Vector Machine method. In this study, the Support Vector Machine method was applied to the Breast Cancer Coimbra Data Set. The application of Backward Elimination aims to optimize the performance of a model with a backward selection work system and select the most relevant attributes in the classification process. The results of the classification study of breast cancer patients using the Support Vector Machine method resulted in an accuracy value of 65.22% and an AUC value of 0.700 which was included in the Fair Classification category. Meanwhile, the results of the classification research of breast cancer patients using the Support Vector Machine method with Backward Elimination resulted in an accuracy value of 95.65% and an AUC value of 1,000 which is included in the Excellent Classification category.

Keywords: *backward elimination*, breast cancer, classification, *support vector machine*

1 Pendahuluan

Kanker merupakan penyakit yang dapat tumbuh hampir di semua organ atau jaringan tubuh [1]. Kanker adalah penyakit dimana terjadi kerusakan dan tumbuh tidak normal pada gen yang mengontrol regenerasi sel dalam tubuh manusia [2], yaitu terdapat kelainan siklus sel khas yang mengakibatkan

<http://sistemasi.ftik.unisi.ac.id>

kemampuan sel melakukan pembelahan sel lebih dari batas normal. Akibatnya dapat menyerang jaringan biologis di dekatnya dan melalui aliran darah dapat berpindah ke jaringan tubuh lainnya [3].

Salah satu penyebab kematian utama di seluruh dunia adalah penyakit kanker. Sekitar 8,2 juta orang meninggal disebabkan kanker pada tahun 2012 [4]. Berdasarkan data Globocan Tahun 2018, Indonesia berada pada peringkat 8 di Asia Tenggara dan 23 di Asia dalam hal jumlah penyakit kanker. Kanker payudara merupakan jenis kanker yang banyak diderita oleh kaum wanita di Indonesia dengan jumlah kejadian 42,1 per 100.000 orang dan jumlah kematian rata-rata 17 per 100.000 orang [5]. Prevalensi kanker bergerak cepat di negara berkembang karena semakin meningkatnya gaya hidup yang dapat memicu penyakit ini, seperti merokok, kurang beraktivitas secara fisik, dan melakukan diet “*westernized*” [6]. Studi epidemiologi juga mendukung bahwa obesitas erat kaitannya dengan perkembangan beberapa jenis kanker, seperti kanker payudara [7].

Kanker payudara adalah penyakit ganas yang menyerang jaringan payudara yang dapat berasal dari epitel duktus maupun lobulusnya [8]. Kanker payudara merupakan tumor ganas yang tumbuh tidak normal pada sel-sel payudara [9] yang dapat menyebar di antara jaringan atau organ di sekitar payudara dan berpindah ke bagian tubuh yang lain [10]. Kanker payudara sangat berpengaruh terhadap fisik dan mental kesehatan seorang wanita [11] seperti keterkejutan mental, takut, depresi bahkan bisa sampai panik dan melakukan hal-hal yang tidak terlalu penting [12]. Kanker payudara termasuk jenis kanker yang paling banyak menyerang kaum wanita di negara-negara dengan penghasilan menengah ke bawah yang peluang untuk melakukan pencegahan dan deteksi dini atau keduanya sedikit [2].

Mendeteksi kanker sejak dini memungkinkan untuk dilakukan penanganan yang lebih baik [10] dan timbulnya sel-sel kanker dapat diatasi dengan segera dan dihentikan penyebarannya [13]. Strategi penting untuk deteksi dini dan memastikan kemungkinan yang lebih besar untuk mendapatkan hasil yang baik dalam pengobatan yaitu skrining kanker payudara [14]. Skrining mammografi merupakan salah satu teknik deteksi dini kanker payudara [15] dan prosedur tunggal yang mampu mendeteksi kanker payudara stadium awal atau sebelum kanker itu jelas. Namun kasus klasifikasi mammografi yang tidak tepat, masih merupakan salah satu area untuk perbaikan dalam pendeteksian kanker payudara. Dengan demikian, masih terdapat tantangan untuk dapat menemukan alat deteksi yang efektif. Parameter tubuh, seperti yang diperoleh dari sampel darah, yaitu analisis darah rutin dapat digunakan untuk mendeteksi keberadaan kanker payudara dan dapat memberikan cara alternatif untuk mendeteksi kanker payudara dengan lebih baik bagi kaum wanita [16],[14].

Dalam bidang kecerdasan buatan, teknik *machine learning* diperkenalkan untuk membantu meningkatkan kemampuan pendeteksian otomatis [17]. Metode klasifikasi *machine learning* salah satu teknik data mining dapat membantu untuk mendeteksi penyakit kanker [9]. Data mining merupakan suatu proses untuk menganalisis dataset besar yang setiap tahunnya jumlah data semakin bertambah, mengekstraksi dan menemukan informasi atau pola dan pengetahuan yang ada dalam data tersebut [18], [19].

Salah satu metode yang baik untuk masalah klasifikasi dan regresi baik linier maupun nonlinier adalah *Support Vector Machine* [20]. Metode *Support Vector Machine* pada awalnya digunakan untuk klasifikasi data berbentuk linier, namun kini dikembangkan agar dapat digunakan untuk data berbentuk nonlinier dengan menarapkan *kernel trick* [21]. Cara kerja metode ini yaitu mencari *hyperplane* dengan margin yang maksimal untuk memisahkan antar kelas klasifikasi. Metode SVM merupakan metode klasifikasi dengan waktu komputasi yang relatif cepat [22]. Beberapa penelitian membuktikan bahwa SVM unggul dalam melakukan klasifikasi, seperti pada penelitian tentang analisa perbandingan tingkat performansi metode *Support Vector Machine* dan *Naïve Bayes classifier* untuk klasifikasi jalur minat SMA [23], kemudian perbandingan metode data mining SVM dan NN untuk klasifikasi penyakit ginjal kronis [24] dan perbandingan metode prediksi *Support Vector Machine* dan *Linear Regression* menggunakan *Backward Elimination* pada produksi minyak kelapa [25]. Dengan kelebihan yang dimiliki metode SVM, ternyata metode ini pun memiliki kelemahan yaitu sulitnya memilih fitur yang relevan dan optimal pada nilai atribut yang digunakan sehingga mengakibatkan nilai akurasi menjadi kecil [26].

Salah satu metode untuk mengoptimalkan akurasi yaitu menggunakan seleksi fitur, salah satu diantaranya adalah *Backward Elimination*. Metode *Backward Elimination* merupakan metode yang bertujuan untuk mengoptimalkan performa suatu model dengan sistem kerja pemilihan mundur [27]. Metode ini digunakan sebagai seleksi fitur dengan cara pemilihan ke depan kemudian menguji semua

<http://sistemasi.ftik.unisi.ac.id>

atribut, atribut-atribut yang dianggap tidak sesuai dan tidak terlalu berpengaruh terhadap data yang diolah akan dibuang [3], [26], [28], [29]. Metode *backward elimination* akan membuang satu atribut yang tidak sesuai dan menghasilkan subset atribut baru setiap melakukan perulangan [30]. Pada saat dibandingkan dengan cara statistik dalam menyeleksi fitur, metode ini menghasilkan performa yang lebih baik yang diperoleh dari sensitivitas, spesifisitas dan keakuratan [27].

Penelitian ini bertujuan untuk mengetahui performa dari metode *Support Vector Machine* dengan seleksi fitur *Backward Elimination* dalam klasifikasi pasien kanker payudara. Dataset yang digunakan yaitu *Breast Cancer Coimbra Data Set* dengan parameter yang digunakan diantaranya *age*, BMI, *glucose*, insulin, HOMA, leptin, *adiponectin*, resistin dan MCP.1.

2 Tinjauan Pustaka

Penelitian tentang klasifikasi kanker payudara dengan metode yang beragam telah banyak dilakukan. Seperti dalam [31], [3], [14], [9] dan [32]. Penelitian Farahdiba dan Nugroho [31] menggunakan algoritma *Gain Ratio* untuk klasifikasi kanker payudara termasuk kanker payudara ganas atau jinak. Dataset yang digunakan yaitu *Winconsin Breast Cancer Dataset*. Atribut yang digunakan dalam klasifikasi yaitu ketebalan rumpun, keseragaman ukuran sel, keseragaman bentuk sel, adhesi marjinal, ukuran sel epitel tunggal, ukuran asli nuclei, kromatin, keadaan nucleoli normal dan mitosis. Algoritma *gain ratio* dalam klasifikasi jenis kanker payudara menghasilkan nilai akurasi sebesar 95,17%, *recall* sebesar 92,55% dan *precision* sebesar 93,76%.

Ma'arif dan Arifin [3] melakukan optimasi metode SVM dengan seleksi fitur *Backward Elimination* untuk mengklasifikasikan jenis kanker payudara ganas atau jinak. Dataset yang digunakan yaitu *Winconsin Breast Cancer Dataset* dan menyatakan bahwa *Support Vector Machine* (SVM) baik untuk klasifikasi kanker payudara pada dataset WBC dan nilai performansi akurasi dan AUC nya tertinggi diantara algoritma lain yang telah uji. Kemudian optimasi fitur dengan menggunakan algoritma *Backward Elimination* dapat menyeleksi fitur dengan baik diantara seleksi fitur lain yang telah diuji dan penggabungan antara SVM dan *Backward Elimination* mampu meningkatkan akurasi sebesar 14%, sehingga menghasilkan akurasi sebesar 97,14%.

Penelitian oleh Patricio et al. [14] pada tahun 2018 menggunakan metode *Logistic Regression*, *Random Forest* dan *Support Vector Machine* untuk mengembangkan dan menilai model prediksi yang dapat digunakan sebagai *biomarker* kanker payudara, berdasarkan data *antropometrik* dan parameter yang dikumpulkan dalam analisis darah rutin. Penelitiannya menggunakan data 166 peserta dengan beberapa fitur klinis, termasuk usia, BMI, Glukosa, Insulin, HOMA, Leptin, Adiponektin, Resistin dan MCP-1. Metode SVM menggunakan Glukosa, Resistin, Usia dan BMI sebagai prediktor memungkinkan memprediksi keberadaan kanker payudara pada wanita dengan sensitivitas berkisar antara 82 dan 88% dan spesifisitas berkisar antara 85 dan 90%. Interval kepercayaan 95% untuk AUC adalah [0,87, 0,91].

Penelitian Artha et al. [9] menggunakan metode V-ELM untuk mengklasifikasikan pengidap kanker payudara, yaitu menggunakan dataset *Breast Cancer Coimbra*. Penelitiannya menguji banyak *hidden neuron* yaitu pada rentang nilai 5 sampai dengan 20 menghasilkan nilai akurasi yang meningkat, sehingga menghasilkan akurasi tertinggi sebesar 89,564%.

Madyaningrum dan Sulastri [32] menggunakan metode *K-nearest Neighbor* untuk memprediksi kekambuhan penyakit kanker payudara dengan dataset *Breast Cancer*. Pada penelitian ini dilakukan tiga kali percobaan dengan komposisi data *training* dan data *testing* yang berbeda. Pada percobaan pertama yaitu data *training* 80% dan data *testing* 20% menghasilkan akurasi terbaik sebesar 80%.

3 Metode Penelitian

Menurut Dawson [33], metode penelitian yang paling umum digunakan ada empat yaitu penelitian tindakan, eksperimen, studi kasus, dan survei. Pada penelitian ini dilakukan penelitian eksperimen dengan pendekatan kuantitatif. Eksperimen yaitu jenis penelitian yang melakukan penyelidikan hubungan ketergantungan dengan menggunakan tes dimana yang mengendalikan adalah peneliti itu sendiri [33]. Penelitian ini dilakukan dengan tahapan seperti pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini yaitu *Breast Cancer Coimbra Dataset* yang merupakan data publik atau data sekunder yang bukan berasal langsung dari penelitian tetapi dari sumber lain, yaitu diambil dari situs web *University of California Irvine Machine Learning Repository* atau lebih dikenal dengan *UCI Repository* yang merupakan kumpulan *database*, teori *domain* dan *data generator* yang digunakan oleh komunitas *machine learning* untuk analisis empiris dari algoritma *machine learning* [34]. Adapun dataset *Breast Cancer Coimbra* dapat diakses di alamat <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>. Dataset *Breast Cancer Coimbra* merupakan hasil penelitian dari Miguel Patricio, Jose Pereira, Joana Crisostomo, Paulo Matafome, Raquel Seica dan Francisco Caramelo yang melibatkan 166 peserta di *University Hospital Centre of Coimbra* untuk pemeriksaan kanker payudara. Dataset tersebut pada tahun 2018 dipublikasikan melalui situs web *UCI Repository*.

Dataset *Breast Cancer Coimbra* terdiri dari 116 *record*, memiliki 9 atribut dan 1 label atau atribut target (deskripsi atribut dapat dilihat pada Tabel 1. Label pada dataset ini terdiri dari dua kelas yaitu 1 (Sehat) dan 2 (Pasien), dimana 52 *record* pada kelas 1 (Sehat) dan 64 *record* pada kelas 2 (Pasien).

Tabel 1. Deskripsi Atribut Dataset *Breast Cancer Coimbra*

No	Atribut	Satuan	Deskripsi
1	Age	Years	Usia pasien
2	BMI	kg/m ²	<i>Body Mass Index</i> (Indeks massa tubuh)
3	Glucose	mg/dL	Kadar glukosa dalam serum
4	Insulin	μU/mL	Kadar insulin dalam serum
5	HOMA	-	Nilai indeks <i>Homeostasis Model Assessment</i>
6	Leptin	ng/mL	Kadar leptin dalam serum
7	Adiponectin	μU/mL	Kadar adiponektin dalam serum
8	Resistin	mg/dL	Kadar resistin dalam serum
9	MCP.1	pg/dL	Kadar <i>Monocyte Chemoattractant Protein 1</i> dalam serum
10	Classification	-	1 = <i>Healthy Control</i> (Sehat) 2 = <i>Patients</i> (Pasien)

3.2 Pengolahan Data Awal

Pada tahap ini, dibutuhkan eksplorasi atau pendalaman terhadap *Breast Cancer Coimbra Dataset*. Eksplorasi ini dilakukan untuk memastikan semua atribut dan kelas dalam dataset tersebut *valid*, sehingga baik untuk dijadikan objek penelitian. Dengan begitu, tujuan untuk mengetahui hasil klasifikasi terbaik dari pasien kanker payudara dapat tercapai.

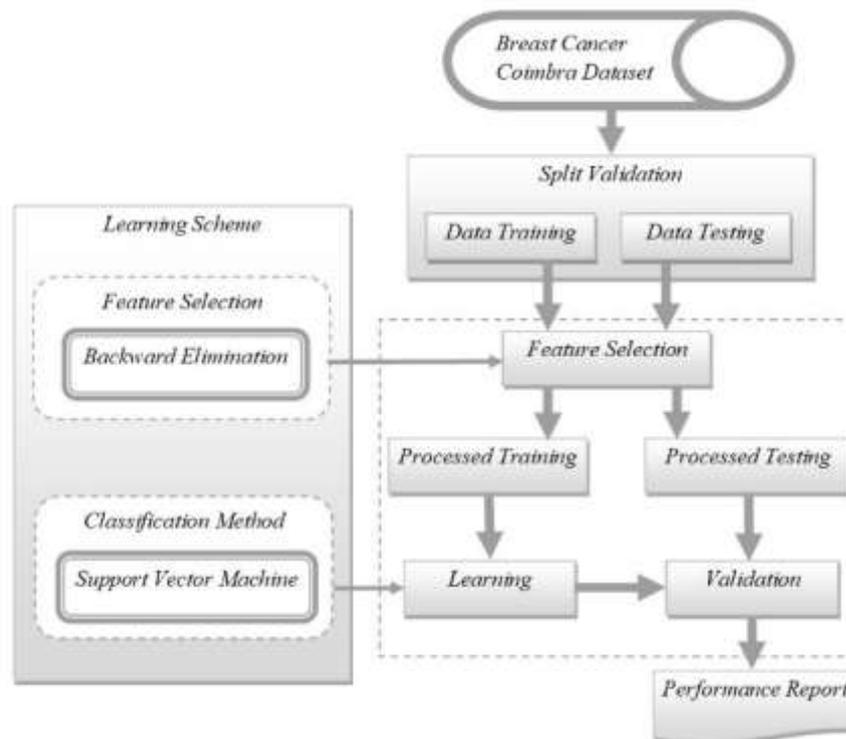
Untuk pengujian model yang digunakan, data akan dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan menggunakan *Split Validation*, yaitu besaran data *training* dan data *testing* ditentukan secara manual. Adapun pada penelitian ini, besaran pembagian data adalah 80% untuk data *training* dan 20% untuk data *testing*. Data *training* untuk pengembangan model dan data *testing* untuk menguji model. Sebaran dataset dapat dilihat pada Tabel 2.

Tabel 2. Sebaran Dataset Breast Cancer Coimbra

No	Kelas	Jumlah Record	Data Training	Data Testing
1	Healthy Control (Sehat)	52	42	10
2	Patients (Pasien)	64	51	13
	Jumlah	116	93	23

3.3 Model yang Diusulkan

Model yang akan digunakan pada penelitian ini adalah metode klasifikasi *Support Vector Machine* dan *Backward Elimination* sebagai seleksi fitur. Model yang diusulkan dapat dilihat pada Gambar 2.



Gambar 2. Model yang Diusulkan

Pada Gambar 2, penelitian dimulai dengan melakukan pembagian data menjadi dua bagian yaitu data *training* sebanyak 80% dan data *testing* sebanyak 20%. Besaran pembagian data *training* dan *testing* 80:20 karena sebelumnya pernah diuji menggunakan jumlah lain seperti 60:40, 70:30, 90:10, dan hasilnya tidak lebih baik dari 80:20 atau 80% untuk data *training* dan

20% untuk data *testing*. Kemudian diterapkan seleksi fitur *Backward Elimination* untuk menghilangkan atribut yang tidak relevan dalam proses klasifikasi. Tahapan *Backward Elimination* adalah sebagai berikut [35]:

1. Mulailah dengan menggunakan semua variabel prediktor dalam model.
2. Hapus variabel prediktor yang memiliki *p-value* tertinggi atau lebih besar dari nilai target yang telah ditentukan.
3. Perbaiki model dan ulangi langkah kedua.
4. Hentikan proses eliminasi jika *p-value* kurang dari nilai target.

Metode *Support Vector Machine* digunakan untuk mengklasifikasikan pasien kanker payudara dan menghasilkan bobot dari setiap atribut. Tahapan *Support Vector Machine* adalah sebagai berikut:

1. Mencari nilai *weight* (w) dan bias (b)
Untuk mencari nilai *weight* dan bias dapat menggunakan persamaan (1).

$$Y_i (w \cdot x_i + b) \geq 1 \quad (1)$$

2. Menentukan nilai *hyperplane*
Untuk menentukan nilai *hyperplane* dapat menggunakan persamaan (2).

$$f(x) = w_1x_1 + w_2x_2 + w_nx_n + b = 0 \quad (2)$$

3. Menguji dan mengklasifikasi data
Untuk menguji dan mengklasifikasi data dapat menggunakan persamaan (3).

$$f(x) = \text{sign}(w_1x_1 + w_2x_2 + w_nx_n + b) \quad (3)$$

Proses seleksi fitur menggunakan *Backward Elimination* dilakukan bersamaan dengan proses klasifikasi menggunakan *Support Vector Machine* agar mendapatkan set atribut yang paling relevan. Kemudian, hasil performa pada model ini akan diukur dengan *Area Under Curve* (AUC).

3.4 Eksperimen dan Pengujian Model

Pada tahap ini, dilakukan eksperimen dengan cara pemodelan data yang akan diproses menggunakan metode yang telah diusulkan pada sebuah *software*. Penelitian ini, melakukan eksperimen terhadap klasifikasi pasien kanker payudara menggunakan metode *Support Vector Machine* dengan *Backward Elimination* menggunakan *Software Rapidminer Studio 9.6*.

Spesifikasi *hardware* yang digunakan pada penelitian ini yaitu sebagai berikut:

1. Laptop Acer Aspire E1 471
2. Processor Intel core i3 2348M (2.3GHz, 3MB L3 cache)
3. Intel HD Graphics 3000
4. 14.0" HD LED LCD
5. RAM 4 GB
6. HDD 500 GB

3.5 Evaluasi dan Validasi

Validasi menggunakan *confusion matrix* dan kurva ROC (*Receiver Operating Characteristic*). Untuk menghitung *confusion matrix* dinyatakan dalam persamaan 4 sampai dengan persamaan 6. Parameter TP (*True Positive*) menunjukkan sampel yang bernilai *true* diklasifikasi secara benar, FN (*False Negative*) menunjukkan sampel yang bernilai *false* diklasifikasi secara salah, FP (*False Positive*) menunjukkan sampel yang bernilai *false* diklasifikasi secara benar, TN (*True Negative*) menunjukkan sampel yang bernilai *true* diklasifikasi secara salah. Adapun untuk AUC atau kurva ROC dinyatakan sebagai berikut [28]:

- a. Akurasi 0.90 – 1.00 = *Excellent classification*
- b. Akurasi 0.80 – 0.90 = *Good classification*

- c. Akurasi 0.70 – 0.80 = *Fair classification*
- d. Akurasi 0.60 – 0.70 = *Poor classification*
- e. Akurasi 0.50 – 0.60 = *Failure classification*

Nilai *accuracy*, *sensitivity* dan *specificity* dapat dihitung menggunakan persamaan (4), (5) dan (6) berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

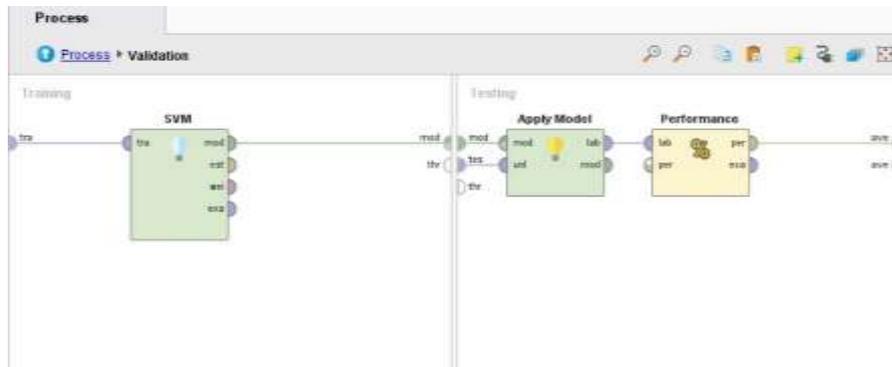
$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

4 Hasil dan Pembahasan

Penerapan metode *Support Vector Machine* dan seleksi fitur *Backward Elimination* pada *Breast Cancer Coimbra Dataset* bertujuan untuk mengetahui dan meningkatkan nilai akurasi dari penelitian-penelitian sebelumnya, sehingga akan dihasilkan set atribut terbaik untuk klasifikasi pasien kanker payudara. Meningkat atau tidaknya nilai akurasi dapat dilihat pada hasil eksperimen. Eksperimen ini dilakukan dengan dua tahap, pertama adalah eksperimen menggunakan metode *Support Vector Machine* dan kedua adalah eksperimen menggunakan metode *Support Vector Machine* dengan *Backward Elimination*.

4.1 Hasil Metode *Support Vector Machine*

Eksperimen ini dilakukan pada *tools Rapidminer* dengan pemodelan seperti pada Gambar 3.



Gambar 3. Pemodelan Metode *Support Vector Machine*

Hasil eksperimen menggunakan metode *Support Vector Machine* dengan menampilkan nilai *weight* dari setiap atribut pada dataset *Breast Cancer Coimbra* yang terdiri dari 9 atribut. *Kernel model* yang dihasilkan oleh metode *Support Vector Machine* dapat dilihat pada Tabel 3.

Tabel 3. *Kernel Model Support Vector Machine*

No	Atribut	Weight
1	w[Age]	-0.162
2	w[BMI]	-0.421
3	w[Glucose]	1.046
4	w[Insulin]	0.326
5	w[HOMA]	0.287
6	w[Leptin]	-0.300
7	w[Adiponectin]	-0.084
8	w[Resistin]	0.520
9	w[MCP.1]	-0.031
10	<i>Bias (offset)</i>	0.222

Pada Tabel 3 merupakan *kernel model* dengan nilai *weight* dan bias dari *Support Vector Machine* yang diperoleh dengan *tools Rapidminer*. Selanjutnya akan dijelaskan nilai akurasi yang diperoleh metode *Support Vector Machine*.

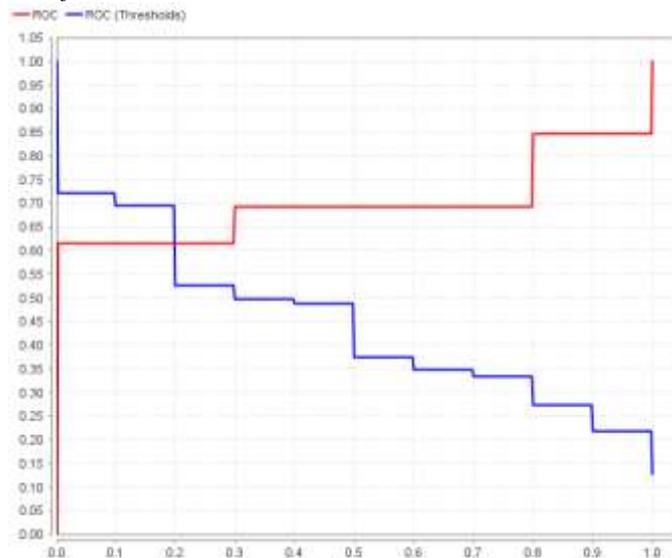
Tabel 4. Confusion Matrix Metode Support Vector Machine

	True 1 (Healthy Control)	True 2 (Patients)	Precision
Pred.1 (Healthy Control)	7	5	58.33%
Pred. 2 (Patients)	3	8	72.73%
Recall	70.00%	61.54%	

Pada Tabel 4 memberikan keterangan prediksi 1 untuk *Healthy Control* (Sehat) dan prediksi 2 untuk *Patients* (Pasien) yang merupakan hasil klasifikasi dengan metode *Support Vector Machine*. Sedangkan untuk *True 1 (Healthy Control)* dan *True 2 (Patients)* merupakan hasil klasifikasi dari data *testing*. Dari Tabel 4 dapat dihitung persentase nilai akurasi sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{7+8}{7+8+3+5} \times 100\% \\
 &= 65,22\%
 \end{aligned}$$

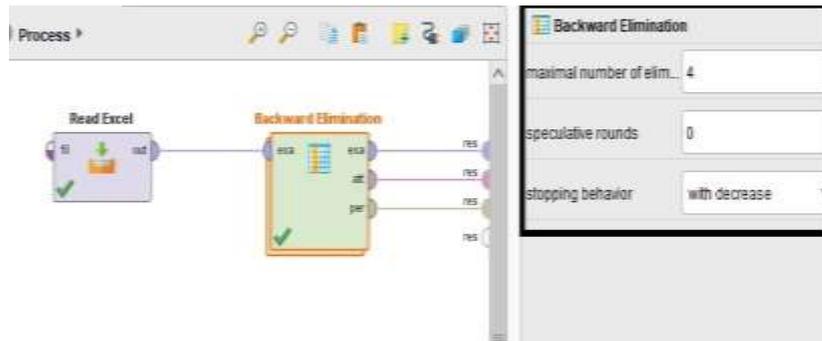
Selain menghasilkan *confussion matrix*, pengujian tersebut menghasilkan kurva ROC (*Receiver Operating Characteristic*) seperti pada Gambar 4 dengan nilai AUC sebesar 0,700 dan termasuk ke dalam kategori *Fair Classification*.



Gambar 4. Grafik ROC Metode Support Vector Machine

4.2 Hasil Metode Support Vector Machine dengan Backward Elimination

Penggunaan seleksi fitur dimaksudkan untuk menghilangkan atribut yang tidak relevan pada suatu dataset dan untuk meningkatkan nilai akurasi dari klasifikasi pasien kanker payudara. Adapun penerapan metode *Backward Elimination* pada *tools Rapidminer* dapat dilihat pada Gambar 5. Hasil eksperimen terhadap dataset *Breast Cancer Coimbra* yang sudah dilakukan seleksi fitur dengan *Backward Elimination* dapat dilihat pada Tabel 5 dan Tabel 6.



Gambar 5. Penerapan Seleksi Fitur *Backward Elimination*

Pada Tabel 5 menampilkan hasil eksperimen terhadap dataset *Breast Cancer Coimbra* yang sudah dilakukan seleksi fitur dengan *Backward Elimination*. Sebelum dilakukan seleksi fitur, jumlah atribut sebanyak 9 atribut. Sedangkan setelah dilakukan seleksi fitur jumlah atribut berkurang menjadi 8 atribut.

Tabel 5. Atribut Hasil Seleksi Fitur *Backward Elimination*

No	Atribut
1	BMI
2	Glucose
3	Insulin
4	HOMA
5	Leptin
6	Adiponectin
7	Resistin
8	MCP.1

Seleksi fitur dilakukan dengan menghilangkan suatu atribut bersamaan dengan melakukan pemodelan menggunakan metode *Support Vector Machine*. *Backward Elimination* akan menghilangkan atribut yang tidak relevan dan kurang berpengaruh terhadap nilai akurasi, dimana jika suatu atribut dihilangkan dari model maka akan menunjukkan peningkatan nilai akurasi. Atribut yang dihilangkan oleh *Backward Elimination* yaitu atribut *Age*.

Tabel 6. Confusion Matrix Metode *Support Vector Machine* Dengan *Backward Elimination*

	True 1 (Healthy Control)	True 2 (Patients)	Precision
Pred. 1 (Healthy Control)	9	0	100%
Pred. 2 (Patients)	1	13	92.86%
Recall	90.00%	100.00%	

Dari Tabel 6 dapat dihitung persentase nilai akurasi sebagai berikut:

$$\begin{aligned}
 Accuracy &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{9+13}{9+13+1+0} \times 100\% \\
 &= 95,65\%
 \end{aligned}$$

Kurva ROC (*Receiver Operating Characteristic*) yang dihasilkan pada pengujian tersebut dapat dilihat pada Gambar 6 dengan menghasilkan nilai AUC sebesar 1,000 dan termasuk ke dalam kategori *Excellent Classification*.



Gambar 6. Grafik ROC Metode Support Vector Machine Dengan Backward Elimination

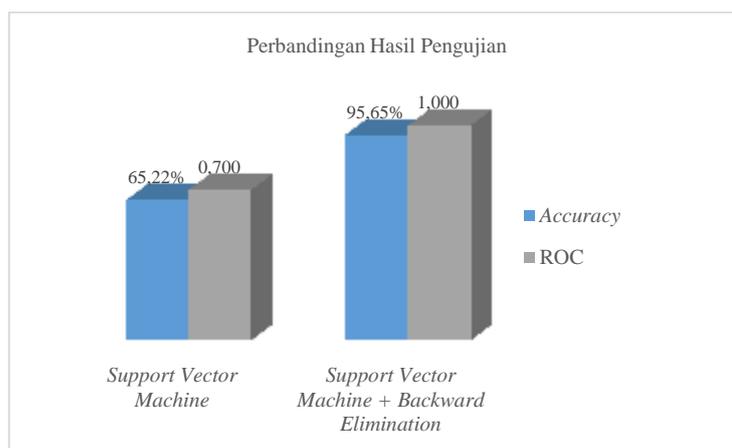
Berdasarkan hasil penelitian yang telah dijelaskan di atas, dilakukan evaluasi untuk mengetahui peningkatan nilai akurasi dari pengujian klasifikasi yang dihasilkan oleh metode *Support Vector Machine* sebelum dilakukan seleksi fitur dan *Support Vector Machine* setelah dilakukan seleksi fitur dengan *Backward Elimination*, yaitu dengan cara menghitung jumlah data *testing* yang diklasifikasikan secara benar dalam bentuk nilai akurasi dan dengan grafik ROC.

Secara keseluruhan penerapan seleksi fitur *Backward Elimination* mampu meningkatkan nilai akurasi dari metode *Support Vector Machine*, hal ini sejalan dengan penelitian [27]. Perbandingan nilai akurasi dan AUC pada *Support Vector Machine* sebelum dan setelah dilakukan seleksi fitur dengan *Backward Elimination* dapat dilihat pada tabel 7.

Tabel 7. Perbandingan Hasil Pengujian

No	Metode	Accuracy	AUC
1	<i>Support Vector Machine</i>	65,22%	0,700
2	<i>Support Vector Machine + Backward Elimination</i>	95,65%	1,000

Pada Gambar 7 dapat dilihat bahwa hasil klasifikasi metode *Support Vector Machine* mengalami peningkatan sebesar 30,43% setelah dilakukan seleksi fitur dengan *Backward Elimination*. Hal ini karena *Backward Elimination* akan menghilangkan atribut yang tidak relevan dalam proses klasifikasi, seperti pada penelitian [3], [26], [28], [29]. Sehingga terbukti bahwa penggunaan seleksi fitur *Backward Elimination* dapat meningkatkan nilai akurasi metode *Support Vector Machine* pada klasifikasi pasien kanker payudara dan membuktikan bahwa atribut *Age* adalah atribut yang tidak relevan dalam klasifikasi pasien kanker payudara.



Gambar 7. Grafik Perbandingan Hasil Pengujian

5 Kesimpulan

Berdasarkan uraian di atas tentang penerapan metode *Support Vector Machine* dengan *Backward Elimination* pada klasifikasi pasien kanker payudara menggunakan *software Rapidminer Studio 9.6*, maka penulis menyimpulkan bahwa penerapan metode *Support Vector Machine* dalam klasifikasi pasien kanker payudara menghasilkan nilai akurasi sebesar 65,22% dan AUC sebesar 0,700 dan termasuk ke dalam kategori *Fair Classification*. Adapun hasil klasifikasi metode *Support Vector Machine* setelah diterapkan seleksi fitur *Backward Elimination* menunjukkan peningkatan nilai akurasi sebesar 30,43%, sehingga nilai akurasi yang dihasilkan sebesar 95,65% dan AUC sebesar 1,000 dan termasuk ke dalam kategori *Excellent Classification*.

Bagi penelitian berikutnya terdapat beberapa hal yang dapat dilakukan untuk meningkatkan nilai akurasi dan melakukan klasifikasi pasien kanker payudara dengan lebih baik lagi, diantaranya dapat menggunakan metode lain seperti metode *Regresi Linear*, *Decision Tree* dan metode lainnya. Selain itu, dapat menerapkan seleksi fitur lainnya seperti *Forward Selection* untuk mendapatkan set atribut terbaik yang terpilih oleh seleksi fitur yang berbeda sehingga mampu meningkatkan nilai akurasi.

Referensi

- [1] WHO, "Cancer," WHO, 2019. https://www.who.int/health-topics/cancer#tab=tab_1 (accessed Apr. 03, 2020).
- [2] N. P. W. P. Sari, "Women Living With Breast and Cervical Cancer in the Community: The Face of Surabaya Nowadays," *Indones. J. Cancer*, vol. 12, no. 4, pp. 116–122, 2019, doi: 10.33371/ijoc.v12i4.605.
- [3] F. Ma'arif and T. Arifin, "Optimasi Fitur Menggunakan Backward Elimination Dan Algoritma SVM Untuk Klasifikasi Kanker Payudara," *J. Inform.*, vol. 4, no. 1, pp. 46–53, 2017, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/1548/pdf>.
- [4] Pusdatin, "Situasi Penyakit Kanker," *Buletin Jendela Data & Informasi Kesehatan*, Jakarta, pp. 1689–1699, 2015.
- [5] K. Khadijah and R. Kusumaningrum, "Ensemble Classifier untuk Klasifikasi Kanker Payudara," *It J. Res. Dev.*, vol. 4, no. 1, pp. 61–71, 2019, doi: 10.25299/itjrd.2019.vol4(1).3540.
- [6] Yuliana, "Risiko dan Deteksi Dini Kanker Payudara," *OPINI*, vol. 45, no. 2, pp. 144–149, 2018, [Online]. Available: [http://www.kalbemed.com/Portals/6/21_261Opini-Risiko dan Deteksi Dini Kanker Payudara.pdf](http://www.kalbemed.com/Portals/6/21_261Opini-Risiko%20dan%20Deteksi%20Dini%20Kanker%20Payudara.pdf).
- [7] J. Crisóstomo *et al.*, "Hyperresistinemia and metabolic dysregulation: a risky crosstalk in obese breast cancer," *Endocrine*, vol. 53, no. 2, pp. 433–442, 2016, doi: 10.1007/s12020-016-0893-x.
- [8] Komite Penanggulangan Kanker Nasional, *Panduan Penatalaksanaan Kanker Payudara*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2015.
- [9] D. T. Artha, S. Adinugroho, and P. P. Adikara, "Klasifikasi Pengidap Kanker Payudara

- Menggunakan Metode Voting Based Extreme Learning Machine (V-ELM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 3, pp. 2180–2186, 2019.
- [10] T. A. Y. Siswa and Prihandoko, “Analisis Penerapan Optimasi Perbandingan Kinerja Algoritma C4.5 dan Naïve Bayes Berbasis Particle Swarm Optimization (PSO) Untuk Mendeteksi Kanker Payudara,” *Bangkit Indones.*, vol. 7, no. 2, pp. 1–9, 2018.
- [11] R. Preetha and S. V. Jinny, “A research on breast cancer prediction using data mining techniques,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 11S2, pp. 362–370, 2019, doi: 10.35940/ijitee.K1058.09811S219.
- [12] A. Wibowo, “Aplikasi Diagnosis Penyakit Kanker Payudara Menggunakan Algoritma Sequential Minimal Optimization,” *J. Teknol. dan Sist. Komput.*, vol. 5, no. 4, pp. 153–158, 2017, doi: 10.14710/jtsiskom.5.4.2017.153-158.
- [13] F. Shahura, O. Soesanto, and F. Indriani, “Penerapan Metode Rbpnn Untuk Klasifikasi Kanker Payudara,” *Klik - Kumpul. J. Ilmu Komput.*, vol. 3, no. 2, pp. 135–145, 2016, doi: 10.20527/klik.v3i2.51.
- [14] M. Patrício *et al.*, “Using Resistin, glucose, age and BMI to predict the presence of breast cancer,” *BMC Cancer*, vol. 18, no. 1, pp. 1–8, 2018, doi: 10.1186/s12885-017-3877-1.
- [15] K. N. Setiawan and I. M. S. Putra, “Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM),” *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 6, no. 1, pp. 13–24, 2018, doi: 10.24843/jim.2018.v06.i01.p02.
- [16] Y. D. Austria, M. L. Goh, L. Sta. Maria Jr., J.-A. Lalata, J. E. Goh, and H. Vicente, “Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset,” *Int. J. Simul. Syst. Sci. Technol.*, pp. 1–8, 2019, doi: 10.5013/ijssst.a.20.s2.23.
- [17] F. S. Nugraha, M. J. Shidiq, and S. Rahayu, “Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Kanker Payudara,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 149–156, 2019, doi: 10.33480/pilar.v15i2.601.
- [18] Y. Ramdhani and A. Mubarak, “Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi,” *RESPONSIF*, vol. 1, no. 1, pp. 77–82, 2019.
- [19] A. Supriyatna and W. P. Mustika, “Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 2, no. 2, pp. 152–161, 2018, doi: 10.30645/j-sakti.v2i2.78.
- [20] Fitriyani and R. Sanjaya, “Komparasi Algoritma Lr , K-Nn Dan Svm Untuk Estimasi,” *Infotronik*, vol. 3, no. 2, pp. 103–110, 2018, [Online]. Available: <http://jurnal.usbypkp.ac.id/index.php/infotronik/article/view/109>.
- [21] J. Mase, M. T. Furqon, and B. Rahayudi, “Penerapan Algoritme Support Vector Machine (SVM) Pada Pengklasifikasian Penyakit Kucing,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3648–3654, 2018.
- [22] S. Aprilla, M. T. Furqon, and M. A. Fauzi, “Klasifikasi Penyakit Skizofrenia dan Episode Depresi Pada Gangguan Kejiwaan Dengan Menggunakan Metode Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5611–5618, 2018.
- [23] O. Arifin and T. B. Sasongko, “Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier untuk klasifikasi jalur minat SMA,” *Semin. Nas. Teknol. Inf. dan Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018, [Online]. Available: <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/2059>.
- [24] H. Amalia, “PERBANDINGAN METODE DATA MINING SVM DAN NN UNTUK KLASIFIKASI PENYAKIT GINJAL KRONIS,” *J. Pilar Nusa Mandiri*, vol. 14, no. 1, pp. 1–6, 2018, doi: <https://doi.org/10.33480/pilar.v14i1.80>.
- [25] A. Bode, “Perbandingan Metode Prediksi Support Vector Machine Dan Linear Regression Menggunakan Backward Elimination Pada Produksi Minyak Kelapa,” *J. Sist. Inf. dan Tek. Komput.*, vol. 4, no. 2, pp. 104–107, 2019, [Online]. Available: <http://ejournal.catursakti.ac.id/index.php/simtek/article/view/57>.
- [26] A. Byna and F. N. Anisa, “Backward Elimination Untuk Meningkatkan Akurasi Kejadian Stunting Dengan Analisis Algoritma Support Vector Machine,” *Din. Kesehat.*, vol. 9, no. 2, pp. 217–225, 2018, doi: <https://doi.org/10.33859/dksm.v9i2>.
- [27] A. Bode, “K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward <http://sistemasi.ftik.unisi.ac.id>

- Elimination Untuk Prediksi Harga Komoditi Kopi Arabika,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017, doi: 10.33096/ilkom.v9i2.139.188-195.
- [28] A. D. Ghani, N. Salman, and Mustikasari, “Algoritma k-Nearest Neighbor Berbasis Backward Elimination Pada Client Telemarketing,” *Pros. Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. 8, no. 2, pp. 141–150, 2019.
- [29] R. Sulaehani, “Prediksi Keputusan Klien Telemarketing Untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination,” *Ilk. J. Ilm.*, vol. 8, no. 3, pp. 182–189, 2016, doi: 10.33096/ilkom.v8i3.83.182-189.
- [30] M. Ary and D. A. F. Rismiati, “Ukuran Akurasi Klasifikasi Penyakit Mesothelioma Menggunakan Algoritma K-Nearest Neighbor dan Backward Elimination,” *SATIN – Sains dan Teknol. Inf.*, vol. 5, no. 1, pp. 11–18, 2019.
- [31] B. A. Farahdiba and Y. S. Nugroho, “Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio,” *J. Tek. Elektro*, vol. 8, no. 2, pp. 43–46, 2016.
- [32] N. A. Madyaningrum and Sulastri, “Analisa Prediksi Kekambuhan Kanker Payudara Dengan Menggunakan Algoritma K-Nearest Neighbor,” *Proceeding SINTAK*, vol. 3, no. 3, pp. 180–185, 2019.
- [33] C. W. Dawson, *Projects in Computing and Information Systems*, 2nd ed., vol. 2. London: Pearson Education Limited, 2009.
- [34] UCI Repository, “UCI Machine Learning Repository: About,” *UCI Repository*. archives.ics.uci.edu/ml/about.html (accessed Apr. 18, 2020).
- [35] Yunita, “Seleksi Fitur Menggunakan Backward Elimination Pada Prediksi Cuaca Dengan Neural Network,” *Indones. J. Comput. Inf. Technol.*, vol. 2, no. 1, pp. 26–37, 2017.