

**IMPLEMENTASI METODE SUPPORT VECTOR MACHINE  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT KANKER  
PAYUDARA**

**SKRIPSI**

**OLEH  
SITI HURRIYATI  
NIM. 16610041**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**IMPLEMENTASI METODE SUPPORT VECTOR MACHINE  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT KANKER  
PAYUDARA**

**SKRIPSI**

**Diajukan kepada  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh**

**Siti Hurriyati  
NIM. 16610041**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2023**

**IMPLEMENTASI METODE SUPPORT VECTOR MACHINE  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT KANKER  
PAYUDARA**

**SKRIPSI**

**Oleh  
Siti Hurriyati  
NIM. 16610041**

Telah Diperiksa dan Disetujui untuk Diuji  
Malang, 26 Juni 2023

Dosen Pembimbing I



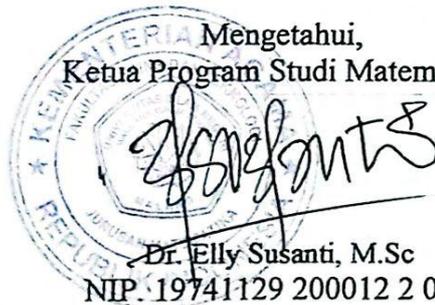
Ria Dhea Layla Nur Karisma, M.Si  
NIDT.19900709 20180201 1 228

Dosen Pembimbing II



M. Nafie Jauhari, M.Si  
NIDT.19870218 20160801 1 056

Mengetahui,  
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005

**IMPLEMENTASI METODE SUPPORT VECTOR MACHINE  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT KANKER  
PAYUDARA**

**SKRIPSI**

Oleh  
**Siti Hurriyati**  
NIM. 16610041

Telah Dipertahankan di Depan Dosen Penguji Skripsi  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S.Mat)  
Tanggal 30 Juni 2023

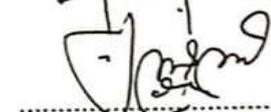
Ketua penguji : Angga Dwi Mulyanto, M.Si



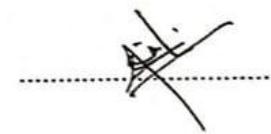
Anggota Penguji 1 : Dr. Elly Susanti, M.Sc



Anggota Penguji 2 : Ria Dhea Layla Nur Karisma, M.si



Anggota Penguji 3 : Mohammad Nafie Jauhari, M.Si



Mengetahui,  
Ketua Program Studi Matematika



  
Dr. Elly Susanti, M.Sc  
NIP. 19741129 200012 2 005

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Siti Hurriyati

NIM : 16610041

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Implementasi Metode *Support Vector Machine* pada Klasifikasi  
Diagnosis Penyakit Kanker Payudara

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 30 Juni 2023

Yang membuat pernyataan,



Siti Hurriyati  
NIM. 16610041

## **MOTO**

*“Karena sesungguhnya sesudah kesulitan itu ada kemudahan.”*

*(Q.S. Al-Insyirah: 5)*

*“My life ain't no fun, but it's still worth living and trying”*

*(Kim Namjoon)*

## **PERSEMBAHAN**

Skripsi ini penulis persembahkan untuk:

Ayahanda H. Fadlullah dan Ibunda Subhiah, yang tiada hentinya mendoakan dan menyemangati penulis dalam menyelesaikan skripsi ini.

Kepada dosen pembimbing penulis, Bu Ria Dhea Layla Nur Karisma, M.Si dan Bapak M. Nafie Jauhari, M.Si yang terus memberikan dorongan dan motivasi agar penulis dapat menyelesaikan skripsi.

Kepada sahabat penulis, Rini, yang bersedia menjadi tempat berbagi suka dan duka dalam berbagai proses hidup yang penulis jalani selama ini.

## KATA PENGANTAR

*Assalamualaikum Warahmatullahi Wabarakatuh.*

Segala puji bagi Allah SWT. yang telah memberikan nikmat sehat dan kesempatan sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi Metode *Support Vector Machine* pada Klasifikasi Diagnosis Penyakit Kanker Payudara” sebagai salah satu syarat untuk memperoleh gelar sarjana dalam bidang matematika di Fakultas Sains dan Teknologi, Universitas Islam Maulana Malik Ibrahim Malang. Shalawat serta salam selalu tercurah kepada Nabi Muhammad SAW. yang telah membawa umatnya dari jalan kegelapan menuju jalan yang terang benderang.

Ucapan syukur dan terima kasih penulis sampaikan kepada semua pihak yang telah membantu, membimbing, dan memberikan bantuan, baik secara langsung maupun tidak langsung. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Prof. Dr. M. Zainuddin, MA., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Harini, M. Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Elly Susanti, M.Sc., selaku ketua Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Ria Dhea Layla Nur Karisma, M.Si., selaku dosen pembimbing I yang telah memberikan ilmu, arahan, dan nasihat dalam menyelesaikan skripsi ini.
5. Mohammad Nafie Jauhari, M.Si., selaku dosen pembimbing II yang telah memberikan ilmu, arahan, dan nasihat dalam penulisan skripsi ini.
6. Dr. Sri Harini, M. Si., selaku ketua penguji dalam ujian skripsi yang telah memberikan arahan serta solusi dalam menyelesaikan skripsi ini.
7. Angga Dwi Mulyanto, M.Si., selaku anggota penguji dalam ujian skripsi yang telah memberikan masukan dan saran dalam skripsi ini.
8. Seluruh dosen dan staf Universitas Islam Negeri Maulana Malik Ibrahim Malang.

9. Ayahanda H. Fadlullah dan Ibunda Subhiah selaku orang tua penulis serta seluruh keluarga dan sahabat yang senantiasa mendoakan dan memberikan motivasi bagi penulis.
10. Teman-teman mahasiswa Matematika angkatan 2016 yang telah bersama-sama berjuang dalam perkuliahan dan memberikan motivasi kepada penulis.

Semoga Allah SWT. senantiasa memberikan berkah kepada kita semua.

Penulis berharap skripsi ini dapat bermanfaat dan menambah wawasan bagi penulis dan pembaca.

*Wassalamualaikum Warahmatullahi Wabarakatuh.*

Malang, 30 Juni 2023

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGAJUAN .....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>PERNYATAAN KEASLIAN TULISAN .....</b>	<b>v</b>
<b>MOTO .....</b>	<b>vi</b>
<b>PERSEMBAHAN.....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR SIMBOL .....</b>	<b>xiv</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xv</b>
<b>ABSTRAK .....</b>	<b>xvi</b>
<b>ABSTRACT .....</b>	<b>xvii</b>
<b>مستخلص البحث.....</b>	<b>xvii</b>
<b>.....</b>	<b>xvii</b>
<b>i</b>	
<b>BAB IPENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	5
1.5 Batasan Masalah .....	5
1.6 Definisi Istilah .....	5
<b>KAJIAN PUSTAKA .....</b>	<b>7</b>
2.1 <i>Support Vector Machine</i> .....	7
2.1.1 <i>Linear Separable Data</i> .....	8
2.1.2 <i>Linear Non Separable Data</i> .....	10
2.2 <i>Non Linear Support Vector Machine</i> .....	12
2.3 Ketepatan Model Klasifikasi .....	14
2.4 Kanker Payudara.....	15
2.5 Diagnosis Kanker Payudara.....	16
2.6 <i>Wisconsin Diagnostic Breast Cancer</i> .....	17
2.7 Kajian Agama .....	18
<b>BAB III METODE PENELITIAN .....</b>	<b>21</b>
3.1 Jenis Penelitian .....	21
3.2 Data dan Sumber Data .....	21
3.3 Variabel Penelitian.....	21
3.4 Teknik Analisis Data .....	22
<b>BAB IV PEMBAHASAN.....</b>	<b>25</b>
4.1 Analisis Deskriptif .....	25
4.2 Implementasi Support Vector Machine .....	28
4.2.1 <i>Split Data</i> .....	28
4.2.2 Klasifikasi Diagnosis Kanker Payudara dengan SVM.....	28
4.2.3 Akurasi Model Klasifikasi Kanker Payudara .....	31

4.3 Kajian Agama Terkait Penelitian.....	32
<b>BAB V PENUTUP .....</b>	<b>34</b>
5.1 Kesimpulan .....	34
5.2 Saran .....	34
<b>DAFTAR PUSTAKA .....</b>	<b>35</b>
<b>LAMPIRAN</b>	
<b>RIWAYAT HIDUP</b>	

## DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i> .....	15
Tabel 3. 1 Variabel Penelitian .....	21
Tabel 4.1 Analisis Deskriptif Variabel Penelitian.....	26
Tabel 4. 2 Model SVM dengan $C = 0,01$ .....	29
Tabel 4. 3 Model SVM dengan $C = 0,1$ .....	29
Tabel 4. 4 Model SVM dengan $C = 1$ .....	29
Tabel 4. 5 Model SVM dengan $C = 10$ .....	29
Tabel 4. 6 Model SVM dengan $C = 100$ .....	29
Tabel 4. 7 <i>Confusion Matrix</i> pada Diagnosis Kanker Payudara .....	31

## DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>Support Vector Machine</i> .....	8
Gambar 2.2 Pemetaan Data <i>Input</i> ke <i>Fitur Space</i> .....	12
Gambar 3.1 <i>Flowchart Penelitian</i> .....	24
Gambar 4.1 <i>Pie Chart</i> Diagnosis Kanker Payudara.....	25
Gambar 4. 2 Plot SVM dengan $C = 10$ .....	30

## DAFTAR SIMBOL

$b$	:	Bias
$w$	:	Vektor pembobot
$x_i$	:	<i>Dataset</i> ke- $i$
$y_i$	:	Label ke- $i$
$K(x_i, x_j)$	:	Fungsi Kernel
$\alpha_i$	:	Pengali <i>lagrange</i>
$\xi_i$	:	Variabel <i>slack</i>

## DAFTAR LAMPIRAN

Lampiran 1 Data <i>Wisconsin Breast Cancer Diagnostic</i> .....	37
Lampiran 2 Plot SVM .....	39
Lampiran 3 Script SVM di R Studio.....	43

## ABSTRAK

Hurriyati, Siti. 2023. **Implementasi *Support Vector Machine* pada Klasifikasi Diagnosis Penyakit Kanker Payudara.** Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Mohammad Nafie Jauhari, M.Si.

**Kata Kunci:** Kanker payudara, *Support Vector Machine*, *Wisconsin Breast Cancer Diagnostic, benign, malignant*

*Support Vector Machine* merupakan salah satu jenis *machine learning* yang banyak digunakan saat ini. Penerapan metode ini mencakup berbagai bidang, salah satunya bidang medis. Pada bidang medis, *Support Vector Machine* diterapkan pada klasifikasi diagnosis suatu penyakit. Hasil dari metode ini diharapkan dapat memudahkan pihak-pihak terkait dalam proses penanganan pasien sejak dini. Penelitian ini berfokus pada penyakit kanker payudara. Hasil penelitian menunjukkan bahwa model SVM terbaik ketika nilai  $C = 10$ . Variabel radius ( $X1$ ) dan *perimeter* ( $X3$ ) memiliki pengaruh yang besar pada diagnosis kanker payudara kategori *benign*, sedangkan variabel *texture* dan *smoothness* memiliki pengaruh besar terhadap diagnosis kanker payudara kategori *malignant*. Tingkat akurasi metode *Support Vector Machine* pada klasifikasi diagnosis kanker payudara sebesar 96,49%. Hal ini menunjukkan bahwa metode *Support Vector Machine* bekerja dengan baik pada diagnosis kanker payudara.

## ABSTRACT

Hurriyati, Siti. 2023. **On the Implementation of Support Vector Machine Method in Classification of Breast Cancer Diagnosis.** Thesis. Departement of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Malang. Advisors: (I) Ria Dhea Layla Nur Karisma, M.Si. (II) Mohammad Nafie Jauhari, M.Si.

**Keywords:** Breast cancer, Support Vector Machine, Wisconsin Breast Cancer Diagnostic, benign, malignant

Support Vector Machine is one of the type of machine learning that widely used recently. The application of this method covers various fields, medical field is one of them. In medical field, Support Vector Machine (SVM) method is applied to the disease diagnosis classification. The results of this method are expected to facilitate related parties in the process of treating patients from an early age. This research focuses on breast cancer. The results show that the SVM model is best when the value is  $C = 10$ . The radius ( $X1$ ) and perimeter ( $X3$ ) variables have a large influence on the diagnosis of benign breast cancer, while the texture and smoothness variables have a large influence on the diagnosis of malignant breast cancer. The accuracy rate of the SVM method in the classification of breast cancer diagnosis is 96.49%. This shows that the SVM method works well in the diagnosis of breast cancer.

## مستخلص البحث

حريتي، ستي. ٢٠٢٣. تنفيذ طريقة آلة الناقلات الدعم في تصنيف سرطان الثدي. البحث الجامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية بمالانج. المشرفة الاولى: (١) ريا ديا ليلي نور كرسما، الماجستير. (٢) محمد نافع جوهرى، الماجستير.

**الكلمات المفتاحية:** سرطان الثدي، آلة الناقلات الدعم (*Support Vector Machine*)، تشخيص سرطان الثدي في ويسكونسن، حميدة، خبيث

تعد آلة الناقلات الدعم من أنواع التعلم الآلة المستخدمة للنطاق الواسع اليوم. يغطي تطبيق هذه الطريقة في المجالات المختلفة، أحدها هو المجال الطبي. في هذا، يتم تطبيق دعم شاحنات النقل لتصنيف تشخيص المرض. ومن المتوقع أن تسهل النتائج من هذه الطريقة الأطراف ذات العلاقة في عملية علاج المرضى منذ الآن. يركز هذا البحث على سرطان الثدي. تظهر النتائج أن نموذج *SVM* هو الأفضل عندما تكون القيمة  $C = 10$  متغيرات نصف القطر ( $X1$ ) والمحيط ( $X3$ ) لها تأثير كبير على تشخيص سرطان الثدي الحميد، في حين أن متغيرات النسيج والنعومة لها تأثير كبير على تشخيص سرطان الثدي الخبيث. معدل دقة طريقة جهاز المتجهات الداعمة في تشخيص سرطان الثدي هو 96.49%. هذا يدل على أن طريقة دعم شاحنات النقل تعمل بشكل جيد في تشخيص سرطان الثدي.

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Seiring dengan perkembangan informasi tentang *machine learning*, penggunaan *machine learning* ikut mengalami peningkatan. *Machine learning* memberi kemudahan pada pengguna untuk dapat menyelesaikan berbagai permasalahan terkait dengan analisis data. Data dengan jumlah kecil sampai dengan *big data* menggunakan *machine learning* dalam mengolah dan menganalisis data yang meliputi masalah klasifikasi, regresi, sampai dengan prediksi.

Salah satu metode *machine learning* yang banyak diaplikasikan pada masalah klasifikasi adalah *Support Vector Machine*. Inti dari metode ini adalah mencari *hyperplane* (pemisah) yang membagi data menjadi dua kelas yang berbeda. *Support Vector Machine* sering digunakan dalam berbagai masalah klasifikasi karena mampu bekerja pada data berdimensi tinggi dan menjadi nilai lebih untuk metode ini.

Metode *Support Vector Machine* (SVM) telah digunakan pada berbagai penelitian dan bidang, diantaranya pengidentifikasian jenis kayu menggunakan metode berbasis data citra oleh Gunawan, dkk (2014). Hasil penelitian tersebut menunjukkan akurasi tingkat klasifikasi sebesar 95,83% pada kernel polinomial. Di bidang ekonomi, Kasim, dkk (2019) menggunakan metode ini untuk mengklasifikasi penduduk yang berhak menerima bantuan pemerintah agar tepat sasaran. Hasil klasifikasi menunjukkan tingkat akurasi model SVM sebesar 98%. Adapun implementasi metode *Support Vector Machine* di bidang pendidikan

salah satunya dalam penentuan akreditasi sekolah. Penelitian Octaviani, dkk (2014) menunjukkan bahwa ketepatan klasifikasi SVM menggunakan kernel *Radial Basis Function* (RBF) sebesar 93,902%.

Implementasi metode *Support Vector Machine* juga diterapkan pada bidang medis. Salah satunya pada penentuan klasifikasi diagnosis kanker payudara. Kanker payudara merupakan salah satu penyakit yang banyak diderita di Indonesia maupun di dunia. Hal ini dibuktikan dengan kenaikan jumlah penderita kanker payudara tiap tahunnya yang sebagian besar dialami oleh wanita. *World Health Organization* (WHO) menyatakan bahwa di tahun 2012 kanker payudara menempati urutan kedua penyebab kematian terbesar wanita setelah kanker leher rahim (Permadi, dkk, 2016).

Peningkatan pasien penderita kanker payudara menjadi salah satu peringatan, khususnya di kalangan perempuan agar lebih memperhatikan kesehatan masing-masing. Tidak jarang pasien telah dideteksi mengidap kanker payudara stadium akhir sehingga tidak dapat tertolong karena keterlambatan pemeriksaan dan kurangnya informasi terkait gejala, pemeriksaan, dan pengobatan kanker payudara. Oleh karena itu, perlu dilakukan pemeriksaan sejak awal untuk mengetahui apakah seseorang mengidap kanker payudara atau tidak.

Penyebaran sel kanker dapat dapat diatasi dan dicegah dengan melakukan pemeriksaan agar memperoleh diagnosis dini. Beberapa teknik diagnosis yang dilakukan antara lain anamnesis dan pemeriksaan fisik (klinis), radiologi (*imaging*), dan histopatologi (Ketut & Kartika, 2022). Penelitian ini berfokus pada teknik histopatologi dengan pemeriksaan melalui *Fine Needle Aspiration* (FNA) *Biopsy*. FNA *Biopsy* dilakukan dengan mengambil sebagian kecil jaringan tumor

kemudian diperiksa menggunakan mikroskop.

Allah SWT. berfirman dalam QS. Al-Baqarah ayat 233 yang artinya:

*“Para ibu menyusukan anaknya selama 2 tahun penuh, yaitu bagi yang ingin menyempurnakan penyusuan. Dan menjadi kewajiban atas bayi itu yang dilahirkan untuknya (ayah sang bayi) memberi rezeki (makanan) dan pakaian kepada para ibu dengan cara yang ma’ruf. Seseorang tidak dibebani melainkan menurut kadar kesanggupannya. Tidaklah seorang ibu menderita kesengsaraan karena anaknya dan juga seorang ayah karena anaknya, dan waris pun berkewajiban demikian. Apabila keduanya ingin menyapah berdasarkan kerelaan keduanya dan permusyawaratan, maka tidak ada dosa atas keduanya. Dan jika kamu ingin anak kamu disusukan oleh orang lain, maka tidak ada dosa bagi kamu apabila kamu memberikan pembayaran menurut yang patut. Bertakwalah kepada Allah dan ketahuilah bahwa Allah Maha Melihat apa yang kamu kerjakan.” (Q.S. Al-Baqarah ayat 233).*

Shihab (2002) dalam tafsir Al Mishbah menjelaskan bahwa ayat ini memerintahkan kepada para ibu agar menyusui anak-anaknya. Ayat ini menggunakan kata “*walidat*” bukan “*ummahaat*” yang berarti air susu ibu, baik yang berasal dari ibu kandung maupun bukan merupakan makanan terbaik bagi bayi hingga ia berusia 2 tahun.

Ayat tersebut memiliki keterkaitan dalam penelitian ini. Penelitian ini berfokus kepada pengklasifikasian kanker payudara, yang sebagian besar dialami oleh wanita. Sudah menjadi hal umum bahwa peran wanita ketika menjadi seorang ibu adalah menyusui. ASI sangat dibutuhkan bagi tumbuh kembang bayi. Jika terdapat masalah kesehatan pada ibu, maka hal tersebut akan berpengaruh pada produksi ASI. Akibatnya, proses menyusui yang dianjurkan selama 2 tahun dapat terganggu.

Adapun untuk klasifikasi diagnosis kanker payudara, penelitian-penelitian sebelumnya menggunakan metode *Radial Basis Probabilistic Neural Network* (RBPNN) oleh Shahura, dkk (2016), *K-Nearest Neighbor* (KNN) dan *Naive Bayes*

*Classifier* oleh Amran, dkk (2018).

Berdasarkan penelitian-penelitian tersebut, *Support Vector Machine* pada umumnya digunakan untuk mengklasifikasikan berbagai jenis data, baik itu bidang medis maupun non medis. Oleh karena itu, peneliti tertarik menggunakan metode *Support Vector Machine* dalam menentukan diagnosis kanker payudara. Proses pengklasifikasian ini akan memudahkan penanganan yang tepat bagi penderita kanker payudara di masa yang akan datang. Data yang digunakan adalah data *Wisconsin Breast Cancer*. Data inilah yang digunakan dalam proses pengklasifikasian dengan metode *Support Vector Machine*.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang tersebut, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana hasil klasifikasi *Support Vector Machine* pada diagnosis kanker payudara?
2. Bagaimana akurasi model *Support Vector Machine* pada klasifikasi diagnosis kanker payudara?

## **1.3 Tujuan Penelitian**

Tujuan dilakukan penelitian ini sebagai berikut:

1. Mendapatkan hasil klasifikasi *Support Vector Machine* diagnosis kanker payudara.
2. Menganalisis akurasi model *Support Vector Machine* pada klasifikasi diagnosis kanker payudara.

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat antara lain:

### 1. Bagi peneliti

Meningkatkan pengetahuan dan wawasan terutama tentang *Machine Learning* di bidang medis serta proses pengklasifikasian data dengan *Support Vector Machine*.

### 2. Bagi Instansi Terkait

Sebagai tambahan referensi tentang prosedur diagnosis penyakit kanker payudara dengan waktu yang lebih singkat sehingga dapat memaksimalkan waktu yang ada.

### 3. Bagi Pembaca

Menjadi tambahan referensi bacaan tentang *Support Vector Machine* dan pengaplikasiannya pada penyakit kanker payudara.

## 1.5 Batasan Masalah

Penelitian ini terbatas pada beberapa aspek sebagai berikut:

1. Data yang digunakan adalah *dataset* diagnosis kanker payudara *Wisconsin*.
2. Metode yang digunakan adalah *Support Vector Machine*.
3. Kernel yang digunakan dalam penelitian ini adalah kernel linear.

## 1.6 Definisi Istilah

Beberapa istilah yang digunakan dalam penelitian ini antara lain:

<i>Hyperplane</i>	:	Batas pemisah antar kelas
<i>Margin</i>	:	Jarak terdekat antara <i>Support Vector</i> dengan masing-masing kelas

<i>Support Vector</i>	:	Data yang berada pada <i>hyperplane</i>
<i>Data Training</i>	:	Data yang digunakan untuk membangun model klasifikasi
<i>Data Testing</i>	:	Data yang digunakan untuk mengevaluasi model klasifikasi
TP	:	Prediksi benar dikelompokkan ke kelas positif
TN	:	Prediksi benar dikelompokkan ke kelas negatif
FP	:	Prediksi salah dikelompokkan ke kelas positif
FN	:	Prediksi salah dikelompokkan ke kelas negatif
<i>Benign</i>	:	Tumor jinak
<i>Malignant</i>	:	Tumor ganas (kanker)
<i>Dataset Breast Cancer Wisconsin</i>	:	Data yang digunakan untuk menentukan diagnosis kanker payudara melalui pemeriksaan FNAB
Kanker Payudara	:	Tumor ganas yang tumbuh di sekitar jaringan payudara akibat pertumbuhan sel abnormal dan tak terkendali

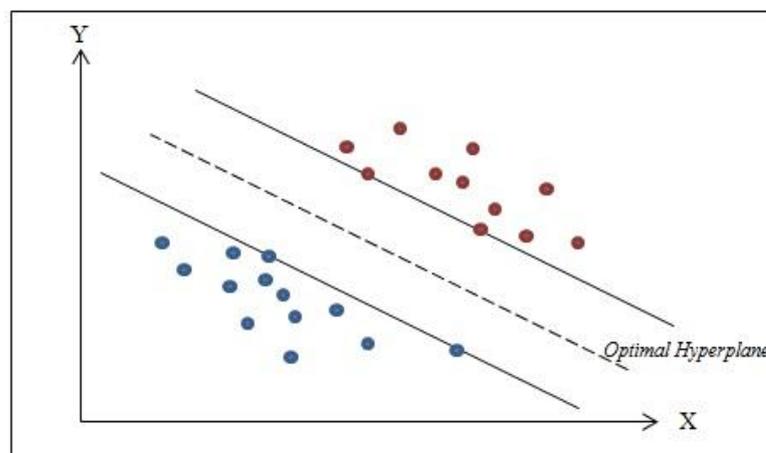
## **BAB II KAJIAN PUSTAKA**

### **2.1 *Support Vector Machine***

*Support Vector Machine* (SVM) merupakan salah satu metode klasifikasi yang didasarkan pada *Structural Risk Minimization*, yang merupakan prinsip induktif dalam *machine learning* (Indriani, dkk, 2017). *Support Vector Machine* (SVM) dikembangkan pada awal 1990an sebagai suatu solusi non linear untuk masalah klasifikasi dan regresi (Vapnik, 1998). *Support Vector Machine* (SVM) merupakan salah satu metode yang relatif baru dibandingkan dengan *machine learning* yang lain, akan tetapi memiliki kinerja yang cukup baik dalam berbagai bidang, seperti bioinformatika, klasifikasi teks dan dokumen, pengenalan tulisan tangan, dan lain sebagainya. Ciri dari metode ini adalah menemukan fungsi pemisah optimal yang dapat memisahkan dua data set data dari dua kelas yang berbeda. *Support Vector Machine* akan membangun suatu *hyperplane* (bidang pemisah) yang memisahkan data terhadap kelasnya. Bidang pemisah yang baik adalah bidang pemisah dengan batasan (*margin*) yang paling besar. Proses pencarian bidang pemisah terbaik inilah yang menjadi inti dari *Support Vector Machine* (Munawarah, dkk, 2016).

Kesuksesan metode *Support Vector Machine* (SVM) dilatarbelakangi oleh beberapa alasan, diantaranya *Support Vector Machine* (SVM) memiliki kemampuan pembelajaran yang baik walaupun dengan jumlah fitur yang sangat kecil. Selain itu, metode ini juga mampu mengatasi model yang *error*, serta memiliki efisiensi komputasi dibanding *machine learning* yang lain. *Support Vector Machine* dapat diterapkan pada data non linear dengan menggunakan

metode *kernel trick* yang mencari bidang pemisah dengan mentransformasikan data ke ruang vektor dengan dimensi yang lebih tinggi (*feature space*) sehingga data dapat dipisah secara linear pada *feature space* tersebut. Secara umum, *Support Vector Machine* (SVM) dibagi menjadi dua, yaitu *Support Vector Classification* (SVC) dan *Support Vector Regression* (SVR) (Gholami & Fakhari, 2017).



Gambar 2.1 Ilustrasi Support Vector Machine

### 2.1.1 Linear Separable Data

Tujuan *Support Vector Machine* adalah untuk menemukan *hyperplane* yang dapat memisahkan dua kelas dengan kriteria tertentu. Untuk data dengan klasifikasi dua kelas, dapat diperoleh *hyperplane* berupa garis lurus dengan margin dari masing-masing kelas klasifikasi. Misal  $x_i$  merupakan data *training* dengan  $i = 1, 2, 3, \dots, n$ , maka dapat didefinisikan sebagai (Gholami & Fakhari, 2017):

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (2.1)$$

dimana  $w$  merupakan vektor pembobot dan  $b$  merupakan bias. *Hyperplane* ini memiliki dua syarat yaitu:

1. Harus memiliki *error* terkecil dalam pengklasifikasian data
2. Memiliki jarak terdekat dari masing-masing kelas (margin) yang maksimum.

Oleh karena itu, data untuk masing-masing kelas memiliki nilai  $y = 1$  untuk kelas positif dan  $y = -1$  untuk kelas negatif, sehingga margin tersebut dapat dituliskan melalui persamaan berikut:

$$\mathbf{w}^T \mathbf{x} + \mathbf{b} \begin{cases} \geq \mathbf{1} & \text{untuk } y_i = \mathbf{1} \\ \leq -\mathbf{1} & \text{untuk } y_i = -\mathbf{1} \end{cases} \quad (2.2)$$

Untuk menemukan *hyperplane* terbaik jarak ( $d$ ) antar margin ditentukan menggunakan persamaan berikut:

$$d(\mathbf{w}, \mathbf{b}; \mathbf{x}) = \frac{|(\mathbf{w}^T \mathbf{x} + \mathbf{b} - \mathbf{1}) - (\mathbf{w}^T \mathbf{x} + \mathbf{b} + \mathbf{1})|}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (2.3)$$

Maka dari itu, memaksimalkan margin setara dengan meminimalkan dimensi vektor  $w$ . Penentuan *hyperplane* dapat ditulis dengan persamaan berikut:

$$\text{Min}_{\mathbf{w}, \mathbf{b}} = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (2.4)$$

dengan  $\mathbf{y}_i(\mathbf{w}^T \mathbf{x} + \mathbf{b}) \geq \mathbf{1}$

Pengali *lagrange* dapat digunakan pada situasi ini dengan menegakkan kendala sebagai berikut:

$$L(\mathbf{w}, \mathbf{b}, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i (\mathbf{y}_i(\mathbf{w}^T \mathbf{x} + \mathbf{b}) - \mathbf{1}) \quad (2.5)$$

Perlu diketahui bahwa  $\alpha_i \neq 0$  jika dan hanya jika  $x_i$  merupakan *support vector*. *Support vector* merupakan data yang dipilih sebagai batasan dari tiap kelas. Optimasi dilakukan dengan melakukan turunan parsial  $L$  terhadap  $w$  dan  $b$  untuk memenuhi kondisi Karush-Kuhn-Tucker (Smola & Scholkopf, 2002):

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{0}$$

sehingga:

$$\mathbf{w}_0 = \sum_{i=1}^N \alpha_i \mathbf{x}_i \mathbf{y}_i \quad (2.6)$$

$$\frac{\partial L}{\partial b} = 0$$

sehingga:

$$\sum_{i=1}^N \alpha_i \mathbf{y}_i = \mathbf{0} \quad (2.7)$$

Substitusi persamaan (2.6) dan (2.7) ke persamaan (2.5) sehingga diperoleh:

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{j=1}^N \mathbf{y}_j \mathbf{y}_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.8)$$

dengan kendala:

$$\alpha_i \geq 0, \quad \sum_{i=1}^N \alpha_i \mathbf{y}_i = \mathbf{0}$$

Persamaan tersebut digunakan *support vector* pada data yang diinput. Parameter  $w$  ditentukan menggunakan persamaan (2.6). Sedangkan parameter bias  $b$  dapat dihitung dengan persamaan berikut (Gholami & Fakhari, 2017):

$$\mathbf{b}_0 = \frac{1}{N} \sum_{i=1}^l (\mathbf{y}_i - \mathbf{w}^T \mathbf{x}_i) \quad (2.9)$$

### 2.1.2 Linear Non Separable Data

Secara umum selalu terdapat kondisi dimana data tidak terpisah secara

sempurna oleh *hyperplane* karena kemiripan beberapa fitur dari *dataset*. Pada kondisi ini digunakan variabel *slack* dengan  $\xi_i \geq 0$ ,  $i = 1, 2, \dots, n$  (Smola & Scholkopf, 2002). Oleh karena itu, optimasi fungsi yang ada pada persamaan (2.4) dapat dituliskan kembali menjadi (Gholami & Fakhari, 2017):

$$\text{Min}_{\mathbf{w}, b} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (2.10)$$

dengan 
$$\mathbf{y}_i(\mathbf{w}^T \mathbf{x} + b) \geq 1 - \xi_i$$

Parameter  $C$  pada persamaan (2.10) memaksimalkan margin dan meminimalisasi *error* pada klasifikasi. Untuk menyelesaikan masalah optimasi, maka digunakan pengali *lagrange*  $\alpha$  dan  $\beta$ , dengan  $\alpha_i, \beta_i \geq 0$  sehingga:

$$\begin{aligned} L(\mathbf{w}, b, \xi, \alpha, \beta) = & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ & - \sum_{i=1}^N \alpha_i (\mathbf{y}_i(\mathbf{w}^T \mathbf{x} + b) - 1 + \xi_i) - \sum_{i=1}^N \beta_i \xi_i \end{aligned} \quad (2.11)$$

Solusi optimal untuk persamaan (2.11) ditentukan dengan melakukan turunan parsial  $L$  terhadap  $w$ ,  $b$ , dan  $\xi$  yang memenuhi kondisi Karush-Kuhn-Tucker (Smola & Scholkopf, 2002):

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{0} \quad (2.12)$$

sehingga:

$$\begin{aligned} \mathbf{w}_0 = & \sum_{i=1}^N \alpha_i x_i y_i \\ & \frac{\partial L}{\partial b} = \mathbf{0} \end{aligned}$$

sehingga: (2.13)

$$\sum_{i=1}^N \alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi} = 0$$

sehingga:

$$\alpha_i + \beta_i = C \quad (2.14)$$

Substitusi persamaan (2.12), (2.13), dan (2.14) ke persamaan (2.11)

sehingga diperoleh:

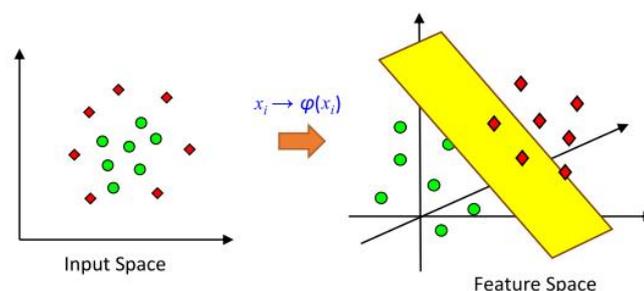
$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{j=1}^N y_j \alpha_j \alpha_i x_i^T x_j \quad (2.15)$$

Dengan kendala:

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^N \alpha_i y_i = 0$$

## 2.2 Non Linear Support Vector Machine

Pada awalnya, teknik *machine learning* terbatas pada asumsi kelinearan. Akibatnya, hasil algoritma-algoritma tersebut terbatas pada kasus-kasus linear. Permasalahan ini dapat diatasi pada *Support Vector Machine* dengan menggunakan fungsi kernel. Pada *Support Vector Machine non linear*, data *input* dipetakan ke dimensi yang lebih tinggi sehingga dapat dipisah secara linear pada ruang fitur (*fitur space*) (Muis & Affandes, 2015).



Gambar 2. 2 Pemetaan Data *Input* ke *Fitur Space*

Berdasarkan gambar tersebut, data *input*  $x$  direpresentasikan sebagai  $\varphi(x)$  pada *fitur space*, sedangkan bentuk fungsi dari pemetaan ini tidak diketahui. Maka dari itu, fungsi kernel didefinisikan sebagai berikut (Gholami & Fakhari, 2017):

$$\varphi(x_i, x_j) = K(x_i, x_j) \quad (2.16)$$

Berdasarkan persamaan (2.15), persamaan umum untuk kasus *non linear support vector* dapat ditulis sebagai berikut:

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{j=1}^N y_i y_j \alpha_i \alpha_j K(x_i, x_j) \quad (2.17)$$

Dengan kendala:

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^N \alpha_i y_i = 0$$

Vektor  $w$  dalam *Hilbert pace* dapat ditulis sebagai berikut:

$$w = \sum_{i=1}^N y_i \alpha_i \varphi(x_i) \quad (2.18)$$

Penggunaan *kernel trick* menyebabkan penentuan vektor pembobot tidak lagi dibutuhkan dan parameter bias  $b$  dapat dihitung menggunakan persamaan sebagai berikut:

$$b = y_i - \sum_{i,j=1}^l \alpha_i y_i K(x_i, x_j) \quad (2.19)$$

Perlu diketahui bahwa *hyperplane* didefinisikan sebagai:

$$w^T \varphi(x) + b \quad (2.20)$$

Dengan substitusi persamaan (2.18) ke persamaan (2.20), maka diperoleh

persamaan umum *hyperplane* sebagai berikut:

$$\mathbf{d}(x) = \sum_{i=1}^N y_i \alpha_i K(x, x_i) + \mathbf{b} \quad (2.21)$$

Beberapa bentuk fungsi kernel yang digunakan pada *Support Vector Machine* (Mustakim, dkk, 2015):

- a. Kernel linear

$$K(x_i, x_j) = x_i^T x_j \quad (2.22)$$

- b. Kernel polinomial

$$K(x_i, x_j) = (x_i^T x_j + 1)^p \quad (2.23)$$

- c. Kernel *Radial Basic Function* (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2.24)$$

$$\text{dengan } \gamma = \frac{1}{2\sigma^2}$$

- d. Kernel sigmoid

$$K(x_i, x_j) = \tanh(\alpha(x_i, x_j) + \vartheta) \quad (2.25)$$

Karena penentuan vektor pembobot  $w$  pada *Hilbert space* tidak diperlukan, maka SVM dapat digunakan secara efisien dalam menyelesaikan kasus non linear dengan memilih fungsi kernel yang sesuai.

### 2.3 Ketepatan Model Klasifikasi

Ketepatan model klasifikasi pada *Support Vector Machine* dapat ditentukan menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan alat pengujian untuk mengetahui keakuratan data prediksi terhadap data aktual (Hidayat, dkk, 2021). *Confusion Matrix* terdiri dari empat komponen utama, diantaranya:

1. *True positive* merupakan nilai positif yang diprediksi benar sesuai dengan nilai aktual.
2. *False positive* merupakan nilai negatif akan tetapi diprediksi positif.
3. *True negative* merupakan nilai negatif yang diprediksi sesuai dengan nilai aktual.
4. *False negative* merupakan nilai positif akan tetapi diprediksi negatif.

*Confusion Matrix* dapat dilihat melalui tabel berikut:

**Tabel 2. 1 Confusion Matrix**

Nilai Prediksi	Nilai Aktual	
	Positif	Negatif
Positif	<i>True Positive</i>	<i>False Positive</i>
Negatif	<i>False Negative</i>	<i>True Negative</i>

Adapun untuk akurasi klasifikasi dapat ditentukan sebagai berikut:

$$Akurasi(\%) = \frac{TP + TN}{TP + FN + FP + TN} \quad (2.26)$$

Tingkat *sensitivity* dapat ditentukan dengan rumus:

$$Sensitivity(\%) = \frac{TP}{TP + FN} \quad (2.27)$$

Sedangkan untuk *specificity* dapat ditentukan dengan:

$$Specificity(\%) = \frac{TN}{TN + FP} \quad (2.28)$$

## 2.4 Kanker Payudara

Tumor pada tubuh manusia terbagi menjadi dua, yaitu tumor jinak (*benign*) dan tumor ganas (*malignant*). Tumor ganas inilah yang disebut sebagai kanker. *Benign dan malignant* sama-sama memiliki arti sebagai pertumbuhan jaringan sel yang tidak normal. Hal yang membedakan kedua jenis tumor ini adalah *benign* tidak menyebar ke jaringan tubuh yang lain. Sedangkan *malignant* dapat menyebar ke jaringan tubuh lainnya. *Malignant* (tumor ganas) inilah yang

disebut sebagai kanker (Putra, 2015).

Yayasan Kanker Indonesia (YKI) mendefinisikan kanker sebagai keadaan tak terkendali pada sel yang mengakibatkan pertumbuhan sel tidak normal, berlangsung cepat dan terus-menerus. Putra (2015) menjelaskan kanker payudara merupakan tumor ganas (*malignant*) yang disebabkan oleh pertumbuhan sel tidak normal dan tidak terkontrol pada jaringan payudara yang bersifat terus menerus sehingga terjadi benjolan. Kanker payudara tumbuh pada kelenjar susu, saluran kelenjar, dan jaringan penunjang payudara seperti jaringan lemak, kelenjar ikat, dan jaringan ikat payudara.

Menurut Iqmy, dkk (2021), beberapa faktor yang dapat meningkatkan resiko terjadinya kanker payudara antara lain:

1. Bertambahnya usia. Hal ini disebabkan karena sebagian besar penderita kanker payudara dialami oleh wanita usia menopause.
2. Genetik. Hal ini berkaitan dengan perubahan gen yang berperan dalam pertumbuhan dan perkembangbiakan sel secara terus-menerus tanpa terkendali.
3. Riwayat reproduksi dan hormonal.
4. Konsumsi rokok. Kandungan asap rokok yang beracun menyebabkan peningkatan pertumbuhan sel payudara menjadi tidak terkendali.
5. Konsumsi alkohol. Konsumsi alkohol menyebabkan proses metabolisme di hati menjadi terganggu.

## **2.5 Diagnosis Kanker Payudara**

Pemeriksaan dini perlu dilakukan untuk mencegah terjadinya penyebaran kanker payudara. Terdapat 3 teknik yang dilakukan dalam diagnosis kanker

payudara diantaranya (Ketut & Kartika, 2022):

1. Klinis (anamnesis dan pemeriksaan fisik)

Pemeriksaan ini dilakukan dengan melakukan pemeriksaan fisik dan mengumpulkan informasi mengenai keluhan yang dirasakan, waktu pertama kali benjolan muncul, letak benjolan, riwayat genetik dan keluarga, serta keluhan-keluhan lain yang dialami selama benjolan muncul seperti nyeri pada tulang dan perut, gangguan pencernaan.

2. *Imaging* (radiologi)

Pemeriksaan radiologi yang dianjurkan antara lain mamografi, ultrasonografi (USG), *Bone Tumor*, dan *Magnetic Resonance Imaging* (MRI).

3. Sitologi (histopatologi)

Teknik ini lebih umum dikenal dengan biopsi. Pengambilan sampel dapat dilakukan melalui tiga cara yaitu:

- a. *Fine Needle Aspirate Biopsy* (FNAB), mengambil sebagian kecil jaringan tumor dan diperiksa menggunakan mikroskop.
- b. *Core Biopsy*, mengambil jaringan tumor dengan ukuran yang lebih besar dibanding FNAB dengan spesimen berbentuk silinder.
- c. Biopsi terbuka dilakukan jika hasil FNAB dan *core biopsy* meragukan, sedangkan hasil radiologi menunjukkan kelainan pada jaringan sel.

## **2.6 *Wisconsin Diagnostic Breast Cancer***

*Wisconsin Diagnostic Breast Cancer* merupakan dataset medis pasien kanker payudara yang diperoleh melalui *Fine Needle Aspirate* (FNA) *biopsy*. *Dataset* ini memiliki dua kelas, yaitu diagnosis *benign* (tumor jinak) dan

*malignant* (tumor ganas). *Fine Needle Aspirate* (FNA) *biopsy* dilakukan dengan mengambil sel benjolan dari massa payudara yang diduga sebagai jaringan tumor. Hasil dari *Fine Needle Aspirate* (FNA) *biopsy* ini berupa citra digital massa payudara yang mendeskripsikan karakteristik inti sel dengan atribut sebagai berikut:

1. Radius merupakan rata-rata dari jarak tepi sel ke pusat sel tumor.
2. *Texture* merupakan perbedaan intensitas keabuan pada piksel sel tumor
3. Perimeter merupakan besar keliling inti sel tumor.
4. Area merupakan luas dari sel tumor atau dapat diartikan sebagai jumlah piksel sel tumor.
5. *Compactness* dapat dihitung dengan membagi antara kuadrat dari perimeter dan area. *Compactness* dapat ditulis sebagai  $\frac{perimeter^2}{area}$
6. *Smoothness* diperoleh dari perbedaan panjang masing-masing garis radial dengan rata-rata dua garis radial yang mengelilinginya.
7. *Concavity* merupakan ukuran kelekukan suatu sel.
8. *Concave points* menunjukkan seberapa banyak lekukan yang ada pada sel.
9. *Symmetry* diperoleh dengan menghitung perbedaan panjang pasangan garis tegak lurus terhadap inti sel.
10. *Fractal Dimension* rasio yang memberikan indeks statistik tentang kompleksitas detail dibandingkan dengan pola yang dibentuk pada saat penskalaan.

## 2.7 Kajian Agama

Allah SWT. berfirman dalam surah Asy-Syu'ara' ayat 80 yang artinya:

*“Dan apabila aku sakit, maka hanya Dia yang menyembuhkan aku”*

Menurut Shihab (2002) dalam tafsir Al Mishbah, penyakit berat atau ringan, baik fisik ataupun mental merupakan keniscayaan dalam hidup. Dalam kehidupan, ada yang disebut *sunnatullah*, yaitu ketetapan-ketetapan Tuhan yang berlaku berdasarkan hukum sebab akibat. Contohnya orang yang sakit, maka hendaklah berobat dan mengikuti saran-saran dari dokter agar sembuh. Akan tetapi, yang perlu dipahami bahwa yang menyembuhkan penyakit adalah Allah SWT. Dokter dan obat merupakan perantara dari penyembuhan tersebut.

Rasulullah SAW. bersabda dalam HR. Bukhari:

*“tidaklah Allah menurunkan penyakit kecuali Dia juga menurunkan penawarnya”*

Berdasarkan hadis tersebut, dapat diketahui bahwa segala macam penyakit ada obatnya. Ini merupakan wujud dari rahmat Allah SWT. kepada hambaNya. Akan tetapi, manusia selalu memiliki keterbatasan sehingga tidak jarang penyakit masih ada penyakit yang belum ditemukan obatnya.

Maka dari itu, ilmu pengetahuan memegang peranan yang penting dalam pengembangan teknologi, khususnya di bidang medis. Sesuai dengan metode yang dipakai dalam penelitian ini yaitu *Support Vector Machine*, dapat menjadi acuan bagi peneliti untuk mengklasifikasikan penyakit ke dalam kelas-kelas tertentu (*malignant* dan *benign*) sehingga penyakit yang diderita dapat tertangani sesuai dengan sasaran.

## BAB III METODE PENELITIAN

### 3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dan studi literatur. Pendekatan kuantitatif dilakukan dengan mengumpulkan data berupa angka-angka yang akan diolah, dianalisis, dan diinterpretasikan hasilnya. Sedangkan studi literatur dilakukan dengan mengumpulkan sumber-sumber literatur berupa buku dan jurnal sebagai referensi pendukung dalam penelitian ini.

### 3.2 Data dan Sumber Data

Data dalam penelitian ini merupakan data sekunder berupa data diagnosis kanker payudara Wisconsin (*Wisconsin Diagnostic Breast Cancer*). Data ini diakses dari laman <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic> dengan memilih opsi *Wisconsin Diagnostic Breast Cancer*.

### 3.3 Variabel Penelitian

Penelitian ini menggunakan 11 variabel yang terdiri atas 10 variabel independen dan 1 variabel dependen. 11 variabel tersebut dapat dilihat pada tabel 3.1.

**Tabel 3. 1 Variabel Penelitian**

No	Variabel	Keterangan	Skala
1.	<i>Y</i>	Diagnosis	Nominal: - <i>Benign</i> (B) - <i>Malignant</i> (M)
2.	<i>X1</i>	Radius (nm)	Rasio
3.	<i>X2</i>	<i>Texture</i> (nm)	Rasio

Lanjutan tabel 3.1

No	Variabel	Keterangan	Skala
4.	X3	<i>Perimeter</i> (nm)	Rasio
5.	X4	<i>Area</i> (nm)	Rasio
6.	X5	<i>Smoothness</i> (nm)	Rasio
7.	X6	<i>Compactness</i> (nm)	Rasio
8.	X7	<i>Concavity</i> (nm)	Rasio
9.	X8	<i>Concave Points</i> (nm)	Rasio
10.	X9	<i>Symmetry</i> (nm)	Rasio
11.	X10	<i>Fractal Dimension</i> (nm)	Rasio

### 3.4 Teknik Analisis Data

Langkah-langkah analisis data dalam penentuan klasifikasi diagnosis kanker payudara sebagai berikut:

1. Melakukan analisis deskriptif data

Analisis deskriptif bertujuan untuk mengetahui deskripsi dari masing-masing variabel yang digunakan dalam penelitian.

2. Membagi dataset menjadi data *training* dan data *testing*.

Data *training* digunakan untuk membentuk model *Support Vector Machine*. Sedangkan data *testing* digunakan untuk mengevaluasi model *Support Vector Machine*. Peneliti membagi data *training* dan data *testing* masing-masing 80% dan 20%.

3. Menentukan nilai parameter fungsi kernel yang optimum.

Penelitian ini menggunakan kernel linear. Oleh karena itu, pada tahap ini akan ditentukan nilai parameter C yang optimum.

4. Mengolah data *training*.

Pada tahap ini, data *training* diolah dengan memasukkan nilai parameter dan fungsi kernel yang digunakan. Sehingga diperoleh model *Support Vector Machine*.

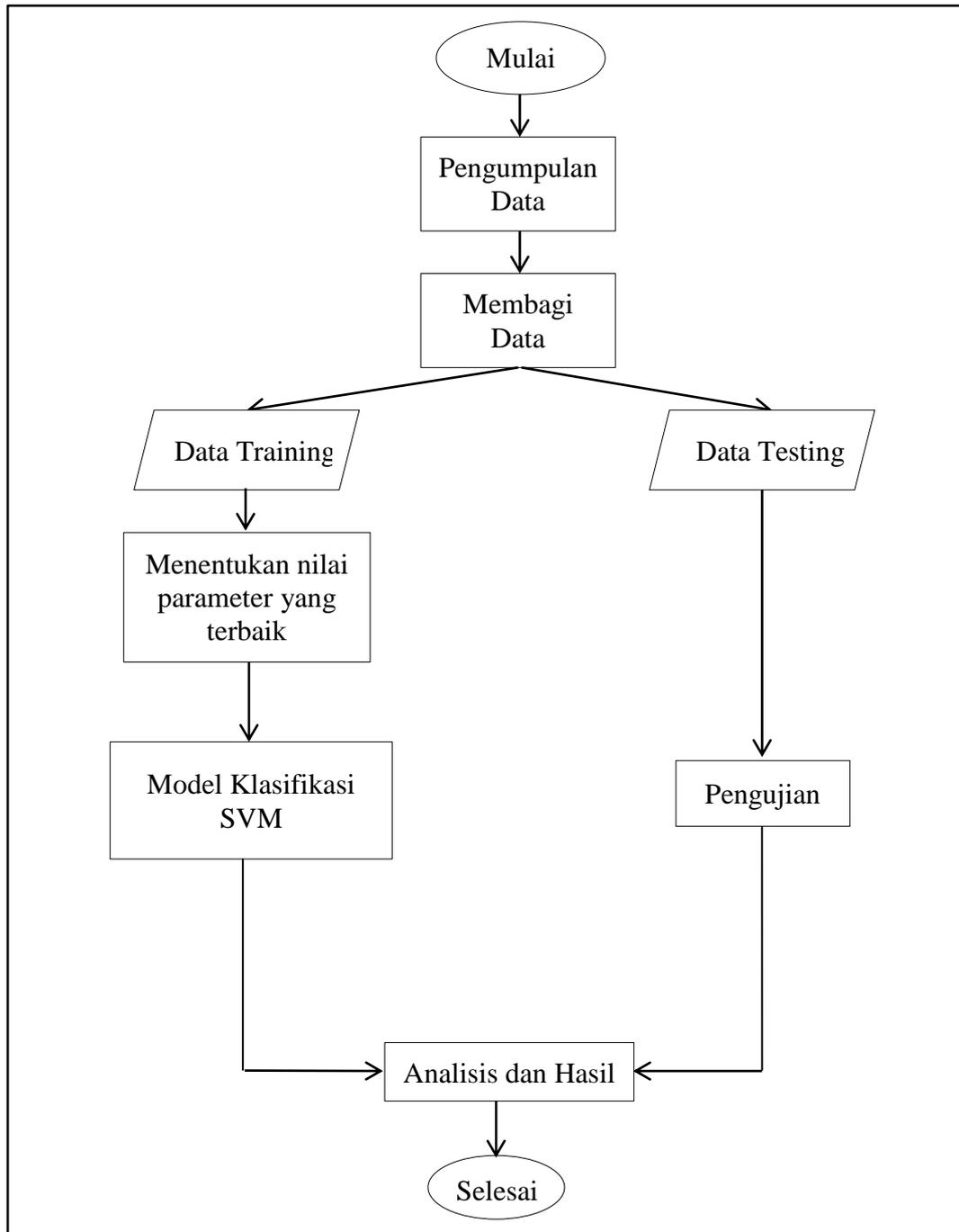
5. Mengolah data *testing*.

Data *testing* digunakan untuk memprediksi dan mengetahui nilai akurasi model klasifikasi data.

6. Menentukan nilai akurasi dari SVM.

Nilai akurasi digunakan untuk mengetahui ketepatan klasifikasi data yang telah diolah.

Alur penelitian dapat dilihat pada *flowchart* berikut:



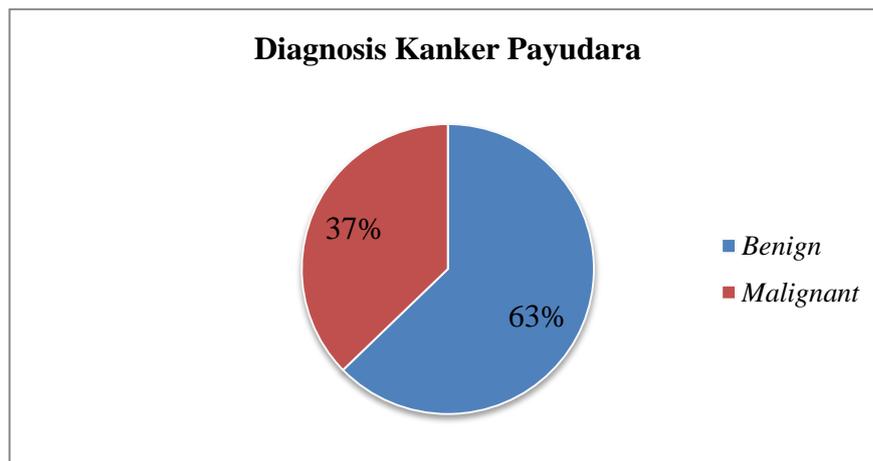
**Gambar 3.1** *Flowchart Penelitian*

## BAB IV PEMBAHASAN

### 4.1 Analisis Deskriptif

Penelitian ini menggunakan 1 variabel dependen dan 10 variabel independen. Variabel dependen pada penelitian ini adalah diagnosis. Sedangkan sepuluh variabel independen yang digunakan antara lain *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*.

Variabel dependen yang digunakan adalah diagnosis kanker payudara. Variabel ini termasuk skala minimal yang terdiri dari dua kategori yaitu *benign* (B) dan *malignant* (M). *Benign* dapat diartikan sebagai tumor jinak dan *malignant* dapat diartikan sebagai tumor ganas. Jika hasil diagnosis menunjukkan *benign*, maka pasien yang telah diperiksa tidak mengidap kanker payudara. Sebaliknya, jika hasil diagnosis menunjukkan *malignant*, maka pasien terdiagnosis kanker payudara.



Gambar 4.1 Pie Chart Diagnosis Kanker Payudara

Berdasarkan *pie chart* pada gambar 4.1, dapat diketahui kategori *benign* lebih tinggi dari kategori *malignant*. Total data pada penelitian ini sebanyak 569

data. Hal ini menunjukkan bahwa dari total 569 data, 63% atau 357 data masuk dalam kategori *benign*. Sementara itu, 37% atau 212 data masuk dalam kategori *malignant*.

Analisis deskriptif dari variabel-variabel dalam penelitian ini dapat dilihat pada tabel berikut:

**Tabel 4.1 Analisis Deskriptif Variabel Penelitian**

Variabel	N	Min.	Max.	Mean	Median	Standar Deviasi
Y	569	-	-	-	-	-
X1	569	6,981	28,110	14,127	13,370	3,524
X2	569	9,71	39,28	19,29	18,84	4,301
X3	569	43,79	188,5	91,97	86,24	24,299
X4	569	143,5	2501	654,9	551,1	351,914
X5	569	0,053	0,163	0,096	0,096	0,014
X6	569	0,019	0,345	0,104	0,093	0,053
X7	569	0	0,427	0,089	0,062	0,080
X8	569	0	0,201	0,049	0,034	0,039
X9	569	0,106	0,304	0,181	0,179	0,027
X10	569	0,05	0,097	0,063	0,062	0,007

Berdasarkan tabel 4.1, seluruh variabel penelitian memiliki jumlah data yang sama yaitu 569 data. Sesuai dengan penjelasan di awal bab ini, variabel Y berskala nominal, sehingga nilai *mean*, *median*, maksimum, minimum, dan standar deviasi tidak ditentukan.

Variabel X1 merupakan variabel radius. Berdasarkan tabel 4.1, radius memiliki nilai terendah sebesar 6,981 nm. Sementara itu, nilai tertinggi untuk variabel radius sebesar 28,110 nm. Dari 569 data, variabel radius memiliki nilai rata-rata sebesar 14,127 nm dengan median 13,370 nm. Sedangkan standar deviasi dari variabel ini sebesar 3,524 nm.

Variabel X2 (*texture*) memiliki rata-rata sebesar 19,29 nm. Variabel ini memiliki nilai terendah sebesar 9,71 nm dan nilai tertinggi sebesar 39,28 nm.

Median dari variabel ini yaitu 18,84 nm dengan standar deviasi 4,301 nm.

Variabel X3 (*perimeter*) memiliki nilai minimum sebesar 43,79 nm, sedangkan nilai maximum sebesar 188,5 nm. X3 (*perimeter*) dengan *mean* 91,97 nm memiliki median 86,24 nm. Adapun standar deviasi dari variabel ini sebesar 24,299 nm.

Variabel X4 yaitu *area*. Berdasarkan tabel 4.1 dapat diketahui bahwa nilai minimum untuk variabel ini sebesar 143,5 nm, sedangkan nilai maksimum sebesar 2502 nm. X4 memiliki nilai rata-rata (*mean*) sebesar 654,9 nm dengan median sebesar 551,1 nm. Dari 569 data, X4 memiliki standar deviasi 351,914 nm.

Variabel X5 (*smoothness*) memiliki nilai minimum 0,053 nm. Nilai maksimum untuk variabel X5 yaitu 0,163 nm. X5 (*smoothness*) memiliki nilai *mean* dan median yang sama yaitu 0,096 nm. Adapun standar deviasi dari variabel ini sebesar 0,014 nm.

Variabel X6 (*compactness*) dengan data sebanyak 569 data memiliki nilai minimum 0,019 nm. Sementara itu, nilai maksimum yang dimiliki 0,345 nm. Nilai *mean* untuk variabel ini 0,104 nm dengan median 0,093 nm. Adapun standar deviasi pada variabel ini yaitu 0,053 nm.

Variabel independen berikutnya yaitu X7 (*concavity*) memiliki nilai minimum 0 nm dan nilai maksimum 0,427 nm. Nilai rata-rata (*mean*) untuk X7 sebesar 0,089 nm dengan median 0,062 nm. Sementara itu, standar deviasi yang dimiliki sebesar 0,080 nm.

Berdasarkan tabel 4.1, variabel X8 (*concave points*) memiliki nilai rata-rata (*mean*) 0,049 nm. Dari total 569 data, variabel X8 memiliki nilai minimum 0

nm dan nilai maksimum 0,201 nm. Sementara itu, median dan standar deviasi dari variabel ini masing-masing sebesar 0,034 nm dan 0,039 nm.

Variabel X9 yaitu *symmetry*. Dari 569 data, variabel *symmetry* memiliki nilai rata-rata (*mean*) 0,181 nm. Nilai minimum dan maksimum variabel ini yaitu 0,106 nm dan 0,304 nm. Adapun median yang dimiliki yaitu 0,179 nm dengan standar deviasi 0,027 nm.

Nilai minimum untuk X10 (*fractal dimension*) sebesar 0,05 nm, sedangkan nilai maksimum 0,097 nm. Dari total 569 data, variabel ini memiliki nilai rata-rata 0,063 nm dengan median 0,062 nm. Adapun standar deviasi dari X10 (*fractal dimension*) sebesar 0,007 nm.

## **4.2 Implementasi Support Vector Machine**

### **4.2.1 Split Data**

Tahap ini merupakan tahap pemisahan data menjadi dua, yaitu data *training* dan data *testing*. Data *training* digunakan untuk membentuk model klasifikasi pada SVM, sedangkan data *testing* digunakan untuk mengevaluasi model klasifikasi kanker payudara. Proses pembagian data menjadi data *training* dan *testing* melalui beberapa presentase dan dipilih perbandingan 80:20 untuk data *training* dan *testing*. Oleh karena itu, dari 569 observasi, 455 diantaranya masuk ke dalam *training set* dan 114 data masuk ke dalam *testing set*.

### **4.2.2 Klasifikasi Diagnosis Kanker Payudara dengan SVM**

Klasifikasi model SVM pada penelitian ini menggunakan fungsi kernel linear. Oleh karena itu, parameter yang digunakan adalah parameter *cost* ( $C$ ). Berdasarkan persamaan (2.15), nilai pengali *lagrange* ( $\alpha_i$ ) berada di antara 0 dan

C. Dengan demikian, dapat diketahui bahwa parameter *cost* bernilai positif. Adapun rentang nilai  $C$  yang dipilih yaitu  $C = 0,01$ ,  $C = 0,1$ ,  $C = 1$ ,  $C = 10$ , dan  $C = 100$ . sehingga diperoleh lima model klasifikasi sebagai berikut:

**Tabel 4. 2 Model SVM dengan  $C = 0,01$**

SVM – type	<i>C-Classificaton</i>
SVM – kernel	<i>Linear</i>
<i>Cost</i>	0,01
<i>Number of Support Vectors</i>	150

**Tabel 4. 3 Model SVM dengan  $C = 0,1$**

SVM – type	<i>C-Classificaton</i>
SVM – kernel	<i>Linear</i>
<i>Cost</i>	0,1
<i>Number of Support Vectors</i>	91

**Tabel 4. 4 Model SVM dengan  $C = 1$**

SVM – type	<i>C-Classificaton</i>
SVM – kernel	<i>Linear</i>
<i>Cost</i>	1
<i>Number of Support Vectors</i>	74

**Tabel 4. 5 Model SVM dengan  $C = 10$**

SVM – type	<i>C-Classificaton</i>
SVM – kernel	<i>Linear</i>
<i>Cost</i>	10
<i>Number of Support Vectors</i>	68

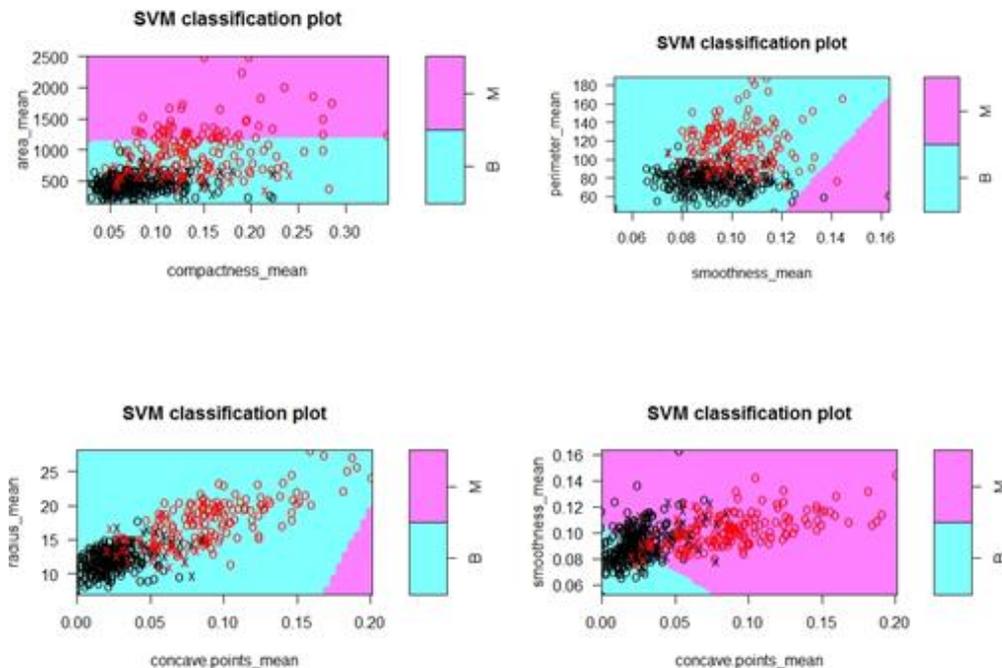
**Tabel 4. 6 Model SVM dengan  $C = 100$**

SVM – type	<i>C-Classificaton</i>
SVM – kernel	<i>Linear</i>
<i>Cost</i>	100
<i>Number of Support Vectors</i>	69

Berdasarkan tabel 4.2 sampai 4.6, diketahui bahwa jumlah *support vector* dari masing-masing nilai *cost* berbeda. Menurut Han, dkk (2012), generalisasi yang baik akan diperoleh apabila jumlah *support vector* pada model SVM lebih kecil. Oleh karena itu, di antara kelima model *Supprt Vector Machine* yang ada pada tabel 4.2 sampai 4.6, model SVM dengan  $C = 10$  merupakan model terbaik.

Visualisasi dari model klasifikasi SVM ditampilkan dalam bentuk plot.

Plot ini terbentuk dari kombinasi dua atribut variabel independen. Dari kombinasi dua variabel independen secara keseluruhan diperoleh plot-plot sebagai berikut:



**Gambar 4. 2** Plot SVM dengan  $C = 10$

Total kombinasi dari dua variabel independen pada penelitian ini diperoleh 45 plot. Plot lengkap dapat dilihat di lampiran 2. Dari 45 plot tersebut, 27 plot diantaranya mampu mengklasifikasi diagnosis kanker payudara dengan jelas. Sebagai contoh, dari gambar 4.2, data yang berada pada area merah muda masuk pada kategori *malignant* (tumor ganas), sedangkan data pada area warna biru masuk kategori *benign* (tumor jinak). Selain itu, titik atau *point* pada data tersebut memiliki bentuk silang dan bulat. Jika *point* data berbentuk silang, maka data tersebut merupakan bagian dari *support vector*.

Berdasarkan plot SVM yang diperoleh, dapat diketahui bahwa variabel *radius* ( $X_1$ ) dan *perimeter* ( $X_3$ ) merupakan variabel yang paling berpengaruh pada klasifikasi diagnosis pasien kategori *benign*. Sedangkan variabel *texture*

( $X_2$ ) dan *smoothness* ( $X_5$ ) merupakan variabel yang paling berpengaruh terhadap diagnosis kanker payudara kategori *malignant* (tumor ganas). Hal ini disebabkan pengamatan dengan citra digital semakin memperjelas tekstur dari sel-sel yang diamati sehingga memperjelas diagnosis tersebut.

#### 4.2.3 Akurasi Model Klasifikasi Kanker Payudara

Ketepatan klasifikasi pada model *Support Vector Machine* dapat ditentukan melalui *confusion matrix*. *Confusion matrix* terdiri dari nilai prediksi dan aktual yang digunakan untuk menghitung tingkat akurasi, *sensitivity*, dan *specificity*. Berikut tabel *confusion matrix* untuk model SVM dengan  $C = 10$ :

**Tabel 4.7 Confusion Matrix pada Diagnosis Kanker Payudara**

<b>Prediksi</b>	<b>Aktual</b>	
	<i>Benign</i>	<i>Malignant</i>
<i>Benign</i>	69	1
<i>Malignant</i>	3	41

Berdasarkan *confusion matrix* pada tabel 4.7, diperoleh bahwa 69 data terprediksi benar masuk dalam kategori *benign* (tumor jinak) dan 41 data terprediksi benar masuk kategori *malignant*. Sementara itu, 1 data yang diprediksi masuk dalam kategori *benign* ternyata merupakan kategori *malignant* (tumor ganas). Sedangkan 3 data yang diprediksi masuk kategori *malignant* pada kenyataannya merupakan kategori *benign*. Dari hasil tersebut, dapat diketahui bahwa 69 merupakan *true negative*, 1 merupakan *false negative*, 3 merupakan *false positive*, dan 41 merupakan *true positive*. Perhitungan nilai akurasi, *sensitivity*, dan *specificity* berdasarkan persamaan (2.26) sampai (2.28) sebagai berikut:

$$Akurasi = \frac{41 + 69}{41 + 1 + 3 + 69} \times 100\% = 96,49\%$$

$$Sensitivity = \frac{41}{41 + 1} \times 100\% = 97,62\%$$

$$Specificity = \frac{69}{69 + 3} \times 100\% = 95,83\%$$

Dari hasil perhitungan tersebut, maka dapat diketahui bahwa tingkat ketepatan model SVM pada diagnosis kanker payudara sebesar 96,49%. Hal ini menunjukkan bahwa metode SVM dapat bekerja dengan baik dalam penentuan diagnosis kanker payudara. Nilai *sensitivity* sebesar 97,62% menunjukkan bahwa model SVM mampu mengidentifikasi diagnosis kanker payudara sebagai kelas positif sebesar 97,62% dengan benar. Sedangkan nilai *specificity* sebesar 95,83% menunjukkan bahwa model SVM mampu mengidentifikasi diagnosis kanker payudara sebagai kelas negatif sebesar 95,83% dengan benar.

### 4.3 Kajian Agama Terkait Penelitian

Proses pengambilan sel dan jaringan dari pasien kemudian dicek menggunakan mikroskop merupakan salah satu bentuk kemajuan di bidang teknologi dan kesehatan. Berkat penggunaan citra digital dan bantuan komputer, tenaga medis dimudahkan dalam melakukan tugasnya dalam mendiagnosis pasien. Hal ini menjadi salah satu langkah preventif, baik bagi pihak pasien maupun tenaga medis. Pasien dapat meminimalisir resiko yang terjadi jika ternyata didiagnosis positif kanker payudara, begitu pula pihak medis dapat lebih cepat melakukan tindakan penanganan dini untuk mencegah pertumbuhan sel kanker. Ini semua merupakan bentuk ikhtiar yang dilakukan agar dapat sembuh dan kembali beraktifitas sebagaimana mestinya.

Allah SWT. berfirman dalam surah *An-Najm* ayat 39-40 yang artinya:

*“dan bahwasanya seorang manusia tiada memperoleh selain apa yang telah diusahakannya. Dan bahwasanya usahanya itu kelak akan diperlihatkan kepadanya.”* (Q.S. An-Najm: 39-40).

Menurut Shihab (2002), Allah SWT. menegaskan apapun yang diusahakan manusia akan diperlihatkan hasilnya. Oleh karena itu, manusia dapat bahagia dengan hal baik yang dilakukan atas usaha tersebut. Dalam kaitannya dengan penelitian ini, jika manusia berikhtiar untuk sembuh, maka Allah SWT. yang akan menunjukkan jalanNya menuju kesembuhan. Jalan ikhtiar yang dilakukan yaitu dengan segera berobat dan mencari tahu lebih dalam mengenai penyakit yang diderita.

Penyakit kanker payudara yang sebagian besar diderita oleh perempuan tentu menjadi kekhawatiran tersendiri bagi pihak perempuan. Penyakit ini tidak mengenal status maupun umur. Tua dan muda tidak luput dari penyakit ini. Terlebih bagi ibu yang sedang menyusui anaknya. Sesuai dengan penjelasan pada bab 1, dalam Q.S. Al-Baqarah ayat 233, Allah SWT. menganjurkan para ibu untuk menyusui anaknya selama 2 tahun. Jika seorang ibu terkena kanker payudara, hal ini dapat mengganggu proses pemberian ASI kepada bayi. Maka dari itu, ikhtiar yang dapat dilakukan sejak dini adalah dengan segera memeriksa diri ke dokter apabila ada gejala-gejala tertentu yang dicurigai untuk mencegah hal-hal yang tidak diinginkan di kemudian hari.

## **BAB V PENUTUP**

### **5.1 Kesimpulan**

Berdasarkan rumusan masalah dan pembahasan bab sebelumnya, maka dapat diperoleh kesimpulan:

1. Hasil klasifikasi diagnosis kanker payudara dengan *Support Vector Machine* menunjukkan model terbaik ketika nilai parameter fungsi kernel linear ( $C$ ) = 10. Berdasarkan plot SVM yang terbentuk, dapat diketahui bahwa variabel *radius* dan *perimeter* memiliki pengaruh yang besar terhadap diagnosis kanker payudara kategori *benign* (tumor jinak), sedangkan variabel *texture* dan *smoothness* memiliki pengaruh besar terhadap diagnosis kanker payudara kategori *malignant*.
2. Tingkat akurasi model SVM dengan nilai parameter  $cost = 10$  sebesar 96,49%. Hal ini menunjukkan bahwa metode SVM dengan kernel linear dapat bekerja dengan baik dalam penentuan klasifikasi diagnosis kanker payudara.

### **5.2 Saran**

Beberapa hal yang dapat dijadikan acuan dalam menyempurnakan penelitian yang akan datang antara lain:

1. Peneliti selanjutnya pembaca dapat menggunakan metode klasifikasi yang berbeda
2. Peneliti selanjutnya dapat menggunakan variabel diagnosis berdasarkan level stadium kanker, sehingga terbentuk 4 kelas klasifikasi yang berbeda.

3. Peneliti dapat menggunakan fungsi kernel dan parameter yang lain.
4. Dengan metode yang sama, peneliti selanjutnya dapat meneliti klasifikasi diagnosis penyakit yang lain.

## DAFTAR PUSTAKA

- Gholami & Fakhari. (2017). *Support Vector Machine: Principles, Parameters, and Applications*.
- Gunawan, dkk. (2014). Identifikasi Jenis Kayu Menggunakan Support Vector Machine. *Jurnal Ilmu Komputer Agri-Informatika*. No. 1. Vol. 3. <https://doi.org/10.29244/jika.3.1.1-8>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-61819-5>
- Hidayat, dkk. (2021). Pengaruh Algoritma ADASYN dan SMOTE terhadap Support Vector Machine pada Ketidakseimbangan Dataset Airbnb. *Jurnal Pendidikan Informatika*. No. 1. Vol. 5. <https://doi.org/10.29408/edumatic.v5i1.3125>
- Indriani, dkk. (2017). Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *Jurnal Infotel*. No. 4. Vol. 9. DOI: 10.20895/infotel.v9i4.312
- Iqmy, dkk. (2021). Faktor Risiko yang Berhubungan dengan Kanker. *Jurnal Kebidanan*. No. 1. Vol. 7. <http://dx.doi.org/10.33024/jkm.v7i1.3581>
- Kasim, dkk. (2019). Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah. *Seminar Nasional APTIKOM (SEMNASTIK)*. <http://publikasi.dinus.ac.id/index.php/semnastik/article/view/2922/1785>
- Ketut & Kartika. (2022). Kanker Payudara: Diagnostik, Faktor Risiko, dan Stadium. *Ganesha Medicina Journal*. No. 1. Vol. 2. <https://doi.org/10.23887/gm.v2i1.47032>
- Muis & Affandes. (2015). Penerapan Metode Support Vector Machine menggunakan Kernel Radial Basis Function (RBF) pada Klasifikasi Tweet. *Jurnal Sains & Teknologi Industri*. No. 2 Vol. 12. <http://dx.doi.org/10.24014/sitekin.v12i2.1010>
- Munawarah, dkk. 2016. Penerapan Metode Support Vector Machine pada Diagnosis Hepatitis. *Jurnal Ilmu Komputer (KLIK)*. No. 01 Vol. 04. <http://dx.doi.org/10.20527/klik.v3i1.39>
- Octaviani, dkk (2014). Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. *Jurnal Gaussian*. No. 3. Vol. 4. <https://doi.org/10.14710/j.gauss.3.4.811-820>

- Permadi, dkk. (2016). Sistem Pakar Diagnosa Penyakit Kanker Payudara Menggunakan Metode Forward Chaining dan Depth First Search. *Jurnal Fasilkom*. No. 2. Vol. 5. Ketut
- Putra, Sitiatava Rizema. (2015). *Buku Lengkap Kanker Payudara*. Yogyakarta: Laksana.
- Shahura, dkk. (2016). Penerapan Metode RBPNN untuk Klasifikasi Kanker Payudara *Kumpulan Jurnal Ilmu Komputer*. No. 02 Vol. 03. <http://dx.doi.org/10.20527/klik.v3i2.51>
- Shihab, M. Quraish. (2002). *Tafsir Al Mishbah Pesan Kesan dan Keserasian Al Qur'an Volume 1*. Jakarta: Lentera Hati.
- Shihab, M. Quraish. (2002). *Tafsir Al Mishbah Pesan Kesan dan Keserasian Al Qur'an Volume 10*. Jakarta: Lentera Hati.
- Shihab, M. Quraish. (2002). *Tafsir Al Mishbah Pesan Kesan dan Keserasian Al Qur'an Volume 13*. Jakarta: Lentera Hati.
- Smola & Scholkopf. (2002). *Support Vector Machine and Kernel Algorithms*.
- Vapnik & Cortes. 1995. Support Vector Networks. *Machine Learning*.
- Wolberg, William H. 1995. *Wisconsin Diagnostic Breast Cancer*. <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>

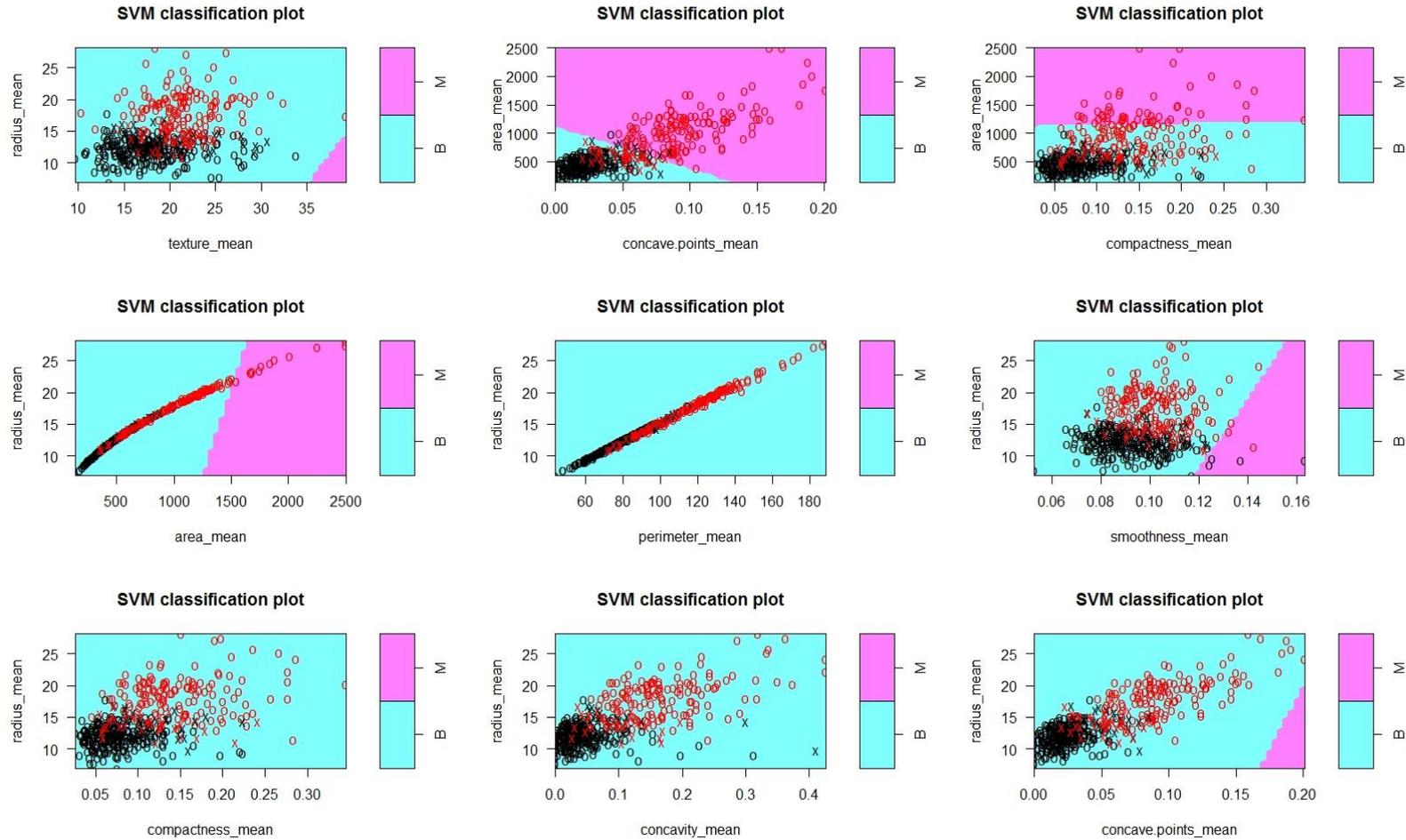
## LAMPIRAN

**Lampiran 1.** *Data Wisconsin Breast Cancer Diagnostic*

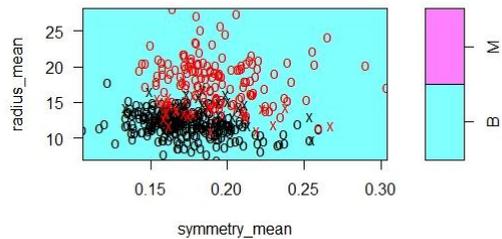
	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
1	M	17.990	10.38	122.80	1001.0	0.11840	0.27760	0.300100	0.147100	0.2419	0.07871
2	M	20.570	17.77	132.90	1326.0	0.08474	0.07864	0.086900	0.070170	0.1812	0.05667
3	M	19.690	21.25	130.00	1203.0	0.10960	0.15990	0.197400	0.127900	0.2069	0.05999
4	M	11.420	20.38	77.58	386.1	0.14250	0.28390	0.241400	0.105200	0.2597	0.09744
5	M	20.290	14.34	135.10	1297.0	0.10030	0.13280	0.198000	0.104300	0.1809	0.05883
6	M	12.450	15.70	82.57	477.1	0.12780	0.17000	0.157800	0.080890	0.2087	0.07613
7	M	18.250	19.98	119.60	1040.0	0.09463	0.10900	0.112700	0.074000	0.1794	0.05742
8	M	13.710	20.83	90.20	577.9	0.11890	0.16450	0.093660	0.059850	0.2196	0.07451
9	M	13.000	21.82	87.50	519.8	0.12730	0.19320	0.185900	0.093530	0.2350	0.07389
10	M	12.460	24.04	83.97	475.9	0.11860	0.23960	0.227300	0.085430	0.2030	0.08243
11	M	16.020	23.24	102.70	797.8	0.08206	0.06669	0.032990	0.033230	0.1528	0.05697
12	M	15.780	17.89	103.60	781.0	0.09710	0.12920	0.099540	0.066060	0.1842	0.06082
13	M	19.170	24.80	132.40	1123.0	0.09740	0.24580	0.206500	0.111800	0.2397	0.07800
14	M	15.850	23.95	103.70	782.7	0.08401	0.10020	0.099380	0.053640	0.1847	0.05338
15	M	13.730	22.61	93.60	578.3	0.11310	0.22930	0.212800	0.080250	0.2069	0.07682
16	M	14.540	27.54	96.73	658.8	0.11390	0.15950	0.163900	0.073640	0.2303	0.07077
17	M	14.680	20.13	94.74	684.5	0.09867	0.07200	0.073950	0.052590	0.1586	0.05922
18	M	16.130	20.68	108.10	798.8	0.11700	0.20220	0.172200	0.102800	0.2164	0.07356
19	M	19.810	22.15	130.00	1260.0	0.09831	0.10270	0.147900	0.094980	0.1582	0.05395
20	B	13.540	14.36	87.46	566.3	0.09779	0.08129	0.066640	0.047810	0.1885	0.05766
21	B	13.080	15.71	85.63	520.0	0.10750	0.12700	0.045680	0.031100	0.1967	0.06811
22	B	9.504	12.44	60.34	273.9	0.10240	0.06492	0.029560	0.020760	0.1815	0.06905

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave.points_mean	symmetry_mean	fractal_dimension_mean
548	B	10.260	16.58	65.85	320.8	0.08877	0.08066	0.043580	0.024380	0.1669	0.06714
549	B	9.683	19.34	61.05	285.7	0.08491	0.05030	0.023370	0.009615	0.1580	0.06235
550	B	10.820	24.21	68.89	361.6	0.08192	0.06602	0.015480	0.008160	0.1976	0.06328
551	B	10.860	21.48	68.51	360.5	0.07431	0.04227	0.000000	0.000000	0.1661	0.05948
552	B	11.130	22.44	71.49	378.4	0.09566	0.08194	0.048240	0.022570	0.2030	0.06552
553	B	12.770	29.43	81.35	507.9	0.08276	0.04234	0.019970	0.014990	0.1539	0.05637
554	B	9.333	21.94	59.01	264.0	0.09240	0.05605	0.039960	0.012820	0.1692	0.06576
555	B	12.880	28.92	82.50	514.3	0.08123	0.05824	0.061950	0.023430	0.1566	0.05708
556	B	10.290	27.61	65.67	321.4	0.09030	0.07658	0.059990	0.027380	0.1593	0.06127
557	B	10.160	19.59	64.73	311.7	0.10030	0.07504	0.005025	0.011160	0.1791	0.06331
558	B	9.423	27.88	59.26	271.3	0.08123	0.04971	0.000000	0.000000	0.1742	0.06059
559	B	14.590	22.68	96.39	657.1	0.08473	0.13300	0.102900	0.037360	0.1454	0.06147
560	B	11.510	23.93	74.52	403.5	0.09261	0.10210	0.111200	0.041050	0.1388	0.06570
561	B	14.050	27.15	91.38	600.4	0.09929	0.11260	0.044620	0.043040	0.1537	0.06171
562	B	11.200	29.37	70.67	386.0	0.07449	0.03558	0.000000	0.000000	0.1060	0.05502
563	M	15.220	30.62	103.40	716.9	0.10480	0.20870	0.255000	0.094290	0.2128	0.07152
564	M	20.920	25.09	143.00	1347.0	0.10990	0.22360	0.317400	0.147400	0.2149	0.06879
565	M	21.560	22.39	142.00	1479.0	0.11100	0.11590	0.243900	0.138900	0.1726	0.05623
566	M	20.130	28.25	131.20	1261.0	0.09780	0.10340	0.144000	0.097910	0.1752	0.05533
567	M	16.600	28.08	108.30	858.1	0.08455	0.10230	0.092510	0.053020	0.1590	0.05648
568	M	20.600	29.33	140.10	1265.0	0.11780	0.27700	0.351400	0.152000	0.2397	0.07016
569	B	7.760	24.54	47.92	181.0	0.05263	0.04362	0.000000	0.000000	0.1587	0.05884

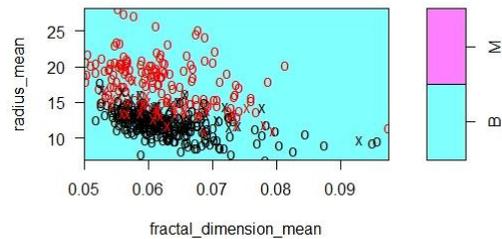
## Lampiran 2. Plot SVM



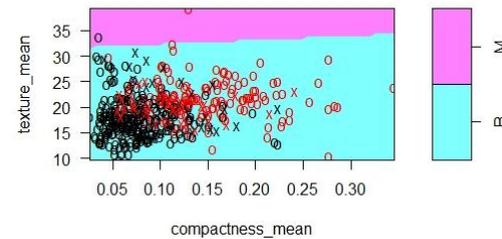
**SVM classification plot**



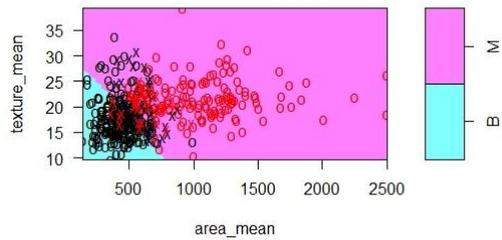
**SVM classification plot**



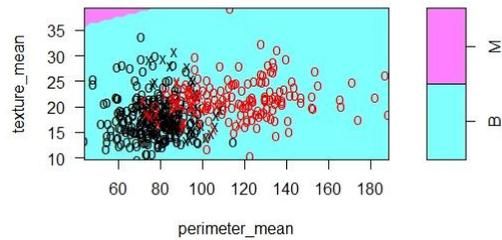
**SVM classification plot**



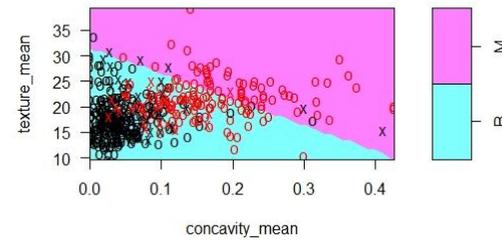
**SVM classification plot**



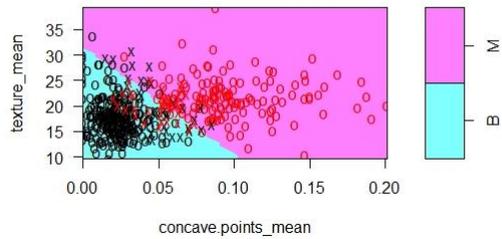
**SVM classification plot**



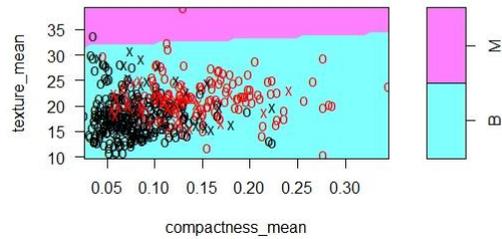
**SVM classification plot**



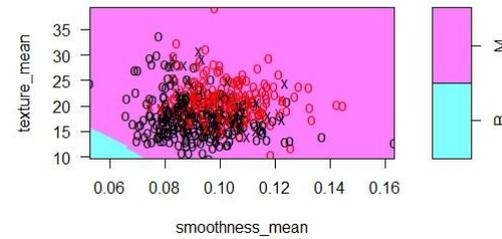
**SVM classification plot**



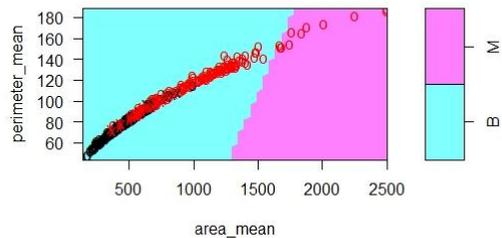
**SVM classification plot**



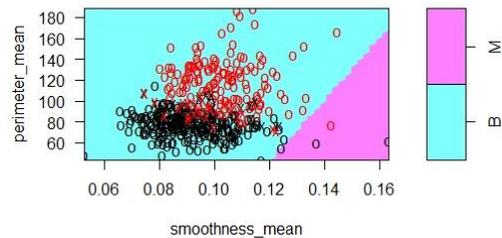
**SVM classification plot**



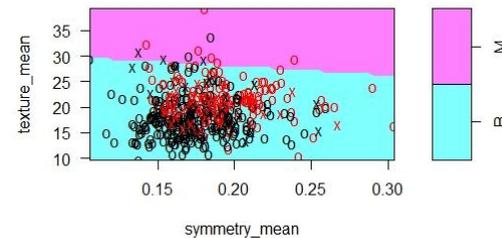
**SVM classification plot**



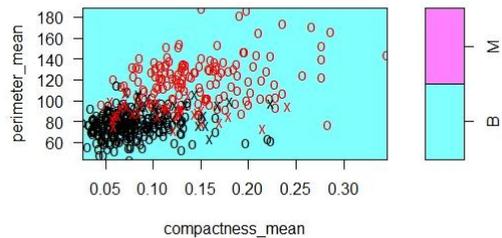
**SVM classification plot**



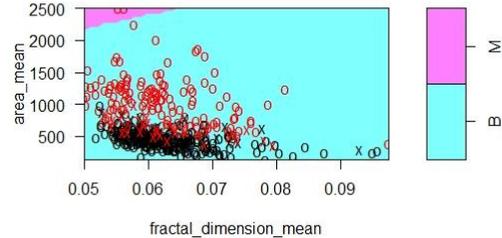
**SVM classification plot**



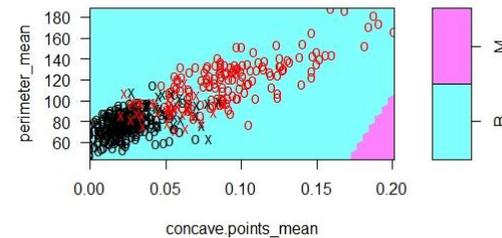
**SVM classification plot**



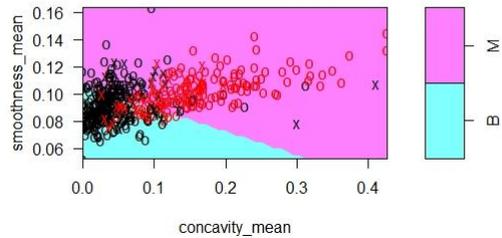
**SVM classification plot**



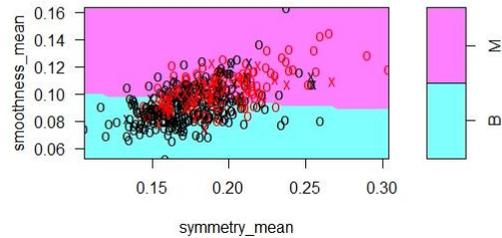
**SVM classification plot**



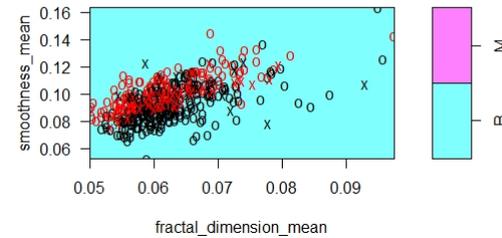
**SVM classification plot**

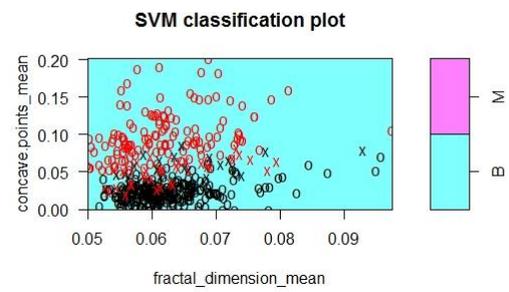
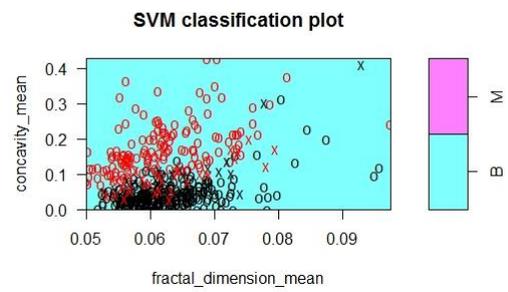
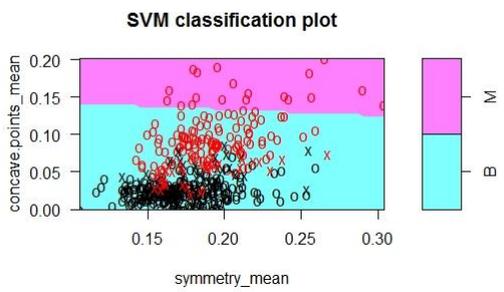
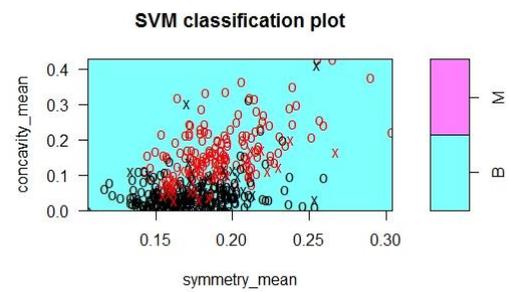
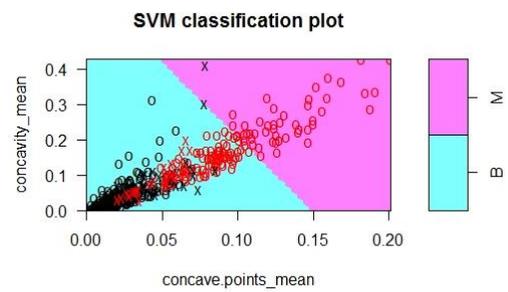
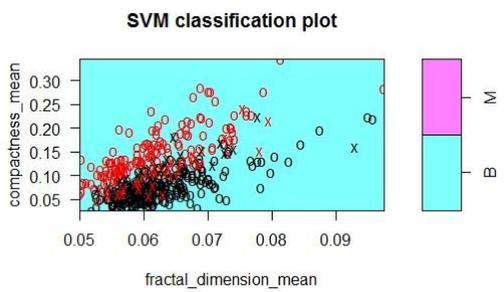
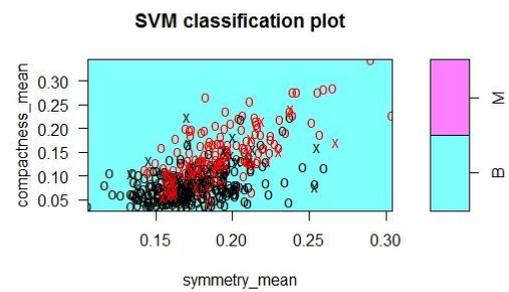
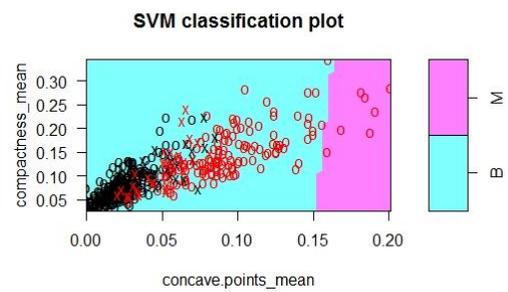
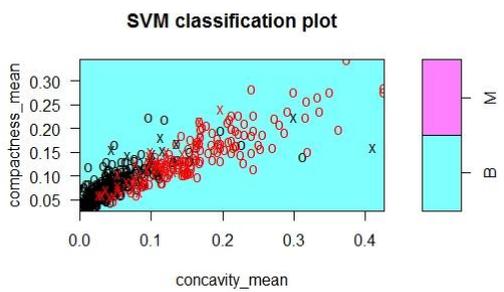


**SVM classification plot**

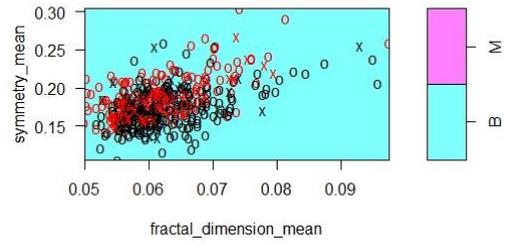


**SVM classification plot**





**SVM classification plot**



### Lampiran 3. Script SVM di R Studio

```
#menyeleksi data yang akan digunakan untuk svm
library(dplyr)
mydata=baru[2:12]
mydata

#split data menjadi training set (80%) dan testing set (20%)
set.seed(123)
split = sample.int(n=nrow(mydata), size=floor(.80*nrow(mydata)), replace=F)
training_set=mydata[split, ]
training_set
training_set$diagnosis = as.factor(training_set$diagnosis)
training_set$diagnosis
testing_set=mydata[-split, ]
testing_set

#Membentuk model SVM
library(kernlab)
library(e1071)
library(RColorBrewer)
library(caret)
library(ggplot2)
classifier = svm(formula = diagnosis ~ .,
                data = training_set,
                type = 'C-classification',
                cost = 0.01,
                kernel = 'linear')
summary(classifier)

classifier2 = svm(formula = diagnosis ~ .,
                 data = training_set,
                 type = 'C-classification',
                 cost = 0.1,
                 kernel = 'linear')
summary(classifier2)

classifier3 = svm(formula = diagnosis ~ .,
                 data = training_set,
                 type = 'C-classification',
                 cost = 1,
                 kernel = 'linear')
summary(classifier3)

classifier4 = svm(formula = diagnosis ~ .,
                 data = training_set,
                 type = 'C-classification',
```

```

        cost = 10,
        kernel = 'linear')
summary(classifier4)

classifier5 = svm(formula = diagnosis ~ .,
                 data = training_set,
                 type = 'C-classification',
                 cost = 100,
                 kernel = 'linear')
summary(classifier5)

#plot svm untuk variabel independen x1 dan x2
plot(classifier4, data=training_set, radius_mean~texture_mean)

#prediksi model svm menggunakan data testing
hij=testing_set[,-1]
hij
prediksi = predict(classifier4, hij)
prediksi

#Confusion Matrix
result_table = table(Predicted = prediksi, actual = testing_set$diagnosis)
result_table

accuracy <- sum(result_table[1], result_table[4]) / sum(result_table[1:4])
accuracy
precision <- result_table[4] / sum(result_table[4], result_table[2])
precision
sensitivity <- result_table[4] / sum(result_table[4], result_table[3])
sensitivity
fscore <- (2 * (sensitivity * precision)) / (sensitivity + precision)
fscore
specificity <- result_table[1] / sum(result_table[1], result_table[2])
specificity

```

## RIWAYAT HIDUP



Siti Hurriyati, lahir di Selong, Kabupaten Lombok Timur pada tanggal 21 Agustus 1998. Putri dari pasangan Bapak H. Fadlullah dan Ibu Subhiah ini telah menempuh pendidikan di TK LKMD selama 6 bulan, lalu dilanjutkan di TK Raudatul Athfal selama 6 bulan. Penulis kemudian melanjutkan jenjang pendidikan dasar di MIN Kefamenanu, NTT selama 5 tahun kemudian pindah dan menamatkan jenjang sekolah dasar di SDN 2 Kelayu Selatan, NTB. Penulis menempuh pendidikan menengah pertama di MTsN 1 Lombok Timur pada tahun 2010 dan lulus di tahun 2013. Selanjutnya penulis menempuh pendidikan menengah atas di MAN Selong (yang sekarang menjadi MAN 1 Lombok Timur) dan lulus pada tahun 2016. Di tahun yang sama, penulis diterima menjadi mahasiswa Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama di jenjang kuliah, penulis sempat mengikuti beberapa komunitas di jurusan matematika, diantaranya Mathematics English Club (MEC) dan Mathematics Arab Club (MAC). Selain itu, penulis juga mengikuti organisasi daerah di awal-awal masa kuliah selama 1 tahun.



KEMENTERIAN AGAMA RI  
UNIVERSITAS ISLAM NEGERI  
MAULANA MALIK IBRAHIM MALANG  
FAKULTAS SAINS DAN TEKNOLOGI  
Jl. Gajayana No. 50 Dinoyo Malang Telp./Fax. (0341)558933

### BUKTI KONSULTASI SKRIPSI

Nama : Siti Hurriyati  
NIM : 16610041  
Fakultas / Program Studi : Sains dan Teknologi / Matematika  
Judul Skripsi : Implementasi Metode *Support Vector Machine*  
pada Klasifikasi Diagnosis Penyakit Kanker  
Payudara  
Pembimbing I : Ria Dhea Layla Nur Karisma, M.Si  
Pembimbing II : Mohammad Nafie Jauhari, M.Si

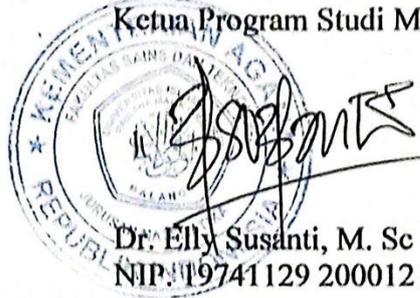
No.	Tanggal	Hal	Tanda Tangan
1.	16 Januari 2023	Konsultasi Bab I	1.
2.	24 Januari 2023	Konsultasi Revisi Bab I	2.
3.	6 Februari 2023	Konsultasi Bab II & III	3.
4.	7 Februari 2023	Konsultasi Kajian Agama Bab I dan II	4.
5.	9 Februari 2023	Konsultasi Revisi Bab II & III	5.
6.	10 Februari 2023	Konsultasi Revisi Kajian Agama Bab I & II	6.
7.	14 Maret 2023	ACC Seminar Proposal	7.
8.	24 Mei 2023	Konsultasi Kajian Agama Bab IV	8.
9.	20 Juni 2023	ACC Seminar Hasil	9.
10.	22 Juni 2023	Revisi Seminar Hasil	10.

11.	23 Juni 2023	ACC Sidang Skripsi	11. <i>[Signature]</i>
12.	30 Juni 2023	ACC Keseluruhan	12. <i>[Signature]</i>

Malang, 30 Juni 2023

Mengetahui,

Ketua Program Studi Matematika



*[Signature]*

Dr. Elly Susanti, M. Sc

NIP. 19741129 200012 2 005