

**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE*  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT HIPERTENSI**

**SKRIPSI**

**OLEH:  
HILDA ZAQYA ELNAZ PUTRI  
NIM. 200601110112**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE*  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT HIPERTENSI**

**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  
Hilda Zaqya Elnaz Putri  
NIM. 200601110112**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2024**

**IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE*  
PADA KLASIFIKASI DIAGNOSIS PENYAKIT HIPERTENSI**

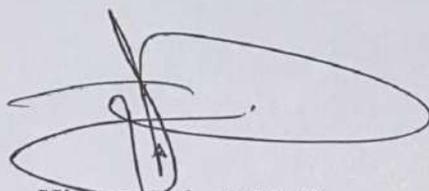
**SKRIPSI**

Oleh  
**Hilda Zaqya Elnaz Putri**  
**NIM. 200601110112**

Telah Disetujui Untuk Diuji

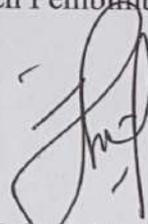
Malang, 16 Mei 2024

Dosen Pembimbing I



Hisyam Fahmi, M. Kom.  
NIP. 19890727 201903 1 018

Dosen Pembimbing II



Juhari, M. Si.  
NIPPPK. 19840209 202321 1 010

Mengetahui,  
Ketua Program Studi Matematika



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

# IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE* PADA KLASIFIKASI DIAGNOSIS PENYAKIT HIPERTENSI

## SKRIPSI

Oleh  
**Hilda Zaqya Elnaz Putri**  
NIM. 200601110112

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Matematika (S. Mat)

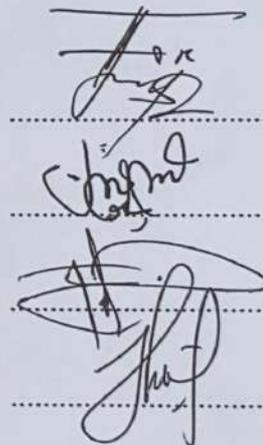
Tanggal 30 Mei 2024

Ketua Penguji : Dr. Fachrur Rozi, M.Si.

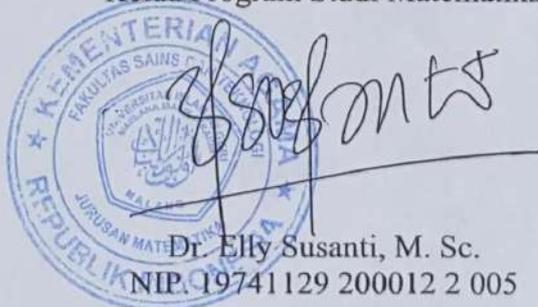
Anggota Penguji 1 : Ria Dhea Layla Nur Karisma, M.Si.

Anggota Penguji 2 : Hisyam Fahmi, M.Kom.

Anggota Penguji 3 : Juhari, 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 : Hilda Zaqya Elnaz Putri  
NIM : 200601110112  
Program Studi : Matematika  
Fakultas : Sains dan Teknologi  
Judul Skripsi : Implementasi Metode *Support Vector Machine* pada  
Klasifikasi Diagnosis Penyakit Hipertensi

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 Mei 2024

Yang membuat pernyataan,



Hilda Zaqya Elnaz Putri

NIM. 200601110112

## **MOTO**

*“Sesungguhnya sesudah kesulitan pasti ada kemudahan”*

*(Q.S. Al-Insyirah:6)*

*“Tidak peduli seberapa sulit atau mustahilnya itu, jangan pernah melupakan  
tujuanmu”*

*(Monkey D Luffy: One Piece)*

## **HALAMAN PERSEMBAHAN**

Skripsi ini penulis persembahkan kepada:

Ayah dan bunda tercinta Nazri dan Elvasari yang senantiasa memberikan doa, dukungan, nasihat dan motivasi terbaik untuk kesuksesan penulis. Kakak tersayang Imelda Zahra Elnaz Putri yang selalu memberikan doa, motivasi dan semangat kepada penulis. Adik tersayang Anindya Khanza Elnaz Putri yang dengan kasih sayangnya senantiasa mendukung penuh penulis dalam segala hal. Serta sahabat-sahabat penulis yang selalu memberikan bantuan dan semangat dalam menyelesaikan skripsi ini.

## KATA PENGANTAR

*Assalamu 'allaikum Warrahmatullahi Wabarakatuh*

Segala puji bagi Allah atas rahmat, taufik serta hidayah-Nya, sehingga penulis mampu menyelesaikan penyusunan proposal skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana dalam bidang Matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Selama proses penyusunan skripsi ini, penulis mendapatkan banyak bimbingan, arahan serta dukungan dari berbagai pihak. Sebagai bentuk terimakasih, penulis mengucapkan terimakasih sebesar-besarnya dan penghargaan yang setinggi-tingginya penulis sampaikan terimakasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., 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. Hisyam Fahmi, M.Kom., selaku dosen pembimbing I yang telah memberikan berbagai pengetahuan, pengalaman, arahan, nasihat, serta motivasi kepada penulis.
5. Juhari, M.Si., selaku dosen pembimbing II yang telah memberikan ilmu, nasihat, bimbingan, pengalaman, serta motivasi kepada penulis.
6. Dr. Fachrur Rozi, M.Si., selaku ketua penguji dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis.
7. Ria Dhea Layla Nur Karisma, M.Si., selaku dosen penguji 1 dalam ujian skripsi yang telah memberikan saran yang bermanfaat bagi penulis.
8. Seluruh dosen Program Studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
9. Bapak Nazri dan Ibu Elvasari selaku orang tua penulis serta seluruh keluarga yang senantiasa mendoakan, memberikan semangat, dukungan, nasihat, serta kasih sayang sehingga penulis dapat menyelesaikan tugas akhir.

10. Rekan penulis yang berjuang bersama dalam mengerjakan skripsi yang senantiasa memberikan dukungan dan semangat kepada penulis.

11. Seluruh teman-teman MAHATMA Program Studi Matematika angkatan 2020 yang senantiasa memberikan bantuan, dukungan dan semangat kepada penulis dalam berbagai kondisi.

Semoga Allah *subhanahu wa ta'ala* selalu memberikan balasan atas segala bantuan dan kebaikan yang telah diberikan kepada penulis. Penulis berharap agar laporan ini dapat bermanfaat bagi penulis serta pembaca untuk menambah wawasan keilmuan yang selalu berkembang.

*Wassalamu 'alaikum Warahmatullahi Wabarakatuh*

Malang, 30 Mei 2024

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>HALAMAN PENGESAHAN</b> .....	<b>iv</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>xviii</b>
<b>BAB I PENDAHULUAN</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 .....	6
<b>BAB II KAJIAN TEORI</b> .....	<b>7</b>
2.1 Teori Pendukung.....	7
2.1.1 <i>Data Mining</i> .....	7
2.1.2 <i>Exploratory Data Analysis (EDA)</i> .....	9
2.1.3 <i>Support Vector Machine (SVM)</i> .....	10
2.1.3.1 SVM Linier .....	10
2.1.3.2 SVM Non-linier .....	15
2.1.4 <i>Confusion Matrix</i> .....	17
2.1.5 <i>K-Fold Cross Validation</i> .....	19
2.1.6 Hipertensi.....	20
2.2 Kajian Integrasi Penyakit Hipertensi dengan Al-Qur'an.....	22
2.3 Kajian Hipertensi dengan Metode <i>Support Vector Machine</i> .....	24
<b>BAB III METODE PENELITIAN</b> .....	<b>25</b>
3.1 Jenis Penelitian .....	25
3.2 Data dan Sumber Data .....	25
3.3 Tahapan Penelitian .....	26
3.3.1 Tahapan Klasifikasi Diagnosis Penyakit Hipertensi Menggunakan SVM.....	26
3.3.2 Tahapan Perhitungan Nilai Akurasi Metode SVM Pada Pengklasifikasian Potensi Penyakit Hipertensi.....	27

<b>BAB IV PEMBAHASAN.....</b>	<b>29</b>
4.1 <i>Exploratory Data Analysis (EDA)</i> .....	29
4.2 <i>Data Preprocessing</i> .....	33
4.2.1 <i>Data Cleaning</i> .....	33
4.2.2 <i>Data Transformation</i> .....	34
4.2.3 <i>Oversampling Data</i> .....	35
4.3 Klasifikasi dengan Metode SVM .....	37
4.3.1 Perhitungan Manual dalam Pengklasifikasian Menggunakan Metode SVM.....	37
4.3.2 Hasil Klasifikasi Menggunakan Metode SVM.....	40
4.4 Evaluasi .....	43
4.5 Analisis Performa Model Klasifikasi .....	44
4.6 Kajian Keislaman dengan Hasil Penelitian .....	45
<b>BAB V KESIMPULAN .....</b>	<b>47</b>
5.1 Kesimpulan.....	47
5.2 Saran .....	48
<b>DAFTAR PUSTAKA .....</b>	<b>49</b>
<b>LAMPIRAN</b>	
<b>RIWAYAT HIDUP</b>	

## DAFTAR TABEL

Tabel 2.1	Contoh Tabel <i>Confusion Matrix</i> .....	18
Tabel 2.2	Rumus Klasifikasi <i>Confusion Matrix</i> .....	19
Tabel 2.3	Kategori Tekanan Darah.....	21
Tabel 3.1	Variabel Penelitian .....	25
Tabel 4.1	Statistik Deskriptif Pada Atribut Numerik .....	32
Tabel 4.2	Tampilan Data .....	38
Tabel 4.3	Tampilan Perubahan Atribut Data.....	38
Tabel 4.4	Korelasi Antar Atribut .....	37
Tabel 4.5	Contoh Data Untuk Perhitungan Manual .....	38
Tabel 4.6	Nilai <i>Weight</i> .....	39
Tabel 4.7	Nilai Akurasi Parameter Model SVM Kernel Linier .....	41
Tabel 4.8	Nilai Akurasi Parameter Model SVM Kernel RBF.....	41
Tabel 4.9	Hasil Evaluasi Menggunakan <i>5-Fold Cross Validation</i> .....	44
Tabel 4.10	<i>Confusion Matrix</i> Hasil Penelitian .....	44

## DAFTAR GAMBAR

Gambar 2.1	Konsep <i>Hyperplane</i> .....	11
Gambar 2.2	Konsep <i>Soft Margin</i> .....	14
Gambar 2.3	Proses Transformasi Data ke Ruang Berdimensi Tinggi .....	15
Gambar 2.4	Skema Pengujian <i>K-Fold Cross Validation</i> .....	20
Gambar 3.1	<i>Flowchart</i> Alur Penelitian .....	28
Gambar 4.1	<i>Pie Chart</i> Diagnosis Hipertensi.....	29
Gambar 4.2	Frekuensi Hubungan Antara Atribut-Atribut Kategorikal .....	30
Gambar 4.3	Korelasi Antar Atribut .....	36
Gambar 4.4	Tampilan Data Sebelum Dan Sesudah <i>Oversampling</i> .....	36
Gambar 4.5	Tampilan Data Sebelum Dan Sesudah <i>Oversampling</i> Kedua .....	37
Gambar 4.6	Plot Klasifikasi SVM Kernel RBF dengan $\gamma = 0,1$ dan $C = 100$ .....	42

## DAFTAR SIMBOL

$x_i$	: <i>Dataset</i>
$y_i$	: Label kelas
$w$	: Vektor pembobot
$b$	: Bias
$N$	: Banyaknya data
$\alpha_i$	: Nilai bobot pada data
$\xi_i$	: Variabel <i>slack</i>
$K(x_i, x_j)$	: <i>Kernel trick</i>
$\phi$	: Parameter fungsi kernel <i>phi</i>
$h$	: Parameter fungsi kernel <i>h</i>
$k$	: Parameter fungsi kernel <i>k</i>
$\delta$	: Parameter fungsi kernel <i>delta</i>
$\gamma$	: Parameter fungsi kernel <i>gamma</i>
$C$	: Parameter fungsi kernel <i>cost</i>
$sign$	: Fungsi tanda atau fungsi <i>signum</i>

## DAFTAR LAMPIRAN

Lampiran 1	Hasil Prediksi Klasifikasi Diagnosis Hipertensi Menggunakan SVM.....	52
Lampiran 2	<i>Script Code Python</i> .....	62

## ABSTRAK

Putri, Hilda Zaqya Elnaz. 2024. **Implementasi Metode *Support Vector Machine* pada Klasifikasi Diagnosis Penyakit Hipertensi**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Juhari, M.Si.

**Kata Kunci:** Hipertensi, *Support Vector Machine*, SVM, *Data Mining*, Klasifikasi.

Hipertensi merupakan salah satu penyakit yang menjadi penyebab utama kematian di dunia. Penyakit ini sering kali disebut *the silent killer* karena penyakit ini dapat menyebabkan kematian tanpa disadari yang mana orang yang mengidap hipertensi tidak memiliki keluhan. Oleh karena itu, deteksi dan penanganan dini hipertensi adalah hal yang sangat penting. Penelitian ini bertujuan untuk mendapatkan klasifikasi penyakit hipertensi menggunakan metode *Support Vector Machine* (SVM) dengan menggunakan berbagai atribut seperti usia, kebiasaan merokok, gaya hidup, tekanan darah, dan diagnosis hipertensi serta mengetahui tingkat akurasi dari hasil klasifikasi penyakit hipertensi dengan metode SVM. Metode SVM dilatih dengan berbagai parameter kernel dan *hyperparameter* untuk menemukan model terbaik. Hasil penelitian menunjukkan bahwa model terbaik untuk klasifikasi penyakit hipertensi dengan metode SVM menggunakan kernel RBF dengan parameter  $C = 100$  dan  $\gamma$  (*gamma*) = 0,1 dengan akurasi sebesar 97,15%. Hal ini menunjukkan bahwa metode SVM mampu mengklasifikasikan penyakit hipertensi dengan sangat baik serta mampu memberikan kontribusi yang signifikan dalam deteksi dini dan penanganan penyakit hipertensi.

## ABSTRACT

Putri, Hilda Zaqya Elnaz. 2024. **On Implementation of Support Vector Machine Method in the Classification of Hypertension Diagnosis**. Thesis. Mathematics Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Juhari, M.Si.

**Keywords:** Hypertension, Support Vector Machine, SVM, Data Mining, Classification.

Hypertension is one of the leading causes of death worldwide. This disease is often referred to as the silent killer because it can lead to death without noticeable symptoms, leaving those affected unaware of their condition. Therefore, early detection and management of hypertension are crucial. This research aims to obtain the classification of hypertension using the Support Vector Machine (SVM) method by utilizing various attributes such as age, smoking habits, lifestyle, blood pressure, and hypertension diagnosis, as well as determining the accuracy level of hypertension classification results using the SVM method. The SVM method is trained with various kernel parameters and hyperparameters to find the best model. The research findings indicate that the best model for classifying hypertension using the SVM method employs the RBF kernel with parameters  $C = 100$  and  $\gamma$  (*gamma*) = 0,1, achieving an accuracy of 97.15%. This demonstrates that the SVM method is capable of classifying hypertension very well and significantly contributes to the early detection and management of hypertension.

## مستخلص البحث

بوتري، هيلدا زقيا الناز. ٢٠٢٤. تطبيق طريقة آلة الدعم الموجه في تصنيف تشخيص ارتفاع ضغط الدم. البحث الجامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (١) هشام فهمي، الماجستير. (٢) جوهاري، الماجستير.

**الكلمة المفتاحية:** إرتفاع ضغط الدم، آلة ناقلات الدم، التنقيب عن البيانات، التصنيف.

ارتفاع ضغط الدم هو أحد الأسباب الرئيسية للوفاة في العالم. وغالباً ما يطلق على هذا المرض اسم "القاتل الصامت" لأنه يمكن أن يتسبب الوفاة دون علم الشخص المصاب بارتفاع ضغط الدم عندما لا يشكو من ارتفاع ضغط الدم. لذلك، فإن الكشف عن ارتفاع ضغط الدم وعلاجه امر مهم للغاية. يهدف هذا البحث إلى الحصول على تصنيف للأشخاص المصابين بارتفاع ضغط الدم باستخدام طريقة آلة المتجهات (SVM) باستخدام سمات مختلفة مثل العمر، وعادة التدخين، ونمط الحياة، وضغط الدم، وتشخيص ارتفاع ضغط الدم لتحديد دقة نتائج تصنيف مرض ارتفاع ضغط الدم باستخدام طريقة SVM. تم تدريب طريقة SVM باستخدام معلمات نواة مختلفة والمعلمات فائقة للعثور على أفضل نموذج. أظهرت النتائج أن أفضل نموذج لتصنيف البذور الزيتية باستخدام طريقة SVM استخدم نواة RBF مع المعلمة  $C = 100$  و  $\gamma = 0,1$  (gamma) بدقة  $97,15\%$  وهذا يدل على أن طريقة SVM قادرة على تصنيف المصابين بارتفاع ضغط الدم بشكل جيد للغاية، و يمكن أن يسهم بشكل كبير في الكشف المبكر عن المصابين بارتفاع ضغط الدم وعلاجهم.

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Kemajuan teknologi memiliki peranan yang penting dalam hampir seluruh aspek kehidupan. Pertumbuhan yang pesat ini tak terlepas dari adanya pengaruh dari komputer. Komputer telah mengubah segala hal menjadi lebih efisien dan lebih cepat melalui sistemnya. Salah satu aspek penting dari pengolahan informasi ini adalah *data mining*, dimana komputer menggunakan teknik-teknik khusus untuk menggali wawasan berharga dari kumpulan data besar menjadi informasi yang berharga. *Data mining* dapat digunakan untuk mengekstrak informasi dari data yang berjumlah besar menjadi bentuk yang berguna dan dapat dipahami (Septiani, 2017).

*Data mining* bertujuan untuk mengidentifikasi pola yang dapat membantu dalam pengambilan keputusan, klasifikasi, dan prediksi. Klasifikasi merupakan teknik pengelompokan *data mining* ke dalam kelas-kelas tertentu atau kategori berdasarkan karakteristiknya (Setiawati dkk., 2019). Salah satu algoritma/metode yang terkenal dalam melakukan klasifikasi adalah *Support Vector Machine* (SVM). Metode SVM bertujuan untuk mencari *hyperplane* atau fungsi pemisah (*decision boundary*) terbaik yang dapat memisahkan dua atau lebih kelas data pada ruang *input*. Dalam melakukan klasifikasi, SVM memiliki konsep yang jelas secara sistematis dan konsisten jika dibandingkan dengan algoritma-algoritma klasifikasi lainnya. Metode SVM memiliki keunggulan, yaitu dapat menghasilkan nilai akurasi yang baik meskipun data yang dianalisis bersifat tidak seimbang (*unbalanced data*), dapat mengurangi kesalahan klasifikasi untuk sampel yang tidak diketahui dan

meningkatkan kemampuan generalisasi dibandingkan dengan metode JST (Jaringan Syaraf Tiruan) dan kepekaan terhadap *outlier* (Maiyanti dkk., 2018).

*Data mining* dapat digunakan untuk melakukan klasifikasi dan prediksi suatu penyakit. Terdapat beberapa penelitian terdahulu yang menggunakan *data mining* dalam mengklasifikasikan atau memprediksi suatu penyakit. Pertama, penelitian yang dilakukan oleh Andriansyah dkk. (2018), yaitu klasifikasi hipertensi menggunakan metode *Fuzzy Decision Tree Iterative Dichotomiser 3* (ID3). Penelitian tersebut menggunakan 12 atribut, yaitu umur, tekanan darah, lingkaran perut, tinggi badan, berat badan, berat massa indeks, merokok, konsumsi gula, konsumsi garam, olahraga dan konsumsi kafein dan diagnosis hipertensi. Dari penelitian tersebut dihasilkan akurasi sebesar 75,56%. Kedua, penelitian yang dilakukan oleh Tino dkk. (2023), yaitu membandingkan algoritma SVM dan *Neural Network* untuk mengklasifikasikan penyakit jantung. Dari percobaan terhadap kedua metode tersebut, metode SVM menghasilkan nilai akurasi 83% dan metode *Neural Network* menghasilkan nilai akurasi 82%. Hal ini mengindikasikan bahwa metode SVM memiliki model yang lebih akurat daripada *Neural Network*. Maka dari itu, dapat disimpulkan bahwa metode *Support Vector Machine* mempunyai kemampuan yang baik dalam melakukan suatu klasifikasi atau prediksi penyakit. Salah satunya untuk pengklasifikasian diagnosis penyakit hipertensi.

Hipertensi adalah salah satu penyakit yang menjadi penyebab utama kematian di dunia. Penyakit ini sering kali disebut *the silent killer* karena penyakit ini dapat menyebabkan kematian tanpa disadari yang mana orang yang mengidap hipertensi tidak memiliki keluhan. Hipertensi adalah sebuah istilah medis dari penyakit tekanan darah tinggi. Umumnya, hipertensi merupakan kondisi

asimtomatik (keadaan tanpa gejala) dimana tekanan di arteri tinggi secara tidak normal yang berpotensi dapat meningkatkan peluang terjadinya stroke, aneurisma, gagal jantung, serangan jantung, kerusakan ginjal, dan kematian. (Khalim dkk., 2023).

Berdasarkan data Riset Kesehatan Dasar 2018, tingkat kematian masyarakat Indonesia akibat hipertensi sebanyak 427.218 kematian (P2TPM Kemenkes RI, 2019). Lalu berdasarkan data Badan Pusat Statistik Kota Malang, hipertensi menduduki peringkat kedua sebagai salah satu penyakit dengan jumlah kasus terbanyak di Kota Malang pada tahun 2019 yaitu sebanyak 13.102 kasus (BPS Kota Malang, 2019). Oleh karena itu, deteksi dan penanganan dini hipertensi adalah hal yang sangat penting. Kemajuan teknologi saat ini telah memudahkan pendiagnosaan penyakit, sehingga penanganan dapat dilakukan lebih cepat dan peluang sembuh pun meningkat. Salah satu cara untuk meningkatkan akurasi dan efisiensi para ahli medis dalam mengklasifikasikan suatu penyakit adalah dengan mempertimbangkan sistem klasifikasi yang dikembangkan menggunakan *data mining*. Hal ini juga memberikan pengingat yang kuat tentang pentingnya menjaga kesehatan sebagai langkah awal untuk mencegah dan mengatasi berbagai penyakit, sehingga meningkatkan peluang hidup sehat bagi masyarakat secara keseluruhan.

Dalam Islam, kesehatan merupakan salah satu hal yang sangat penting. Kesehatan adalah anugerah Allah yang harus disyukuri oleh setiap manusia Seperti yang tertuang pada hadits riwayat Bukhari sebagai berikut:

نِعْمَتَانِ مَغْبُورٌ فِيهِمَا كَثِيرٌ مِنَ النَّاسِ الصِّحَّةُ وَالْفَرَاغُ

“Dua kenikmatan yang sering dilupakan oleh kebanyakan manusia adalah kesehatan dan waktu luang”.

Dalam hadits tersebut, dijelaskan bahwa kesehatan merupakan hal yang penting. Cara menjaga kesehatan juga melibatkan pengembangan pola hidup yang baik untuk menghindari risiko penyakit, terutama penyakit seperti hipertensi. Apabila seseorang tidak memiliki pola hidup yang baik, maka lebih rentan terkena penyakit hipertensi.

Berdasarkan yang telah dipaparkan di atas, maka penulis memilih metode *Support Vector Machine* (SVM) untuk mengklasifikasikan diagnosis penyakit hipertensi. Penelitian ini akan memanfaatkan *tools Jupyter Notebook* dan bahasa pemrograman *python* sebagai alat bantu penelitian. Penelitian ini diharapkan dapat membantu ahli kesehatan dalam menentukan diagnosis penyakit hipertensi.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang, maka rumusan masalah yang diambil dalam penelitian ini sebagai berikut:

1. Bagaimana hasil klasifikasi diagnosis penyakit hipertensi menggunakan metode SVM?
2. Bagaimana tingkat akurasi yang diperoleh dari metode SVM pada pengklasifikasian diagnosis penyakit hipertensi?

## **1.3 Tujuan Penelitian**

Berdasarkan uraian rumusan masalah, maka tujuan dari penelitian ini sebagai berikut:

1. Mengetahui hasil klasifikasi diagnosis penyakit hipertensi menggunakan metode SVM.
2. Mengetahui tingkat akurasi dari hasil klasifikasi diagnosis penyakit hipertensi menggunakan metode SVM.

#### **1.4 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Manfaat Teoritis
  - a. Mampu mengimplementasikan metode SVM dalam mengklasifikasikan diagnosis penyakit hipertensi.
  - b. Mampu mengetahui proses dan kinerja SVM dalam mengklasifikasikan diagnosis penyakit hipertensi.
2. Manfaat Praktis
  - a. Bagi Program Studi, penelitian ini dapat digunakan sebagai wawasan ilmu pengetahuan dan bahan pembelajaran bagi mahasiswa terkait SVM.
  - b. Bagi instansi, penelitian ini dapat digunakan sebagai dasar untuk pengembangan Sistem Pendukung Keputusan (SPK) diagnosis penyakit hipertensi.
  - c. Bagi pembaca, penelitian ini dapat memberikan informasi dan referensi terkait implementasi metode SVM dalam mengklasifikasikan diagnosis penyakit hipertensi.

#### **1.5 Batasan Masalah**

Batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut:

1. Pengumpulan data hipertensi dianggap sebagai proses pengecekan sekali untuk setiap entitas data yang terkait.
2. Atribut usia pada data hipertensi adalah usia pasien pada saat pemeriksaan di Puskesmas Mojolangu Kota Malang.

3. Label pada data hipertensi yang mana berupa tidak berpenyakit hipertensi dan berpenyakit hipertensi diperoleh dari *file* rekam medis Puskesmas Mojolangu Kota Malang.

## 1.6 Definisi Istilah

Terdapat beberapa istilah yang digunakan dalam penelitian ini, yaitu:

- Machine learning* : Ilmu yang mempelajari cara membuat sistem komputer dapat belajar dari data
- Dataset* : Kumpulan data yang berisi objek dan atribut
- Support vector* : Titik data dalam *dataset* yang terletak paling dekat dengan *hyperplane*.
- Hyperplane* : Sebuah bidang datar yang memisahkan antar kelas.
- Margin* : Jarak paling dekat *hyperplane* dengan titik pada tiap kelas.
- Kernel : Fungsi yang digunakan untuk memetakan data dari ruang dimensi rendah ke ruang dimensi tinggi.
- Data training* : Kumpulan data yang digunakan untuk melatih model algoritma.
- Data testing* : Kumpulan data yang digunakan untuk mengevaluasi kinerja model yang telah dilatih.
- Arteri : Pembuluh darah yang berfungsi untuk mengalirkan darah dari jantung ke seluruh tubuh.
- Hipertensi : Kondisi dimana tekanan darah di arteri terlalu tinggi

## BAB II KAJIAN TEORI

### 2.1 Teori Pendukung

#### 2.1.1 *Data Mining*

*Data mining* adalah proses mengekstrak informasi berharga dari data dalam jumlah besar menggunakan metode analisis data tradisional dan algoritma canggih (Tan dkk., 2016). *Data mining* digunakan untuk mengekstrak informasi penting dari data, dan informasi tersebut dapat digunakan untuk memecahkan masalah dan membuat keputusan yang lebih baik. *Data mining* adalah bidang multi-disiplin yang didalamnya mencakup statistik, kecerdasan buatan, *machine learning*, *database*, dan visualisasi data. Penggunaan *data mining* memiliki dampak dalam berbagai aspek, misalnya di bidang kesehatan (Aris & Benyamin, 2019). Pada bidang kesehatan, penggunaan *data mining* dapat membantu memprediksi sebuah penyakit berdasarkan rekam medis pasien yang tersimpan di dalam *database* dan membantu pengelolaan data medis menjadi lebih efisien.

*Data mining* juga dikenal sebagai *Knowledge Discovery in Database* (KDD) karena kemampuannya memperoleh informasi atau pola yang berguna dari kumpulan data mentah, seperti gambar, *website*, teks, *database*, dan sebagainya. Tugas *data mining* yang paling umum adalah klasifikasi, pengelompokan (*clustering*), penemuan pola berurutan, dan penemuan aturan asosiasi (Aris & Benyamin, 2019). Selain itu, terdapat beberapa jenis data yang termasuk *data mining* yaitu sebagai berikut:

1. *Supervised Learning*, yaitu data yang memiliki kelas atau label.
2. *Unsupervised Learning*, yaitu data yang tidak memiliki kelas atau label.

3. *Semi Supervised*, yaitu perpaduan antara data *unsupervised* dan *supervised*.

*Data mining* dapat digunakan untuk mengekstrak informasi dari kumpulan data mentah yang besar dan tidak terstruktur. Algoritma inti *data mining* memungkinkan seseorang untuk menemukan pola dan hubungan yang tersembunyi dalam data, yang dapat digunakan untuk memperoleh wawasan dan pengetahuan. Rangkaian proses *data mining* / KDD yaitu (Tan dkk., 2016):

1. *Data Selection*

*Data Selection* (Seleksi Data) adalah proses memilih data yang relevan untuk digunakan dalam proses *data mining*. Data yang telah dipilih kemudian dipisahkan dari *dataset* operasional.

2. *Data Preprocessing*

*Data Preprocessing* adalah proses mengubah data mentah menjadi data yang siap digunakan dalam proses *data mining*. Beberapa hal yang dapat dilakukan dalam *data preprocessing* yaitu menghilangkan data yang tidak relevan (*data cleaning*) dan mengurangi ukuran data (*data reduction*).

3. *Data Transformation*

*Data Transformation* adalah proses mengubah bentuk data mentah menjadi data yang siap digunakan. Proses ini dilakukan untuk memastikan bahwa data yang digunakan dalam proses *data mining* adalah data yang valid dan dapat diproses oleh algoritma *data mining*.

4. *Data Mining*

*Data mining* adalah proses mengekstrak informasi dari data. Proses ini menggunakan berbagai metode dan teknik untuk menemukan pola dan hubungan dalam data.

## 5. *Evaluation*

*Evaluation/Knowledge Interpretation* yaitu proses mengidentifikasi pola-pola menarik yang telah ditemukan sebelumnya berdasarkan kriteria yang telah ditetapkan. Dalam hal ini digunakan teknik visualisasi data agar informasi dapat lebih mudah dimengerti.

### 2.1.2 *Exploratory Data Analysis (EDA)*

*Exploratory Data Analysis (EDA)* atau dikenal sebagai analisis data eksploratif adalah sebuah metode (cara) untuk menjelajahi data menggunakan teknik aritmatika dan visualisasi grafis agar data dapat lebih mudah dipahami. Tujuan dari EDA adalah untuk mencari tahu informasi/*insight* dari data lebih dalam, menemukan pola, menguji hipotesis dengan bantuan statistik ya kemudian akan dibuat representasi grafis menggunakan *plot* sebagai bentuk visualisasi data (Nawangsari, 2017).

Adapun tahapan-tahapan dalam EDA menurut Nawangsari (2017) yaitu:

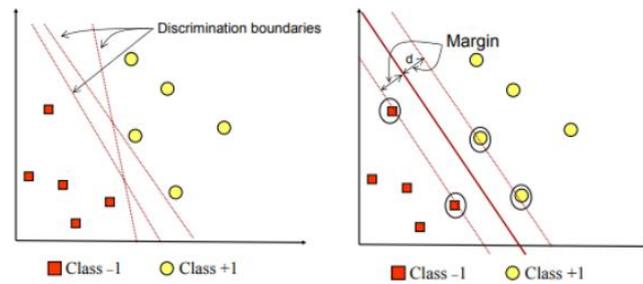
1. Mengobservasi terhadap *dataset* yang ada serta melakukan analisis statistika deskriptif pada *dataset*.
2. Menemukan data yang memiliki *missing value* dan melakukan perbaikan jika diperlukan.
3. Menggolongkan data kedalam kategori, seperti data kategorikal dan numerikal.
4. Mengenali keterkaitan antara variabel-variabel.
5. Mengidentifikasi *outlier*, *skewness* data dan menerapkan statistik deskriptif atau inferensial.

### 2.1.3 *Support Vector Machine (SVM)*

SVM adalah metode pembelajaran mesin yang dapat digunakan untuk melakukan klasifikasi dan regresi. Untuk klasifikasi, metode ini dikenal dengan nama *Support Vector Classification (SVC)*, sedangkan untuk regresi dikenal dengan nama *Support Vector Regression (SVR)*. SVM bekerja untuk mencari *hyperplane* atau fungsi pemisah (*decision boundary*) terbaik yang memisahkan dua buah kelas. Penentuan *hyperplane* ini berdasarkan *support vectors*, yaitu vektor data berjarak paling mendekati *hyperplane* (Maiyanti dkk., 2018). Bentuk *hyperplane* sendiri dapat berupa garis (*line*) pada dua dimensi dan dapat berupa bidang datar (*flat plane*) pada dimensi yang lebih tinggi. SVM dapat digunakan untuk melakukan klasifikasi dan regresi, baik untuk data linier maupun data non linier. Penjelasannya yaitu sebagai berikut:

#### 2.1.3.1 SVM Linier

SVM Linier merupakan salah satu metode SVM yang datanya dapat dipisahkan oleh *hyperlane* secara linier, dimana jika sebuah *dataset* dapat diklasifikasikan menjadi dua kelas dengan menggunakan sebuah garis lurus tunggal. Misalkan  $x_i = \{x_1, x_2, \dots, x_n\}$ ,  $x_i \in R^n$  merupakan *dataset*. Setelah itu, akan dibagi dua kelas di mana untuk kelas positif dinotasikan dengan 1 dan kelas negatif dinotasikan dengan -1. Maka label kelas dinotasikan sebagai  $y_i \in \{+1, -1\}$  dengan  $i = 1, 2, 3, \dots, n$  dimana  $n$  menunjukkan banyak data. Berikut ditunjukkan ilustrasi cara kerja SVM pada Gambar 2.1.



**Gambar 2.1** Konsep *Hyperplane* (Tan dkk., 2016)

Persamaan *hyperplane* yang memisahkan dua kelas secara umum didefinisikan sebagai (Tan dkk., 2016):

$$w \cdot x + b = 0 \quad (2.1)$$

dimana  $w$  merupakan vektor pembobot dan  $b$  merupakan bias. Syarat yang dimiliki oleh *hyperplane* yaitu:

1. Memiliki nilai eror terkecil dalam pengklasifikasian suatu data.
2. Memiliki jarak terdekat dari masing-masing kelas (margin) yang maksimum. Margin maksimum adalah pilihan yang paling aman karena terjadi paling sedikit kesalahan pada klasifikasi.

Oleh karena itu, data untuk masing-masing kelas memiliki nilai  $y = +1$  untuk kelas positif didefinisikan pada Persamaan (2.2) dan  $y = -1$  untuk kelas negatif didefinisikan pada Persamaan (2.3) yaitu (Tan dkk., 2016):

$$w_i x_i + b \geq +1, y_i = +1 \quad (2.2)$$

$$w_i x_i + b \leq -1, y_i = -1 \quad (2.3)$$

Margin sebuah batas keputusan ditentukan oleh jarak antara dua *hyperplane*. Untuk menghitung margin, misalkan  $x_1$  menjadi titik data yang terletak di  $b_1$  dan  $x_2$  menjadi titik data di  $b_2$ , seperti yang ditunjukkan pada Gambar 2.1. Titik-titik ini disubstitusikan ke Persamaan (2.2) dan (2.3). Margin

$d$  dapat dihitung dengan cara mengurangkan persamaan kedua dengan persamaan pertama. Wilayah generalisasi untuk *hyperplane* dapat berkisar antara kelas 1 dan -1 sehingga terdapat banyak margin yang dapat dianggap batas setiap kelas. Oleh karena itu, untuk mendapatkan *hyperplane* terbaik, jarak ( $d$ ) antar margin harus diukur dan dimaksimalkan menggunakan Persamaan (2.4).

$$d(w, b; x) = \frac{|(w_i x_i + b - 1) - (w_i x_i + b + 1)|}{|w|} = \frac{2}{|w|} \quad (2.4)$$

Memaksimalkan margin dalam SVM berarti mencari vektor  $w$  berdimensi terkecil dalam sebuah masalah optimasi konveks untuk menemukan *hyperplane* optimal. Hal ini dapat diformulasikan sebagai masalah pemrograman kuadrat (*Quadratic Programming*) menggunakan Persamaan (2.5) dengan syarat Persamaan (2.6).

$$\min_{w,b} = \frac{1}{2} \|w\|^2 \quad (2.5)$$

$$\text{dengan} \quad y_i(w_i x_i + b) \geq 1 \quad (2.6)$$

Karena fungsi tujuan adalah kuadrat dan kendalanya bersifat linier dalam parameter-parameter  $w$  dan  $b$  ini disebut sebagai masalah optimasi konveks. Hal ini dapat diselesaikan dengan menggunakan metode *Lagrange Multiplier* ( $\alpha$ ) pada Persamaan (2.7).

$$L_P(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i [y_i(w_i \cdot x_i + b) - 1] \quad (2.7)$$

$\alpha_i \neq 0$  jika dan hanya jika  $x_i$  adalah *support vector*. *Support vector* adalah objek data terluar yang paling dekat dengan *hyperplane*. Selanjutnya, optimasi dilakukan dengan melakukan turunan parsial  $L$  terhadap  $w$  dan  $b$  dan

dimaksimalkan terhadap  $\alpha_i > 0$  untuk memenuhi kondisi Karush-Kuhn-Tucker (KKT) pada Persamaan (2.8) dan (2.9) (Gholami dan Fakhari, 2017):

$$\frac{\partial L_p}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i x_i y_i \quad (2.8)$$

$$\frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \quad (2.9)$$

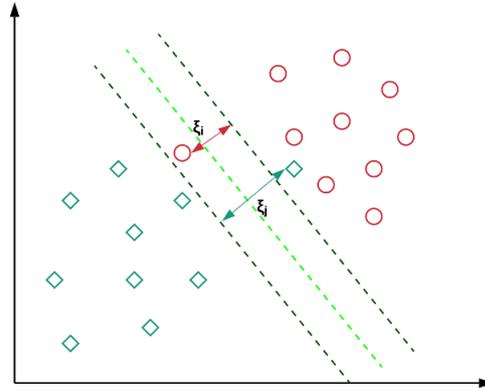
Selain itu, proses optimasi dapat dilakukan dengan memaksimalkan  $L$  terhadap  $\alpha_i$  dengan cara mensubstitusikan Persamaan (2.8) dan Persamaan (2.9) ke dalam Persamaan (2.7) yaitu sebagai berikut.

$$\begin{aligned} L(\alpha) &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i [y_i (w_i x_i + b) - 1] \\ L(\alpha) &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \alpha_i y_i (w_i x_i + b) - \sum_{i=1}^N \alpha_i \\ L(\alpha) &= \frac{1}{2} (w_i \cdot w_j) - \left( \sum_{i=1}^N \alpha_i y_i w_i x_i + \sum_{i=1}^N \alpha_i y_i b - \sum_{i=1}^N \alpha_i \right) \\ L(\alpha) &= \frac{1}{2} \left( \sum_{i=1}^N \alpha_i x_i y_i \cdot \sum_{i=1}^N \alpha_j x_j y_j \right) + \\ &\quad - \left( \sum_{i=1}^N \alpha_i y_i x_i \cdot \sum_{i=1}^N \alpha_j y_j x_j + 0 - \sum_{i=1}^N \alpha_i \right) \\ L(\alpha) &= \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j x_i x_j y_i y_j - \left( \sum_{i,j=1}^N \alpha_i \alpha_j x_i x_j y_i y_j - \sum_{i=1}^N \alpha_i \right) \\ L(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j x_i x_j y_i y_j \end{aligned} \quad (2.10)$$

dengan  $\alpha_i \geq 0$  dan  $\sum_{i,j=1}^N \alpha_i y_i = 0$ .

Secara umum, sebuah *hyperplane* tidak mampu memisahkan dua kelas secara sempurna. Hal ini mengakibatkan *constraint* pada Persamaan (2.6) tidak

terpenuhi, sehingga optimisasi tidak dapat terpenuhi dilakukan. Untuk menyelesaikan masalah ini, SVM dirumuskan ulang menggunakan pendekatan *soft margin*. Konsep *soft margin* dijelaskan dengan ilustrasi dalam Gambar 2.2.



**Gambar 2.2** Konsep *Soft Margin* (Misra, 2019)

Masalah tersebut dapat ditangani dengan cara memodifikasi Persamaan (2.6) menggunakan *soft margin* dengan cara memasukkan variabel *slack*  $\xi_i > 0$  sebagai berikut:

$$y_i(w_i x_i + b) \geq 1 - \xi_i$$

sehingga diperoleh Persamaan (2.11).

$$\min_{w,b} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (2.11)$$

Parameter  $C$  digunakan untuk mengontrol *error* klasifikasi  $\xi$  dan margin. Parameter  $C$  ditentukan dengan mencoba beberapa nilai dan memilih nilai yang menghasilkan akurasi terbaik saat diuji dengan *cross validation*. Nilai  $C$  yang besar akan memberikan penalti yang lebih besar terhadap kesalahan klasifikasi, sehingga *hyperplane* akan dicari dengan lebih cermat untuk menghindari kesalahan klasifikasi.

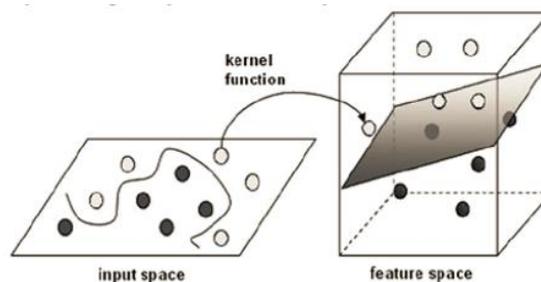
### 2.1.3.2 SVM Non-linier

SVM non-linier digunakan untuk data yang tidak dapat dipisahkan secara linier, yang mana sebuah *dataset* tidak dapat diklasifikasi menggunakan garis, maka data tersebut disebut data non-linier dan klasifikasi yang digunakan disebut sebagai *Non-linier SVM classifier*.

Prinsip dasar ini menggunakan metode *kernel trick* pada fitur berdimensi tinggi. Hasil akurasi data yang dihasilkan algoritma SVM bergantung pada fungsi *kernel* dan parameter yang digunakan. Fungsi kernel adalah fungsi yang digunakan untuk mengubah ruang input data dari dimensi rendah ke dalam ruang berdimensi tinggi (*feature space*), di mana data dapat dipisahkan secara linier. Fungsi kernel didefinisikan sebagai berikut (Gholami dan Fakhari, 2017) pada Persamaan (2.12).

$$\phi(x_i, x_j) = K(x_i, x_j) \quad (2.12)$$

dengan  $\phi$  adalah fungsi transformasi dari ruang input ke ruang fitur. Penggambaran transformasi data dari ruang *input* ke ruang berdimensi tinggi (*feature space*) dapat digambarkan sebagai berikut.



**Gambar 2.3** Proses Transformasi Data ke Ruang Berdimensi Tinggi (Tino, 2023)

Gambar 2.3 menunjukkan bahwa data kelas positif dan kelas negatif yang berdimensi dua tidak dapat dipisahkan secara sempurna oleh *hyperplane* secara

linier. Namun, setelah dilakukan transformasi data ke dalam ruang berdimensi tiga, data kelas positif dan kelas negatif dapat dipisahkan oleh *hyperplane* secara linier. Berdasarkan Persamaan (2.9), Persamaan umum untuk kasus non-linier *support vector* dapat ditulis sebagai berikut (Tan dkk., 2016):

$$L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N y_i y_j \alpha_i \alpha_j \phi(x_i) \cdot \phi(x_j) \quad (2.13)$$

Namun, untuk mendapatkan *hyperplane* yang optimal tidak mudah karena nilai vektor pembobot  $w$  belum diketahui. Untuk menemukan nilai vektor pembobot  $w$  dalam *Hilbert Space* dapat ditulis dalam Persamaan (2.14).

$$w_i = \sum_{i=1}^N \alpha_i y_i \phi(x_i) \quad (2.14)$$

Sedangkan untuk parameter bias  $b$  dapat dihitung menggunakan Persamaan (2.15) yaitu sebagai berikut:

$$b = y_i - \sum_{i,j=1}^N \alpha_i y_i K(x_i, x_j) \quad (2.15)$$

Kemudian, dengan mensubstitusikan Persamaan (2.14) ke Persamaan (2.16), maka diperoleh persamaan umum *hyperplane* sebagai berikut:

$$\begin{aligned} f(\phi(x)) &= \text{sign}(w_i \cdot \phi(x) + b) \\ f(\phi(x)) &= \text{sign} \left( \sum_{i,j=1}^N \alpha_i y_i \phi(x_i) \cdot \phi(x_j) + b \right) \\ f(\phi(x)) &= \text{sign} \left( \sum_{i,j=1}^N \alpha_i y_i K(x_i) \cdot K(x_j) + b \right) \end{aligned} \quad (2.16)$$

dengan  $\alpha_i > 0$ .

Beberapa fungsi kernel yang digunakan pada metode *Support Vector Machine* (SVM) diantaranya yaitu:

a. Kernel Linier

$$K(x_i, x_j) = x_i \cdot x_j \quad (2.17)$$

b. Kernel Polinomial

$$K(x_i, x_j) = (x_i \cdot x_j + 1)^h \quad (2.18)$$

c. Kernel Sigmoid

$$K(x_i, x_j) = \tanh(kx_i \cdot x_j - \delta) \quad (2.19)$$

d. Kernel *Gaussian Radial Basis Function* (RBF)

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \gamma \in \mathbb{R}^+ \quad (2.20)$$

dengan  $h, \gamma, k$ , dan  $\delta$  sebagai parameter kernel.

Fungsi kernel yang digunakan untuk SVM harus dipilih sesuai dengan karakteristik data yang akan diklasifikasikan. Setelah memilih fungsi kernel, perlu ditentukan nilai parameter kernel yang digunakan dan parameter  $C$  untuk menentukan seberapa besar kesalahan klasifikasi yang dapat diterima. Dengan adanya *kernel trick*, maka parameter vektor pembobot  $w$  tidak lagi dibutuhkan sehingga SVM dapat digunakan secara efisien dalam menyelesaikan kasus non-linier dengan memilih fungsi kernel yang sesuai. Pada penelitian ini, akan dilakukan uji coba menggunakan kernel linier dan kernel RBF untuk mengevaluasi akurasi terbaik pada model SVM.

#### 2.1.4 *Confusion Matrix*

*Confusion Matrix* merupakan alat yang digunakan untuk menganalisis kinerja atau menghitung tingkat kebenaran suatu klasifikasi dalam mengenali data dari kelas yang ada (Suyanto, 2019). Hasil klasifikasi menggunakan *confusion matrix* dapat diukur menggunakan 4 komponen, diantaranya yaitu:

1. *True Positive* (TP), artinya data yang bernilai positif diklasifikasikan benar sebagai positif.

2. *True Negative* (TN), artinya data yang bernilai negatif diklasifikasikan benar sebagai negatif.
3. *False Positive* (FP), artinya data yang bernilai negatif tetapi diklasifikasikan sebagai positif.
4. *False Negative* (FN), artinya data yang bernilai positif tetapi diklasifikasikan sebagai negatif.

*Confusion matrix* yang menggambarkan kinerja dari model klasifikasi dapat dilihat pada Tabel 2.1.

**Tabel 2.1** Contoh Tabel *Confusion Matrix*

Data		Aktual	
		<i>True</i>	<i>False</i>
Prediksi	<i>True</i>	<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
	<i>False</i>	<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

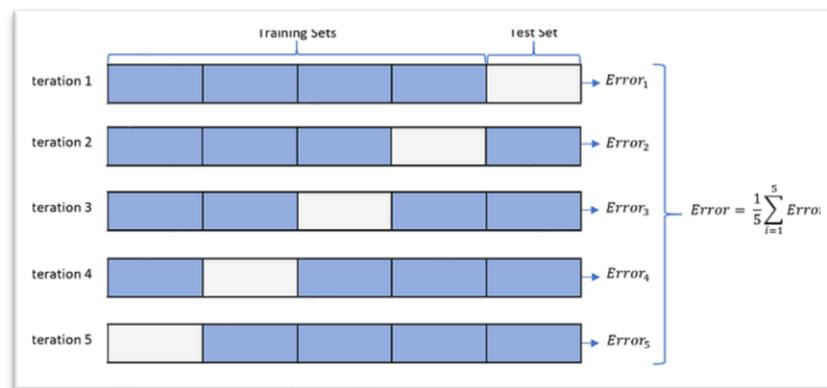
Pengukuran ketepatan sebuah klasifikasi dapat dihitung menggunakan *accuracy*, *precision*, *recall* dan *F1-Score*. *Accuracy* (akurasi) digunakan untuk mengukur sejauh mana model benar dalam memprediksi data, *precision* (presisi) digunakan untuk mengukur presentase data yang benar positif diprediksi sebagai kelas positif. *recall* digunakan untuk mengukur kinerja klasifikasi yang menunjukkan proporsi data positif yang diklasifikasikan dengan benar, dan *F1-Score* digunakan untuk mengukur keseimbangan antar *presisi* dan *recall*. Apabila keempat komponen tersebut memiliki nilai lebih dari 90%, maka sebuah model dapat dianggap memiliki performa yang baik. Lalu terdapat konsep rumus untuk menghitung performa menggunakan *confusion matrix* pada Tabel 2.2 (Tan dkk., 2016).

**Tabel 2.2** Rumus Klasifikasi *Confusion Matrix*

Nilai	Rumus
<i>Accuracy</i>	$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$
<i>Precision</i>	$\frac{TP}{TP + FP} \times 100\%$
<i>Recall</i>	$\frac{TP}{TP + FN} \times 100\%$
<i>F1-Score</i>	$2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$

### 2.1.5 K-Fold Cross Validation

*K-Fold Cross Validation* merupakan salah satu dari jenis *cross validation*. *Cross validation* adalah teknik atau metode untuk memperkirakan kinerja model pada data baru dengan melatih model pada sebagian data dan mengujinya pada sebagian data lainnya. *K-Fold Cross Validation* bekerja dengan cara membagi seluruh data menjadi  $k$  partisi/bagian secara acak ( $P_1, P_2, \dots, P_k$ ). Jika menggunakan  $k = 4$  atau disebut juga *4-fold cross validation*, maka data dibagi menjadi empat bagian yang sama. Pada iterasi pertama, model dilatih pada tiga bagian data pertama dan dievaluasi pada bagian data keempat. Pada iterasi kedua, model dilatih pada tiga bagian data yang berbeda dan dievaluasi pada bagian data keempat yang berbeda. Proses ini diulangi sebanyak 4 kali, sehingga setiap bagian data akan digunakan sebagai data uji sebanyak 1 kali (Peryanto dkk., 2020). Berikut Gambar 2.4 yang berisi cara kerja *K-Fold Cross Validation* (Rebecca Patro, 2021).



**Gambar 2.4** Skema Pengujian *K-Fold Cross Validation* (Rebecca Patro, 2021)

### 2.1.6 Hipertensi

Hipertensi adalah kondisi dimana tekanan di dalam arteri meningkat tinggi secara abnormal. Kondisi ini dapat meningkatkan risiko terhadap berbagai penyakit serius, seperti stroke, gagal jantung, serangan jantung, dan kerusakan ginjal. Adapun gejala terjadinya hipertensi yaitu sesak napas, mimisan, pusing, nyeri dada, gangguan penglihatan, sakit kepala parah, sesak napas, kulit memerah (terutama pada wajah dan leher), gangguan penglihatan, dan adanya darah di dalam urin (Mitra Keluarga, 2022).

Tekanan darah terdiri dari dua angka, yaitu sistolik dan diastolik. Sistolik adalah tekanan darah saat jantung memompa darah, sedangkan diastolik adalah tekanan darah saat jantung beristirahat. Tekanan darah diukur dengan dua angka, yang dipisahkan oleh garis miring. Misalnya 120/70 mmHg. Angka pertama adalah tekanan darah saat jantung memompa darah, sedangkan angka kedua adalah tekanan darah saat jantung beristirahat. Seseorang dinyatakan berpenyakit hipertensi apabila tekanan darah sistolik mencapai 140 mmHg atau lebih, dan tekanan darah diastolik mencapai 90 mmHg atau lebih pada saat pemeriksaan berulang (Awalullaili dkk., 2023). Berikut merupakan kategori

tekanan darah pada orang dewasa menurut P2TPM Kementerian Kesehatan RI pada Tabel 2.3.

**Tabel 2.3** Kategori Tekanan Darah

Kategori Tekanan Darah	Besar Tekanan Darah	
	Sistolik	Diatolik
Normal Dibawah	< 130 mmHg	< 85 mmHg
Normal Tinggi (Stadium 1)	130-139 mmHg	85-89 mmHg
Hipertensi Ringan (Stadium 2)	140-159 mmHg	90-99 mmHg
Hipertensi Sedang (Stadium 3)	160-179 mmHg	100-109 mmHg
Hipertensi Berat (Stadium 4)	180-209 mmHg	110-119 mmHg
Hipertensi Maligna (Tingkat Keparahan yang Tinggi)	$\geq 210$ mmHg	$\geq 120$ mmHg

Adapun faktor terjadinya hipertensi menurut Khalim dkk. (2023), antara lain:

1. Faktor keturunan

Riwayat keluarga dengan hipertensi dapat meningkatkan risiko seseorang memiliki penyakit hipertensi. Variasi genetik tertentu dapat memengaruhi fungsi sistem kardiovaskular dan mengarah pada peningkatan tekanan darah.

2. Gaya hidup yang tidak sehat

Gaya hidup yang tidak sehat, seperti kurang berolahraga/aktivitas fisik, makan makanan tidak sehat, gula berlebih, garam berlebih, lemak berlebih, kelebihan berat badan, dan konsumsi alkohol berlebihan dapat meningkatkan risiko terjadinya hipertensi.

3. Stres yang berlebihan

Stres yang berlebihan dapat memengaruhi tekanan darah. Saat stres, tubuh memproduksi hormon stres seperti kortisol yang dapat menyebabkan peningkatan tekanan darah.

#### 4. Usia

Sejumlah penyakit dapat memengaruhi tekanan darah, termasuk penyakit ginjal, kelainan endokrin (seperti diabetes, hipotiroidisme, atau hipoparatiroidisme), penyakit kardiovaskular (seperti penyakit arteri koroner atau penyakit arteri perifer), dan gangguan tidur seperti sleep apnea dapat berkontribusi terhadap hipertensi.

#### 5. Obat-obatan

Penggunaan obat-obatan tertentu, seperti kontrasepsi oral, obat anti-inflamasi nonsteroid (OAINS), obat penghilang rasa sakit, obat dekonjestan nasal, dan obat-obatan tertentu untuk kondisi kesehatan tertentu, dapat menyebabkan peningkatan tekanan darah.

#### 6. Faktor lingkungan

Paparan polutan lingkungan, seperti polusi udara dan logam berat tertentu, dapat meningkatkan risiko tekanan darah tinggi.

### **2.2 Kajian Integrasi Penyakit Hipertensi dengan Al-Qur'an**

Al-Qur'an tidak membahas persoalan hipertensi secara spesifik, namun, perintah untuk menerapkan pola hidup sehat dapat ditemui secara implisit maupun eksplisit dalam Al-Qur'an. Salah satu cara menjaga pola hidup sehat adalah mengonsumsi makanan dan minuman yang baik. Dalam surat Al-Baqarah ayat 168 dijelaskan bahwa manusia diwajibkan untuk mengonsumsi makanan dan minum yang halal (Aulia, 2023). Dalam Tafsir Ibnu Katsir dijelaskan bahwa

makanan yang tersedia di bumi boleh dikonsumsi manusia, asalkan makanan tersebut halal dan baik untuk dimakan, serta tidak berdampak buruk bagi kesehatan fisik dan mental. Apabila seseorang memakan sesuatu yang memberikan dampak buruk bagi badan, maka dapat memicu berbagai penyakit seperti kolesterol atau darah tinggi. Oleh karena itu, penting untuk memperhatikan kandungan gizi dan nutrisi dalam makanan yang dikonsumsi. Karena makanan-makanan yang tidak sehat, seperti makanan yang tinggi gula, lemak, dan garam, dapat meningkatkan risiko berbagai penyakit kronis, seperti penyakit jantung, stroke atau hipertensi (Aulia, 2023). Memahami dampak makanan terhadap kesehatan tubuh memerlukan pengetahuan yang memadai, yang hanya bisa diperoleh melalui pendidikan dan pemahaman yang mendalam tentang gizi dan kesehatan. Oleh karena itu, pentingnya ilmu pengetahuan dalam kehidupan sehari-hari tidak dapat diabaikan.

Ilmu pengetahuan memiliki peran besar dalam kehidupan seseorang, karena dengan ilmu pengetahuan maka manusia dapat bermanfaat untuk keluarga dan sekitarnya. Dalam sebuah hadis riwayat Muslim pun disebutkan tentang keutamaan mempelajari ilmu pengetahuan dalam Islam, Rasulullah SAW bersabda:

وَمَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ

*"Siapa yang menempuh jalan untuk mencari ilmu, maka Allah akan memudahkan baginya jalan menuju surga"*

Dalil tersebut menjadi bukti bahwa umat Islam wajib menuntut ilmu, karena Allah telah berjanji di dalam Al-Qur'an bahwa orang yang pergi untuk menuntut ilmu maka akan diangkat derajatnya, dan Nabi Muhammad juga menjelaskan bahwa belajar atau berjalan untuk mencari ilmu maka Allah akan memudahkan untuknya jalan masuk surga. Dengan adanya penelitian ini diharapkan dapat membangun sistem klasifikasi penyakit hipertensi sehingga hal ini dapat menunjukkan kemajuan

dalam bidang sains dan teknologi yang dapat bermanfaat bagi umat manusia (Hariyanto, 2016).

### **2.3 Kajian Hipertensi dengan Metode *Support Vector Machine***

Penelitian ini bertujuan untuk mengklasifikasikan diagnosis penyakit hipertensi menggunakan metode SVM. Tahapan yang dilakukan dalam penelitian ini adalah pertama-tama data diimport terlebih dahulu ke dalam *Jupyter Notebook*. Selanjutnya, akan dilakukan tahapan *Exploratory Data Analysis* (EDA) terlebih dahulu untuk mengetahui informasi yang ada di dalam data serta memberikan *insight* yang lebih baik mengenai data. Langkah selanjutnya yaitu, *data preprocessing* berupa *data cleaning* untuk menghapus atribut yang tidak dipakai, *data transformation* untuk mengubah data dari format aslinya ke format lain yang lebih sesuai agar dapat diolah pada tahap klasifikasi dan *data oversampling* untuk meratakan sampel data dari kedua nilai dengan memperbanyak data minoritas hingga setara data mayoritas.

Kemudian dilanjutkan dengan proses pembagian data menjadi dua, yaitu data *training* dan data *testing*. Setelah itu, data *training* akan digunakan untuk proses klasifikasi menggunakan metode SVM sehingga dihasilkan hasil klasifikasi dengan *output* yaitu berupa “1” jika pasien terkonfirmasi berpenyakit hipertensi dan “0” jika pasien tidak terkonfirmasi berpenyakit hipertensi. Tahapan selanjutnya yaitu evaluasi terkait ketepatan model klasifikasi metode SVM menggunakan *k-fold cross validation*. Tahapan terakhir yaitu menganalisis kinerja model klasifikasi metode SVM untuk mengetahui hasil prediksi klasifikasi diagnosis penyakit hipertensi menggunakan *confusion matrix*.

## BAB III METODE PENELITIAN

### 3.1 Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian kuantitatif. Penelitian ini menggunakan data kuantitatif untuk menguji data sekunder berupa data PTM (Penyakit Tidak Menular) hipertensi di Puskesmas Mojolangu Kota Malang.

### 3.2 Data dan Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder merupakan data yang diperoleh dari sumber yang telah tersedia, seperti buku, jurnal atau situs *web*. Penelitian ini menggunakan data sekunder yaitu berupa data *soft file* PTM (Penyakit Tidak Menular) hipertensi di Puskesmas Mojolangu Kota Malang. Jumlah data yang digunakan yaitu sebanyak 1373 data. Tabel 3.1 menyajikan variabel-variabel yang digunakan dalam penelitian ini.

**Tabel 3.1** Variabel Penelitian

No.	Variabel	Keterangan
1.	Usia	Usia pasien pada saat penelitian
2.	Jenis Kelamin	Laki-laki/Perempuan
3.	Kebiasaan Merokok	Ya/Tidak
4.	Kurang Aktivitas Fisik (Min. 30 Menit/Hari)	Ya/Tidak
5.	Konsumsi Gula Berlebih (>4 Sendok Makan/Hari)	Ya/Tidak
6.	Konsumsi Garam Berlebih (>1 Sendok Teh/Hari)	Ya/Tidak
7.	Konsumsi Lemak Berlebih (>4 Potong Gorengan/Hari)	Ya/Tidak
8.	Kurang Makan Buah dan Sayur (<5 Porsi/Hari)	Ya/Tidak
9.	Konsumsi Alkohol	Ya/Tidak
10.	Tekanan Darah Sistolik	Tekanan darah saat jantung memompa darah. Contoh : 140 mmHg
11.	Tekanan Darah Diastolik	Tekanan darah saat jantung beristirahat Contoh: 80 mmHg
12.	Diagnosis Hipertensi	0: Tidak Berpenyakit hipertensi 1: Berpenyakit Hipertensi

### 3.3 Tahapan Penelitian

Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

#### 3.3.1 Tahapan Klasifikasi Diagnosis Penyakit Hipertensi Menggunakan SVM

1. Pengumpulan data penyakit hipertensi yang terdiri dari variabel dependen dan independen.
2. Melakukan *Exploratory Data Analysis* (EDA) untuk mengetahui informasi-informasi serta *insight* yang ada di dalam data.
3. Melakukan *preprocessing data*

Terdapat beberapa tahapan dalam proses *preprocessing data* yaitu:

- a. *Data Cleaning*. *Data cleaning* adalah proses mendeteksi dan memperbaiki (atau menghapus) data yang salah, rusak, duplikat atau format yang tidak sesuai. Tahap ini bertujuan untuk memastikan keakuratan, konsistensi, dan kegunaan data dalam kumpulan data.
- b. *Data Transformation*. *Data Transformation* merupakan tahapan proses mengubah data dari format aslinya ke format lain yang lebih sesuai untuk tujuan tertentu. Dalam hal ini, akan diubah atribut yang memiliki tipe data kategorikal menjadi numerik.
- c. *Oversampling Data*. *Oversampling data* adalah teknik yang digunakan untuk meningkatkan jumlah sampel dari kelas minoritas dalam kumpulan data yang tidak seimbang. Teknik *oversampling* yang akan digunakan SMOTE (*Synthetic Minority Over-sampling Technique*). SMOTE adalah teknik statistik untuk meningkatkan jumlah kasus dalam himpunan data Anda dengan cara yang seimbang.

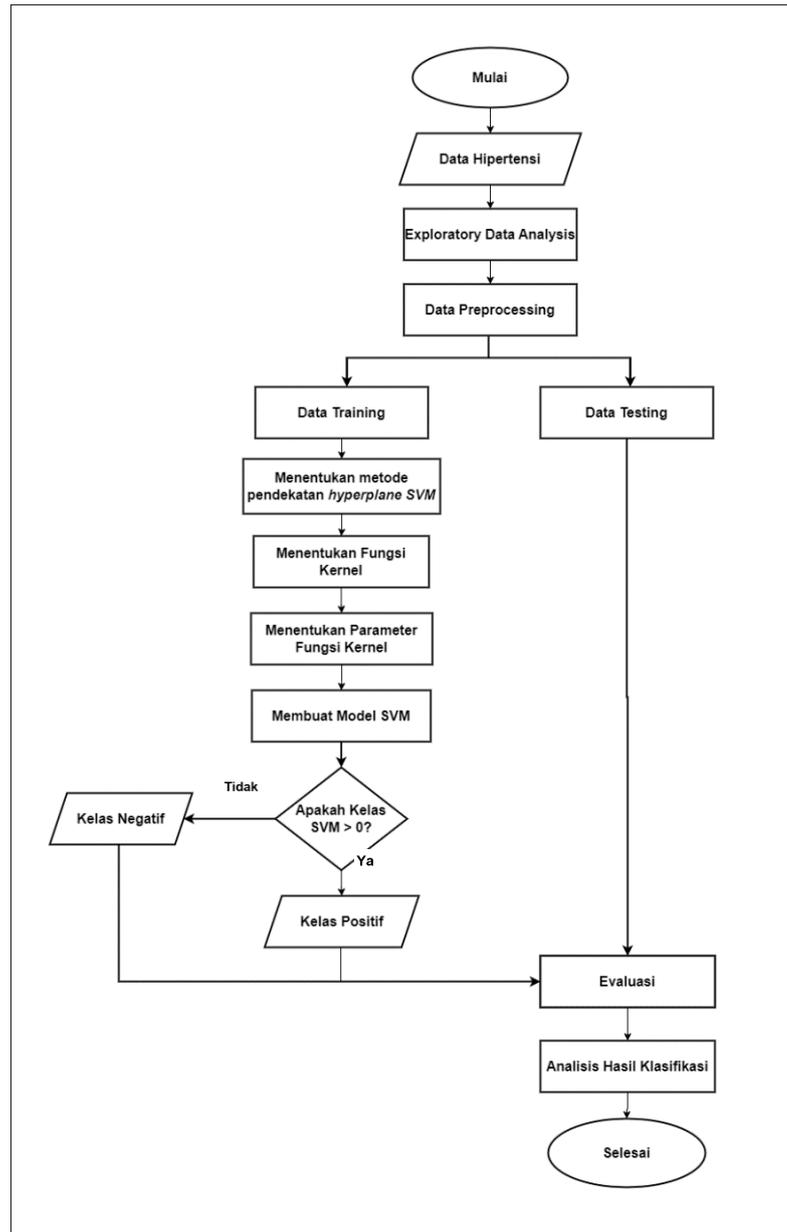
4. Membagi data menjadi dua bagian, yaitu data *training* dan data *testing* dengan rasio 80:20. Rasio tersebut dipilih berdasarkan hasil uji coba yang telah dilakukan, di mana terbukti bahwa model yang dilatih dengan rasio tersebut mencapai tingkat akurasi yang tinggi. Data *training* akan digunakan untuk model klasifikasi dan data *testing* akan digunakan untuk evaluasi.
5. Melakukan proses klasifikasi menggunakan metode SVM

Proses klasifikasi menggunakan metode SVM terdiri dari beberapa langkah, yaitu sebagai berikut:

- a. Memilih metode untuk menentukan *hyperplane* SVM yang akan digunakan pada tahap klasifikasi.
- b. Memilih fungsi kernel untuk membentuk model SVM.  
  
Penelitian ini akan melakukan uji coba terhadap dua kernel, yaitu kernel linier dan kernel RBF.
- c. Menentukan nilai parameter yang akan digunakan untuk model SVM.
- d. Membentuk model SVM menggunakan fungsi kernel.

### **3.3.2 Tahapan Perhitungan Nilai Akurasi Metode SVM Pada Pengklasifikasian Potensi Penyakit Hipertensi**

1. Melakukan evaluasi terkait ketepatan model klasifikasi metode SVM menggunakan *k-fold cross validation*.
2. Menganalisis kinerja model klasifikasi metode SVM untuk mengetahui hasil prediksi klasifikasi diagnosis penyakit hipertensi menggunakan *confusion matrix*.

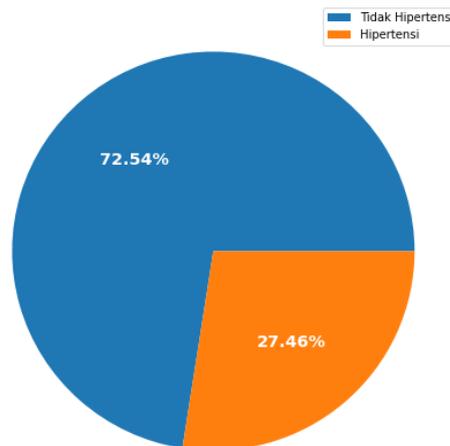


**Gambar 3.1** Flowchart Alur Penelitian

## BAB IV PEMBAHASAN

### 4.1 *Exploratory Data Analysis (EDA)*

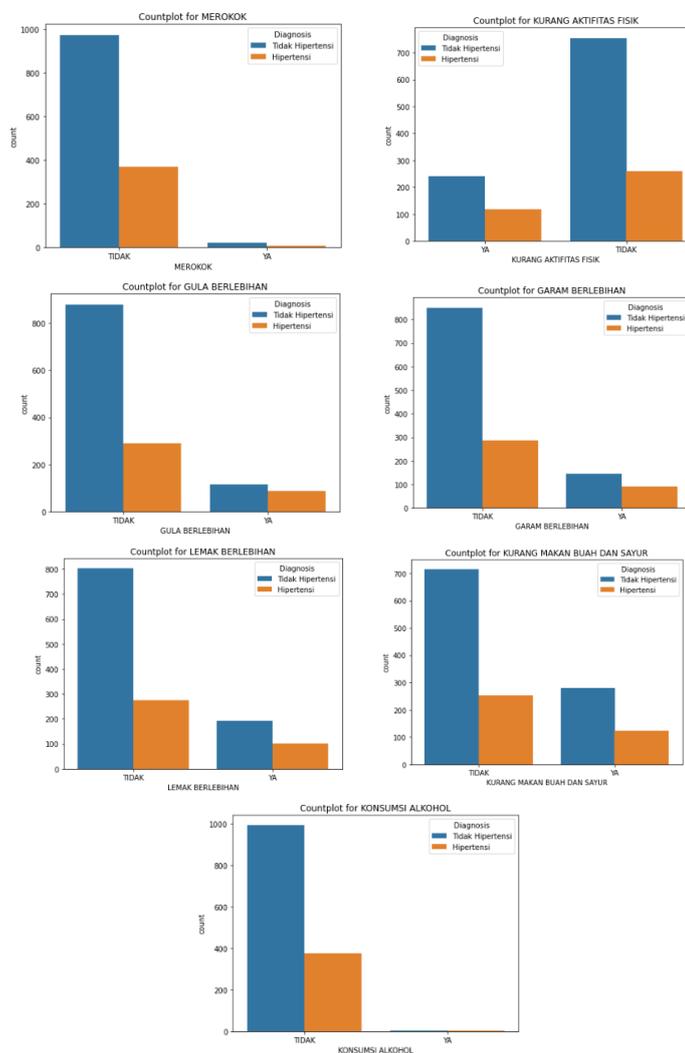
Penelitian ini menggunakan 10 variabel independen dan 1 variabel dependen setelah menghapus variabel jenis kelamin pada tahapan *data cleaning*. Variabel independen yang digunakan yaitu tanggal lahir, kebiasaan merokok, kurang aktivitas fisik, konsumsi gula berlebih, konsumsi garam berlebih, konsumsi lemak berlebih, kurang makan buah dan sayur, konsumsi alkohol, tekanan darah sistolik, dan tekanan darah diastolik. Lalu untuk variabel dependen yang digunakan adalah diagnosis hipertensi. Berikut perbandingan pasien yang terdiagnosis berpenyakit hipertensi dan tidak hipertensi pada Gambar 4.1.



**Gambar 4.1** *Pie Chart* Diagnosis Hipertensi

Berdasarkan *pie chart* pada Gambar 4.1, dapat diketahui bahwa kategori tidak hipertensi lebih banyak daripada hipertensi. Total pasien yang tidak berpenyakit hipertensi sebanyak 996 data, sedangkan yang berpenyakit hipertensi sebanyak 377 data. Hal ini menunjukkan adanya indikasi *imbalanced data* atau disebut juga sebagai data yang tidak seimbang.

Atribut-atribut yang memiliki tipe data kategorikal yaitu kebiasaan merokok, kurang aktifitas fisik, konsumsi gula berlebih, konsumsi garam berlebih, konsumsi lemak berlebih, kurang makan buah dan sayur, dan konsumsi alkohol. Pada atribut dengan tipe data kategorikal dilakukan analisis bivariat menggunakan diagram batang untuk mengetahui frekuensi pasien berpenyakit hipertensi dengan variabel-variabel dependen. Berikut ditunjukkan frekuensi hubungan antara atribut-atribut yang memiliki tipe data kategorikal dengan atribut hipertensi pada Gambar 4.2.



**Gambar 4.2** Frekuensi Hubungan Antara Atribut-Atribut Kategorikal dengan Atribut Diagnosis Hipertensi

Berdasarkan Gambar 4.2, maka dapat diinterpretasikan sebagai berikut:

1. Orang yang tidak merokok memiliki risiko lebih rendah untuk menderita hipertensi. Sebaliknya, orang yang merokok memiliki risiko lebih tinggi untuk menderita hipertensi.
2. Orang yang sering beraktivitas fisik memiliki risiko lebih rendah untuk menderita hipertensi. Sebaliknya, orang yang jarang beraktivitas fisik memiliki risiko lebih tinggi untuk menderita hipertensi.
3. Orang yang tidak memiliki riwayat gula berlebih cenderung tidak menderita hipertensi, sedangkan orang yang memiliki riwayat gula berlebih memiliki risiko lebih tinggi untuk menderita hipertensi.
4. Orang yang tidak memiliki riwayat garam berlebih memiliki risiko lebih rendah untuk menderita hipertensi. Sebaliknya, orang yang memiliki riwayat garam berlebih memiliki risiko lebih tinggi untuk menderita hipertensi.
5. Orang yang tidak memiliki riwayat lemak berlebih memiliki risiko lebih rendah untuk menderita hipertensi, sedangkan orang yang memiliki riwayat lemak berlebih memiliki risiko lebih tinggi untuk menderita hipertensi.
6. Orang yang tidak memiliki riwayat kurang makan buah dan sayur memiliki risiko lebih rendah untuk menderita hipertensi, sedangkan orang yang memiliki riwayat kurang makan buah dan sayur memiliki risiko lebih tinggi untuk menderita hipertensi.
7. Orang yang tidak memiliki riwayat konsumsi alkohol memiliki risiko yang sangat rendah untuk menderita hipertensi.

Dari penjelasan di atas, dapat disimpulkan bahwa setiap atribut yang ada belum bisa dipastikan mempengaruhi hipertensi atau tidak. Selain itu, terlihat bahwa

jumlah pasien yang tidak menderita hipertensi cenderung lebih banyak dibandingkan dengan pasien yang menderita hipertensi. Orang yang tidak menderita hipertensi umumnya memiliki pola atau gaya hidup yang lebih baik dibandingkan dengan mereka yang menderita hipertensi. Kemudian dilakukan analisis *odds ratio* untuk memahami besarnya pengaruh atribut-atribut tersebut terhadap kejadian hipertensi. Analisis *odds ratio* menunjukkan bahwa merokok dan kurang aktivitas fisik memiliki risiko rendah terhadap hipertensi. Hal ini bertentangan dengan fakta bahwa merokok merupakan salah satu faktor risiko hipertensi. Hal ini bisa disebabkan oleh ukuran sampel yang tidak memadai atau variabilitas data yang tidak terkontrol. Oleh karena itu, diperlukan penelitian lebih lanjut dengan metode lebih *robust* dan sampel lebih besar untuk memastikan pengaruh kedua atribut tersebut terhadap hipertensi.

Atribut-atribut yang memiliki tipe data numerik yaitu usia, tekanan darah sistolik, tekanan darah diastolik. Berikut statistik deskriptif dari atribut numerik pada Tabel 4.1.

**Tabel 4.1** Statistik Deskriptif Pada Atribut Numerik

<b>Variabel</b>	<b>Rata-Rata</b>	<b>Standar Deviasi</b>	<b>Minimum</b>	<b>Maksimum</b>
<b>Usia</b>	47,40	15,55	3	102
<b>Tekanan Darah Sistolik</b>	132,37	22,40	81	240
<b>Tekanan Darah Diastolik</b>	80,23	11,59	11	155

Berdasarkan Tabel 4.1, diketahui distribusi usia pasien dalam penelitian ini memiliki rata-rata 47 tahun, dengan rentang usia terentang dari 3 tahun hingga 102 tahun dan standar deviasi usia pasien dalam penelitian ini adalah 15,55 tahun. Kemudian untuk atribut tekanan darah sistolik diketahui memiliki rata-rata sebesar 132,37 mmHg, standar deviasi sebesar 22,40 mmHg, minimum sebesar 81 mmHg

dan maksimum sebesar 240 mmHg. Sedangkan untuk atribut tekanan darah diastolik diketahui memiliki rata-rata sebesar 80,23 mmHg, standar deviasi sebesar 11,59 mmHg, minimum sebesar 11 mmHg dan maksimum sebesar 155 mmHg. Kedua rata-rata ini berada pada batas normal, yaitu 140 mmHg untuk tekanan sistolik dan 90 mmHg untuk tekanan diastolik. Hal ini menunjukkan bahwa sebagian besar pasien dalam data memiliki tekanan darah yang normal.

## 4.2 Data Preprocessing

### 4.2.1 Data Cleaning

Pada tahapan ini dilakukan penghapusan data terhadap atribut Jenis Kelamin karena seluruh nilai atribut memiliki nilai yang sama yaitu Perempuan. Langkah ini bertujuan untuk menyederhanakan data dan mengurangi kompleksitas yang tidak perlu sehingga berpotensi untuk memberikan kontribusi yang lebih signifikan terhadap analisis yang dilakukan. Berikut tampilan data setelah dilakukan penghapusan terhadap atribut jenis kelamin pada Tabel 4.2.

**Tabel 4.2** Tampilan Data

No.	Variabel		Keterangan
1.	Usia	$X_1$	Usia pasien pada saat penelitian
2.	Kebiasaan Merokok	$X_2$	Ya/Tidak
3.	Kurang Aktivitas Fisik	$X_3$	Ya/Tidak
4.	Konsumsi Gula Berlebih	$X_4$	Ya/Tidak
5.	Konsumsi Garam Berlebih	$X_5$	Ya/Tidak
6.	Konsumsi Lemak Berlebih	$X_6$	Ya/Tidak
7.	Kurang Makan Buah dan Sayur	$X_7$	Ya/Tidak
8.	Konsumsi Alkohol	$X_8$	Ya/Tidak
9.	Tekanan Darah Sistolik	$X_9$	Tekanan darah saat jantung memompa darah
10.	Tekanan Darah Diastolik	$X_{10}$	Tekanan darah saat jantung beristirahat
11.	Diagnosis Hipertensi	$Y$	0: Tidak Berpenyakit hipertensi 1: Berpenyakit Hipertensi

#### 4.2.2 Data Transformation

Proses transformasi data dilakukan untuk mengubah data pada *dataset* sesuai dengan format yang dapat diproses oleh *software* yang digunakan. Pada tahapan ini, terdapat proses transformasi data atribut-atribut yang bersifat kategorikal menjadi numerik. Berikut tampilan transformasi data kategorikal ke numerik pada Tabel 4.3.

**Tabel 4.3** Tampilan Perubahan Atribut Data

<b>Atribut</b>	<b>Sebelum Transformasi</b>	<b>Sesudah Transformasi</b>
Kebiasaan Merokok	Ya	1
	Tidak	0
Kurang Aktivitas Fisik	Ya	1
	Tidak	0
Konsumsi Gula Berlebih	Ya	1
	Tidak	0
Konsumsi Garam Berlebih	Ya	1
	Tidak	0
Konsumsi Lemak Berlebih	Ya	1
	Tidak	0
Kurang Makan Buah dan Sayur	Ya	1
	Tidak	0
Konsumsi Alkohol	Ya	1
	Tidak	0
Diagnosis Hipertensi	Berpenyakit Hipertensi	1
	Tidak Berpenyakit Hipertensi	0

Setelah semua atribut diubah menjadi numerik, maka dapat dicari korelasi antar variabel menggunakan korelasi *pearson*. Korelasi *pearson* digunakan untuk mengukur hubungan linear antara dua variabel. Simbol untuk korelasi *pearson* adalah “p” jika diukur dalam populasi, dan “r” jika diukur dalam sampel. Hasil korelasi akan berupa nilai yang berkisar antara -1 sampai +1, dengan interpretasi sebagai berikut:

1. Apabila  $r = -1$  maka korelasi negatif sempurna (hubungan berlawanan arah yang paling kuat)
2. Apabila  $r = 1$  maka korelasi positif sempurna (hubungan searah yang paling kuat)
3. Apabila koefisien korelasi menunjukkan angka 0, maka kedua variabel kemungkinan besar tidak saling mempengaruhi.
4. Apabila nilai  $r$  antara  $-1$  dan  $0$  atau  $0$  dan  $+1$  maka hubungan linear dengan kekuatan yang bervariasi (semakin dekat ke  $-1$  atau  $+1$ , semakin kuat hubungannya).

Berikut tampilan korelasi antar atribut dalam penelitian ini pada Tabel 4.4.

**Tabel 4.4** Korelasi Antar Atribut

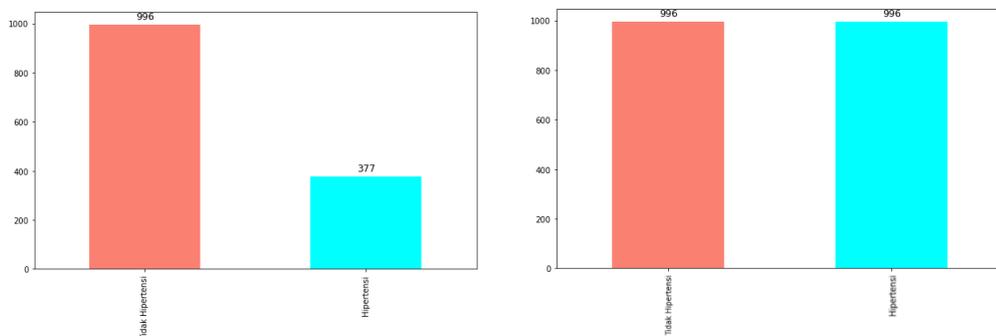
Korelasi Antar Atribut	Merokok	Kurang Aktifitas Fisik	Gula Berlebihan	Garam Berlebihan	Lemak Berlebihan	Kurang Makan Buah dan Sayur	Konsumsi Alkohol	T. D. Sistolik	T. D. Diastolik	Diagnosis	Usia
Merokok	1	-0.02408	0.015135	0.005182	0.028267	0.058346	0.165596	0.025603	-0.01838	-0.01631	0.04057
Kurang Aktifitas Fisik	-0.02408	1	0.044812	0.154405	0.083734	0.249609	0.019365	0.092399	0.062216	0.065778	0.045811
Gula Berlebihan	0.015135	0.044812	1	0.204089	0.183922	0.156397	-0.02515	0.213874	0.086245	0.137332	0.139752
Garam Berlebihan	0.005182	0.154405	0.204089	1	0.186765	0.169607	0.036697	0.197948	0.112999	0.11168	0.119583
Lemak Berlebihan	0.028267	0.083734	0.183922	0.186765	1	0.138794	0.027352	0.141263	0.075619	0.082063	0.062866
Kurang Makan Buah dan Sayur	0.058346	0.249609	0.156397	0.169607	0.138794	1	0.014084	0.083181	0.024886	0.042275	0.063256
Konsumsi Alkohol	0.165596	0.019365	-0.02515	0.036697	0.027352	0.014084	1	0.033331	0.016747	0.017166	0.021785
T. D. Sistolik	0.025603	0.092399	0.213874	0.197948	0.141263	0.083181	0.033331	1	0.531745	0.565757	0.547133
T. D. Diastolik	-0.01838	0.062216	0.086245	0.112999	0.075619	0.024886	0.016747	0.531745	1	0.684391	0.137859
Diagnosis	-0.01631	0.065778	0.137332	0.11168	0.082063	0.042275	0.017166	0.565757	0.684391	1	0.219468
Usia	0.04057	0.045811	0.139752	0.119583	0.062866	0.063256	0.021785	0.547133	0.137859	0.219468	1

Tabel 4.4 menunjukkan bahwa nilai korelasi positif terkuat terdapat pada atribut tekanan darah diastolik dengan diagnosis hipertensi yaitu 0,684391 dan nilai korelasi negatif terkuat terdapat hubungan atribut gula berlebihan dan konsumsi alkohol yaitu -0,025149.

### 4.2.3 Oversampling Data

Berdasarkan Gambar 4.1, dapat disimpulkan bahwa data penelitian yang digunakan tidak seimbang. Hal ini dapat menyebabkan nilai performa yang tidak baik dari akurasi, presisi, *recall* dan *F1-Score*. Oleh karena itu, untuk membuat data tersebut *balance* digunakan teknik *oversampling*, yaitu dengan meratakan

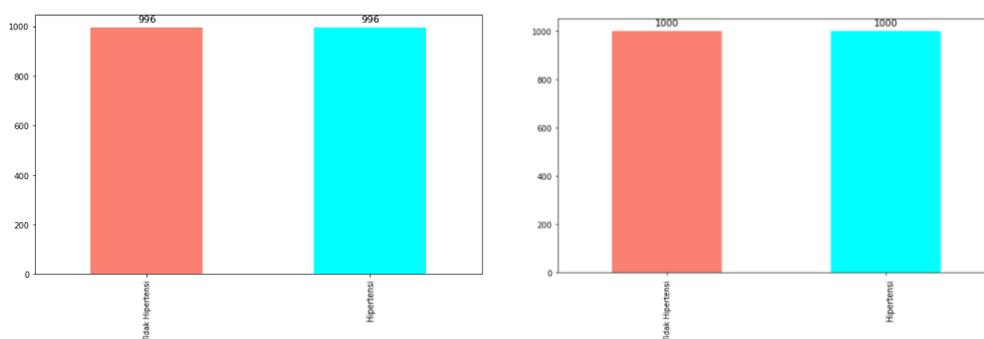
sampel data dari kedua nilai dengan memperbanyak data minoritas hingga setara nilai mayoritas. SMOTE bekerja dengan cara yang terstruktur. Pertama, SMOTE memilih sampel-sampel dari kelas minoritas dalam *dataset*. Kemudian, untuk setiap sampel minoritas yang dipilih, SMOTE mencari tetangga terdekatnya dalam ruang fitur, biasanya menggunakan metode seperti *K-Nearest Neighbors* (KNN). Setelah menemukan tetangga terdekat untuk suatu sampel minoritas, SMOTE membuat sampel sintetis baru di antara sampel minoritas dan tetangga terdekatnya. Sampel sintetis ini berada pada garis yang menghubungkan dua sampel tersebut di dalam ruang fitur. Akhirnya, sampel sintetis baru ini ditambahkan ke *dataset*, meningkatkan jumlah sampel dalam kelas minoritas hingga mencapai tingkat yang diinginkan atau seimbang dengan kelas mayoritas. Berikut gambaran label dari kelas data sebelum dan setelah proses *oversampling* pada Gambar 4.4.



**Gambar 4.4** Tampilan Data Sebelum Dan Sesudah *Oversampling*

Berdasarkan Gambar 4.4, terlihat bahwa sebelum dilakukan *oversampling data*, terdapat 996 data non hipertensi dan 377 data hipertensi. Setelah dilakukan proses *oversampling data*, kedua kelas memiliki jumlah data yang sama, yaitu 996 data untuk setiap kelas. Selanjutnya, dilakukan *oversampling data* yang kedua untuk memastikan perlakuan yang sama terhadap kedua label data. Dalam

proses *oversampling data* kedua, setiap label data ditingkatkan menjadi 1000 data. Penelitian ini akan menggunakan data yang telah mengalami proses *oversampling* kedua. Berikut gambaran label dari kelas data sebelum dan setelah proses *oversampling* kedua pada Gambar 4.5.



**Gambar 4.5** Tampilan Data Sebelum Dan Sesudah *Oversampling* Kedua

### 4.3 Klasifikasi dengan Metode SVM

Proses klasifikasi dimulai dengan membagi data menjadi dua, yaitu data *training* dan *testing* dengan rasio 80:20. Rasio tersebut dipilih berdasarkan hasil uji coba yang telah dilakukan, di mana terbukti bahwa model yang dilatih dengan rasio tersebut mencapai tingkat akurasi yang tinggi. Pada data *training* yang terdiri dari beberapa variabel dan kelas target dijadikan sebagai data sampel untuk membentuk model SVM. Sedangkan data *testing* digunakan untuk mengevaluasi ketepatan model klasifikasi berdasarkan *confusion matrix* yang terbentuk.

#### 4.3.1 Perhitungan Manual dalam Pengklasifikasian Menggunakan Metode SVM

Pada perhitungan manual, dilakukan klasifikasi SVM menggunakan kernel RBF yang didefinisikan pada Persamaan (2.20) dengan parameter  $\gamma$  (*gamma*) = 1. Pada contoh perhitungan manual digunakan sebanyak 4 data, yaitu data ke 7,16,17 dan 28 yang terletak pada Lampiran 1. Adapun rincian data yang digunakan ditampilkan pada Tabel 4.5.

**Tabel 4.5** Contoh Data Untuk Perhitungan Manual

No	$X_1$	$X_2$	Y
1	1	0	0
2	0	1	1
3	0	0	1
4	0	1	0

Langkah pertama yang perlu dilakukan dalam perhitungan klasifikasi SVM adalah mencari nilai matriks kernel K. Kernel K dapat dihitung dengan dimensi  $m \times m$ , Dimana  $m$  adalah banyaknya data. Adapun persamaan yang digunakan dalam mencari kernel K adalah Persamaan (2.20). Parameter yang digunakan dalam perhitungan adalah  $\gamma = 1$  sehingga didapatkan perhitungan sebagai berikut:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right)$$

$$K(x_1, x_1) = \exp\left(-1 \left(\sqrt{(1-1)^2 + (0-0)^2}\right)^2\right) = \exp(0) = 1$$

$$K(x_1, x_2) = \exp\left(-1 \left(\sqrt{(1-0)^2 + (0-1)^2}\right)^2\right) = \exp(-2) = 0,135335$$

$$K(x_1, x_3) = \exp\left(-1 \left(\sqrt{(1-0)^2 + (0-0)^2}\right)^2\right) = \exp(-1) = 0,367879$$

$$K(x_1, x_4) = \exp\left(-1 \left(\sqrt{(1-0)^2 + (0-1)^2}\right)^2\right) = \exp(-2) = 0,135335$$

Kemudian untuk data selanjutnya dilakukan dengan cara yang sama sehingga didapatkan matriks kernel K yang ditunjukkan sebagai berikut.

$$K = \begin{bmatrix} 1 & 0,135335 & 0,367879 & 0,135335 \\ 0,135335 & 1 & 0,367879 & 1 \\ 0,367879 & 0,367879 & 1 & 0,367879 \\ 0,135335 & 1 & 0,367879 & 1 \end{bmatrix}$$

Dengan menggunakan kernel  $K$  sebagai pengganti dot-product  $\phi(x_i) \cdot \phi(x_j)$  dalam persamaan dualitas *Lagrange Multiplier* pada Persamaan (2.13), didapatkan:

$$L(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}(\alpha_1\alpha_2 + \alpha_2^2 + 0,735758\alpha_2\alpha_3 + \alpha_3^2)$$

Dengan syarat  $-\alpha_1 + \alpha_2 + \alpha_3 - \alpha_4 = 0$  dan  $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0$ .

Dalam fungsi tujuan, suku kedua sudah dikalikan dengan  $y_i y_j$ . Persamaan tersebut memenuhi standar *Quadratic Programming* sehingga bisa dibantu penyelesaiannya dengan solver komersial untuk *Quadratic Programming* (QP). Dengan bantuan perangkat lunak, didapatkan hasil sebagai berikut.

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$$

Hasil di atas menunjukkan bahwa semua data dalam contoh ini adalah *support vector*, karena nilai  $a \neq 0$ . Kemudian dapat dilakukan pencarian nilai  $w$  (*weight*) menggunakan Persamaan (2.14) sehingga diperoleh hasil yang ditunjukkan pada Tabel 4.6.

**Tabel 4.6** Nilai *Weight*

$w_i$	Nilai <i>weight</i>
$w_1$	-1,64
$w_2$	2,50
$w_3$	2,10
$w_4$	-2,50

Adapun untuk mendapatkan nilai  $b$  (bias) dilakukan perhitungan menggunakan Persamaan (2.15). Pada penelitian ini didapatkan nilai  $b = -0,115$ . Setelah mendapatkan nilai  $w$  (*weight*) dan  $b$  (bias) maka selanjutnya dapat dibentuk model SVM yang digunakan dalam proses klasifikasi dengan menggunakan Persamaan (2.16) yaitu sebagai berikut.

$$f(\phi(x)) = \text{sign}\left(\sum_{i,j=1}^N \alpha_i y_i K(x_i) \cdot K(x_j) - 0,115\right)$$

Untuk mengklasifikasikan data pertama di Tabel 4.5, maka dapat dilakukan perhitungan sebagai berikut:

$$f(\phi(1)) = \text{sign}((1. -1. K(x_1). K(x_1) - 0,115) + (1. -1. K(x_1). K(x_2) - 0,115) \\ + (1. -1. K(x_1). K(x_3) - 0,115) + (1. -1. K(x_1). K(x_4) - 0,115))$$

$$f(\phi(1)) = \text{sign}((1. -1.1 - 0,115) + (1. -1.0,135335 - 0,115) \\ + (1. -1.0,367879 - 0,115) + (1. -1.0,135335 - 0,115))$$

$$f(\phi(1)) = \text{sign}(-2,10)$$

$$f(\phi(1)) = -1$$

Berdasarkan perhitungan di atas, maka data pertama diklasifikasikan sebagai kelas negatif. Proses perhitungan diatas akan diterapkan pada seluruh data *training* untuk mendapatkan model SVM yang optimal dalam klasifikasi penyakit hipertensi.

#### 4.3.2 Hasil Klasifikasi Menggunakan Metode SVM

Klasifikasi SVM untuk keseluruhan data menggunakan bahasa pemrograman *python* dan *software Jupyter Notebook*. Penelitian ini menggunakan metode SVM linier dan non-linier untuk uji coba mencari akurasi terbaik. Fungsi kernel yang digunakan untuk klasifikasi SVM adalah kernel linier dengan parameter  $C = 0,01; 0,1; 1; 10$  dan  $100$ , kernel RBF dengan parameter  $C = 0,01; 0,1; 1; 10$  dan  $100$  serta parameter  $\gamma$  (*gamma*) =  $0,01; 0,1; 1; 10$  dan  $100$ . Pemilihan parameter ini berdasarkan penelitian yang dilakukan oleh Kurnia dkk. (2018). Uji coba dilakukan sebanyak 1 kali dengan menggunakan 5 parameter kernel linier yang berbeda dan 25 parameter kernel RBF yang berbeda.

Setelah model SVM terbentuk dengan menggunakan data *training*, maka diperoleh nilai akurasi model menggunakan data *testing* pada parameter kernel linier sebagai berikut pada Tabel 4.7.

**Tabel 4.7** Nilai Akurasi Parameter Model SVM Kernel Linier

<i>C</i> ( <i>Cost</i> )	Akurasi	Jumlah <i>Support Vector</i>
0,01	92,75%	634
0,1	94,75%	332
1	95,25%	257
10	95,25%	245
100	95,25%	246

Sedangkan untuk parameter RBF diperoleh sebagai berikut pada Tabel 4.8.

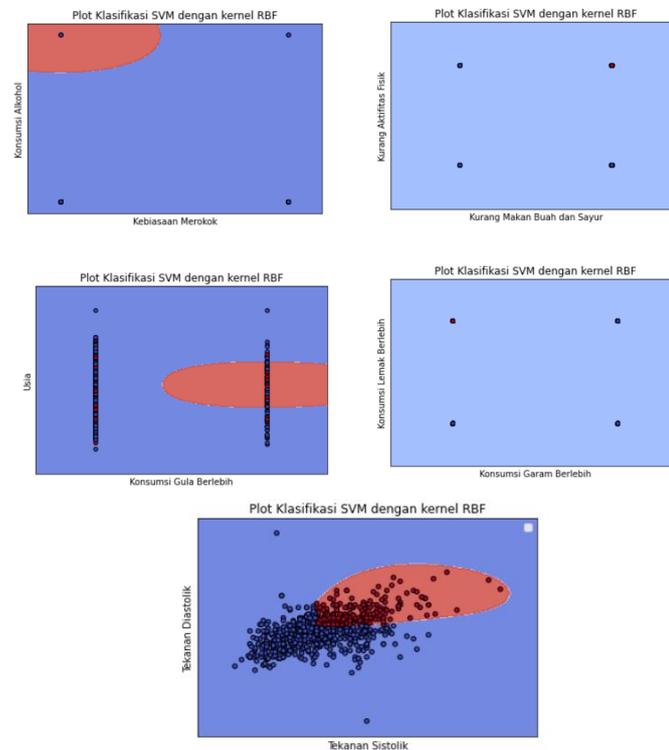
**Tabel 4.8** Nilai Akurasi Parameter Model SVM Kernel RBF

Parameter	Akurasi			
	$\gamma = 0,01$	$\gamma = 0,1$	$\gamma = 1$	$\gamma = 10$
$C = 0,01$	76,25%	89,75%	69,75%	49,75%
$C = 0,1$	90,75%	93,25%	89,25%	71,50%
$C = 1$	94,00%	95,25%	95,75%	90,25%
$C = 10$	95,25%	96,50%	96,75%	90,75%
$C = 100$	96,50%	98,00%	97,50%	90,75%

Berdasarkan *trial and error* yang dilakukan, diperoleh beberapa nilai akurasi dari beberapa percobaan menggunakan kernel linier dan RBF. Dari Tabel 4.6 dan Tabel 4.7, didapatkan akurasi tertinggi pada kernel RBF dengan parameter  $C = 100$  dan  $\gamma$  (*gamma*) = 0,1 dengan nilai akurasi sebesar 98%. Dari model tersebut, diperoleh jumlah *support vector* sebanyak 133 titik. Model tersebut dianggap sudah sangat baik, karena nilai *cost* ( $C$ ) yang kecil cenderung membuat margin lebih lebar dengan mengabaikan data yang tidak sesuai pada data *training*, sedangkan nilai *cost* ( $C$ ) yang besar cenderung menyesuaikan data *training* sehingga mengurangi kesalahan klasifikasi (Don, 2018).

Model klasifikasi SVM yang telah terbentuk dapat divisualisasikan dalam bentuk plot. Plot dibentuk dengan kombinasi dua variabel independen. Dari

kombinasi dua variabel independen secara keseluruhan diperoleh plot-plot sebagai berikut pada Gambar 4.6.



**Gambar 4.6** Plot Klasifikasi SVM Kernel RBF dengan  $\gamma = 0,1$  dan  $C = 100$

Berdasarkan Gambar 4.6, kombinasi dua variabel independen berhasil membentuk plot klasifikasi diagnosis penyakit hipertensi menggunakan SVM dengan kernel RBF. Pada plot tersebut, area berwarna merah ditandai kategori hipertensi, sedangkan area berwarna biru ditandai kategori tidak hipertensi. Titik berwarna merah menunjukkan data yang diklasifikasikan sebagai hipertensi, sedangkan titik berwarna biru menunjukkan data yang diklasifikasikan sebagai tidak hipertensi. Berdasarkan plot SVM yang diperoleh, dapat diinterpretasikan yaitu sebagai berikut:

1. Atribut kebiasaan merokok dan konsumsi alkohol dapat dipisahkan secara sempurna menggunakan *hyperplane*, seperti yang terlihat dari adanya garis pemisah yang mampu memisahkan kedua kelas dengan jelas.
2. Atribut kurang makan buah dan sayur tidak dapat dipisahkan menggunakan *hyperplane* karena tidak ada garis pemisah yang memisahkan kedua kelas dengan jelas.
3. Atribut usia dan konsumsi gula berlebih dapat dipisahkan menggunakan *hyperplane*, namun dengan pemisahan yang tidak sempurna. Ini terlihat dari adanya garis pemisah yang mampu memisahkan kedua kelas, meskipun tidak secara tuntas.
4. Atribut konsumsi lemak berlebih dan konsumsi garam berlebih tidak dapat dipisahkan menggunakan *hyperplane* karena tidak ada garis pemisah yang memisahkan kedua kelas dengan jelas.
5. Atribut tekanan darah sistolik dan tekanan darah diastolik dapat dipisahkan secara sempurna menggunakan *hyperplane*, seperti yang terlihat dari adanya garis pemisah yang mampu memisahkan kedua kelas dengan jelas.

#### **4.4 Evaluasi**

Evaluasi terkait ketepatan model klasifikasi yang diperoleh dapat ditentukan menggunakan *K-Fold Cross Validation* dengan  $k$  sebanyak 5 *fold*. Setiap bagian data akan digunakan sebagai data uji sebanyak 1 kali, sedangkan data lainnya akan digunakan sebagai data latih. Dengan demikian, model klasifikasi akan dilatih sebanyak 5 kali dan diuji sebanyak 5 kali. Berikut hasil akurasi, presisi dan *recall* yang didapatkan menggunakan *5-Fold Cross Validation* pada Tabel 4.9.

**Tabel 4.9** Hasil Evaluasi Menggunakan *5-Fold Cross Validation*

Parameter $k$	Presisi	Recall	Akurasi
$k = 1$	0.957983	0.942149	0.970000
$k = 2$	0.975904	0.975904	0.990000
$k = 3$	0.972477	0.946429	0.977500
$k = 4$	1.000000	0.936620	0.955000
$k = 5$	1.000000	0.965000	0.965000
Rata-rata	0.981273	0.953220	0.971500

Berdasarkan Tabel 4.8, diperoleh rata-rata presisi sebesar 0.981273 atau 98,12%, rata-rata *recall* sebesar 0.953220 atau 95,32% dan rata-rata akurasi sebesar 0.971500 atau 97,15%. Hasil ini menunjukkan bahwa model SVM dengan kernel RBF memiliki performa yang baik dalam mengklasifikasikan data.

#### 4.5 Analisis Performa Model Klasifikasi

Perhitungan ketepatan klasifikasi diagnosis penyakit hipertensi menggunakan metode SVM kernel RBF dengan parameter  $C = 100$  dan  $\gamma = 0.1$  dapat dianalisis menggunakan *confusion matrix* pada Tabel 4.10.

**Tabel 4.10** *Confusion Matrix* Hasil Penelitian

Data		Aktual	
		Non Hipertensi	Hipertensi
Prediksi	Non Hipertensi	192	7
	Hipertensi	1	200

Berdasarkan Tabel 4.9 dari *confusion matrix* menunjukkan bahwa 192 pasien non hipertensi diprediksi benar sebagai non hipertensi, 7 pasien hipertensi diprediksi salah sebagai non hipertensi, 200 pasien hipertensi diprediksi benar sebagai hipertensi, dan 1 pasien non-hipertensi diprediksi salah sebagai hipertensi. Kemudian didapatkan perhitungan nilai akurasi, presisi, *recall* dan *F1-Score* sebagai berikut.

$$\text{Akurasi} = \frac{192+200}{192+200+7+1} \times 100\% = 98\%$$

$$\text{Presisi} = \frac{192}{192+7} \times 100\% = 96.48\%$$

$$\text{Recall} = \frac{192}{192+1} \times 100\% = 99.48\%$$

$$\text{F1-Score} = 2 \times \frac{96.48\% \times 99.48\%}{96.48\% + 99.48\%} \times 100\% = 97.95\%$$

Berdasarkan hasil perhitungan tersebut, dapat disimpulkan bahwa klasifikasi diagnosis hipertensi menggunakan SVM memiliki akurasi sebesar 98%. Hal ini membuktikan bahwa metode SVM mampu mengklasifikasikan diagnosis hipertensi dengan sangat baik. Kemudian untuk nilai presisi didapatkan sebesar 96.48%, yang artinya dari 199 pasien non hipertensi yang diprediksi oleh model SVM, sebanyak 192 pasien di antaranya benar-benar non hipertensi. Kemudian untuk nilai *recall* didapatkan sebesar 99.48%, yang artinya dari 193 pasien non hipertensi yang ada, sebanyak 192 pasien berhasil diidentifikasi dengan benar oleh model SVM. Terakhir nilai *F1-Score* diperoleh sebesar 97.95% yang artinya model SVM mampu menyeimbangkan antara presisi dan *recall*. Hasil prediksi klasifikasi diagnosis penyakit hipertensi menggunakan SVM terdapat pada Lampiran 1.

#### **4.6 Kajian Keislaman dengan Hasil Penelitian**

Hipertensi, atau tekanan darah tinggi, merupakan penyakit kronis yang dapat memicu berbagai komplikasi serius, seperti stroke, serangan jantung, dan gagal ginjal. Untuk mencegah tekanan darah tinggi dibutuhkan pola hidup yang sehat dan seimbang. Islam sangat menekankan pentingnya kesehatan, karena kesehatan merupakan salah satu unsur penunjang utama dalam melaksanakan berbagai hal, baik itu bekerja maupun dalam pelaksanaan aktivitas ibadah kepada Allah SWT. Dari dilakukannya penelitian klasifikasi penyakit hipertensi menggunakan metode

SVM ini dan hasil yang sudah dihasilkan hendaknya bisa menjadi pengingat agar senantiasa bersyukur terhadap nikmat sehat yang sudah diberikan Allah SWT.

Berdasarkan hasil penelitian di atas, terbukti bahwa metode SVM dapat mengklasifikasikan penyakit hipertensi dengan tingkat akurasi mencapai 97,15%. Hasil penelitian ini sejalan dengan Hadis Riwayat Muslim yang menekankan pentingnya ilmu pengetahuan. Hal ini menunjukkan bahwa ilmu pengetahuan modern, khususnya di bidang kecerdasan buatan, mampu memberikan kontribusi nyata dalam meningkatkan kualitas layanan kesehatan. Kemampuan SVM dalam mengklasifikasikan hipertensi dengan baik dapat membantu tenaga medis dalam meningkatkan akurasi diagnosis, sehingga meminimalisir kesalahan diagnosis dan mempercepat proses pengobatan. Hal ini tentunya membawa manfaat besar bagi pasien, karena mereka dapat mendapatkan diagnosis yang lebih tepat dan pengobatan yang lebih efektif. Selain itu, penelitian ini diharapkan dapat menekan angka kematian akibat hipertensi dan dapat dikembangkan lebih lanjut agar dapat bermanfaat bagi masyarakat.

## BAB V KESIMPULAN

### 5.1 Kesimpulan

Berdasarkan rumusan masalah dan pembahasan pada bab sebelumnya maka diperoleh kesimpulan sebagai berikut:

1. Hasil klasifikasi metode SVM pada klasifikasi diagnosis penyakit hipertensi menunjukkan model terbaik menggunakan kernel RBF. Nilai parameter yang digunakan adalah  $C = 100$  dan  $\gamma$  (*gamma*) = 0,1. Hasil analisis klasifikasi menggunakan *confusion matrix* menunjukkan bahwa 192 pasien non hipertensi diprediksi benar sebagai non hipertensi, 7 pasien hipertensi diprediksi salah sebagai non hipertensi, 200 pasien hipertensi diprediksi benar sebagai hipertensi, dan 1 pasien hipertensi diprediksi salah sebagai non hipertensi.
2. Tingkat akurasi yang diperoleh dari klasifikasi diagnosis penyakit hipertensi menggunakan metode SVM kernel RBF dengan parameter  $C = 100$  dan  $\gamma$  (*gamma*) = 0,1 sebesar 97,15%. Hal ini menunjukkan bahwa metode SVM mampu mengklasifikasikan diagnosis penyakit hipertensi dengan sangat baik. Berdasarkan hasil penelitian ini, metode SVM dapat menjadi alternatif metode klasifikasi yang dapat digunakan untuk mengklasifikasikan diagnosis penyakit hipertensi.

## 5.2 Saran

Adapun beberapa hal yang dapat dijadikan acuan untuk penyempurnaan penelitian selanjutnya maka penulis memberikan saran berikut:

1. Pada penggunaan data penelitian, penelitian selanjutnya diharapkan dapat menambah beberapa atribut lainnya, seperti faktor keturunan, riwayat penggunaan obat-obatan dsb., karena hal itu merupakan faktor terjadinya hipertensi sehingga diharapkan dapat memiliki tingkat akurasi yang lebih baik.
2. Pada metode yang digunakan, penelitian selanjutnya diharapkan dapat menggunakan metode yang lain sehingga dapat diketahui perbandingan hasil keakuratan setiap metode.

## DAFTAR PUSTAKA

- Al-Albani, Muhammad Nashiruddin. Kitab: *As-Silsilah Ash-Shahîhah Hadits Riwayat Ath-Thabrani, Al-Mu'jam Al-Ausath, Dari Jabir Bin Abdullah r.A.*. pustaka imam syafi'i, n.d.
- Al-Bukhari, Abu Abdullah Muhammad bin Ismail, Ensiklopedia Hadits; Shahih al-Bukhari 1, Terj. Masyhar dan Muhammad Suhadi, Jakarta: Almahira, Cet. I, 2011.
- Andriansyah, Mochamad Rafli, Edy Santoso, and Sutrisno. (2018). "Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3)." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2(Vol 2 No 12 (2018)):7088–96.
- Aris, F., & Benyamin. (2019). Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi. *Jurnal Data Mining dan Sistem Informasi*, 1(1), 1-6. doi: 10.25123/jdmsi.v1i1.10
- Aulia, I. (2023). Al-Qur'an dan Tuntutan Pola Hidup (1): Jaga Kesehatan Jasmani. <https://jakarta.nu.or.id/akhlak%20tasawuf/al-quran-dan-tuntunan-pola-hidup-1-jaga-kesehatan-jasmani-Gii6g>. Diakses pada tanggal 25 Oktober 2023.
- Awalullaili, F. O., Ispriyanti, D., & Widiharih, T. (2023). Klasifikasi Penyakit Hipertensi Menggunakan Metode Svm Grid Search Dan Svm Genetic Algorithm (GA). *Jurnal Gaussian*, 11(4), 488-498. doi:10.14710/j.gauss.11.4.488
- BPS Kota Malang. (2019). Jumlah Kasus Penyakit Terbanyak di Kota Malang (Jiwa), 2019. <https://malangkota.bps.go.id/indicator/30/371/1/jumlah-kasus-penyakit-terbanyak-di-kota-malang.html>. Diakses pada tanggal 5 Oktober 2023.
- Don, D. P. W. R. (2018). Multiclass Classification Using Support Vector Machines [Tesis, Georgia Southern University]. <https://digitalcommons.georgiasouthern.edu/etd/1845/>. Diakses pada tanggal 5 Januari 2024.
- Gorunescu, F. (2011). *Data Mining Concepts, Models, And Techniques*. Springer.
- Gholami, R., & Fakhari, N. (2017). Support Vector Machine: Principles, Parameters, And Applications. In *Handbook of neural computation* (pp. 515-535). Academic Press.
- Hardi P, Wiwit. (2023). 2 Nikmat Yang Banyak Dilalaikan. <https://muslimah.or.id/7233-2-nikmat-yang-banyak-dilalaikan.html>. Diakses pada tanggal 20 Oktober 2023.
- Hariyanto, Muhsin. (2016). Jadilah Orang Yang "Bermanfaat". <https://muallimin.sch.id/2016/01/20/jadilah-orang-yang-bermanfaat/>. Diakses pada tanggal 15 Februari 2024.
- Kemenag. (2022). Qur'an Kemenag. <https://quran.kemenag.go.id/>. Diakses pada tanggal 15 Januari 2024.

- Khalim, K. A., Hayati, U., & Bahtiar, A. (2023). Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest Dan Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 498-504. doi:10.36040/jati.v7i1.6376
- Kurnia, A. I., Furqon, M. T., & Rahayudi, B. (2018). Klasifikasi Kualitas Susu Sapi Menggunakan Algoritme Support Vector Machine (SVM)(Studi Kasus: Perbandingan Fungsi Kernel Linier dan RBF Gaussian). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4453-4461
- Maiyanti, S. I., Zayanti, D. A., Andriani, Y., Suprihatin, B., Desiani, A., Salsabila, A., & Marselina, N. C. (2023). Perbandingan Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Support Vector Machine Dan K-Nearest Neighbor. *Jurnal Ilmiah Teknologi Kesehatan*, 18(1), 54-62. doi: 10.33998/processor.2023.18.1.700
- Misra, Rishabh. (2019). Support Vector Machines — Soft Margin Formulation and Kernel Trick. <https://towardsdatascience.com/support-vector-machines-soft-margin-formulation-and-kernel-trick-4c9729dc8efe>. Diakses pada 1 Maret 2024.
- Mitra Keluarga, 2022. 4 Penyebab Hipertensi Serta Gejala, Pengobatan, dan Pencegahannya. <https://www.mitrakeluarga.com/artikel/penyebab-hipertensi>. diakses pada 16 Juni 2023.
- Mucholladin, A. W., Bachtiar, F. A., & Furqon, M. T. (2021). Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 5(2), 622-633.
- Muslim ibn al-Hajjaj al-Naisaburi, Shahih Muslim, Libanon: Dar al-Kutub alIlmiah, 2000.
- Nawang Sari, T. (2017). Analisis Data Eksploratif Prestasi Mahasiswa Pada Mata Kuliah Statistika Matematika II. *Prosiding SNasPPM*, 2(1), 303-308.
- Nurchaya, E. D. (2017). Klasifikasi Penyakit Ayam Menggunakan Metode Support Vector Machine. *Jurnal Ilmiah Pendidikan Teknik Elektro*, 2(1), 45. doi:10.30870/volt.v2i1.1316
- P2TPM Kemenkes Kesehatan RI. (2016). Tekanan darah tinggi (hipertensi). Diakses pada 25 Oktober 2023, dari <https://p2ptm.kemkes.go.id/uploads/2016/10/Tekanan-Darah-Tinggi-Hipertensi.pdf>
- P2TPM Kemenkes Kesehatan RI. (2019). Hari Hipertensi Dunia 2019 : “Know Your Number, Kendalikan Tekanan Darahmu dengan CERDIK.” <https://p2ptm.kemkes.go.id/tag/hari-hipertensi-dunia-2019-know-your-number-kendalikan-tekanan-darahmu-dengan-cerdik>. Diakses pada tanggal 15 Oktober 2023.
- Panindre, P., Gandhi, V., & Kumar, S. (2021). Instantaneous Heart Rate-Based Automated Monitoring Of Hypertension Using Machine Learning. In Proceedings of the 2021 IEEE International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 732-737. doi: 10.1109/ICCCIS51004.2021.9397126

- Peryanto, A., Yudhana, A., & Umar, R. (2020). Klasifikasi Citra Menggunakan Convolutional Neural Network Dan K-Fold Cross Validation. *Journal of Applied Informatics and Computing*, 4(1), 45-51. doi:10.30871/jaic.v4i1.2017
- Pratiwi, S. N. D., & Ulama, B. S. S. (2016). Klasifikasi Email Spam dengan Menggunakan Metode Support Vector Machine dan k-Nearest Neighbor. *Jurnal Sains Dan Seni ITS*, 5(2), 344–349.
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Journal of Computing and Information System*, 13(1), 76-84.
- Setiawati, I., Wibowo, A. P., & Hermawan, A. (2019). Pendahuluan tinjauan pustaka penelitian sebelumnya klasifikasi. *Jurnal of Information System Management*, 1(1), 13-17.
- Suyanto. (2019). *Data Mining Untuk Klasifikasi Dan Klasterisasi Data*. Bandung: Informatika.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining (3rd ed.)*. Harlow, England: Pearson Education Limited.
- Tino, M. D. F., Hasanah, H., & Santosa, T. D. (2023). Perbandingan Algoritma Support Vector Machines (Svm) Dan Neural Network Untuk Klasifikasi Penyakit Jantung. *INFOTECH journal*, 9(1), 232-235. doi: 10.31949/infotech.v9i1.5432

## LAMPIRAN

**Lampiran 1** Hasil Prediksi Klasifikasi Diagnosis Hipertensi Menggunakan SVM

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
1	0	0	0	0	0	0	0	160	90	42	1	1
2	0	0	0	0	0	0	0	134	96	52	1	1
3	0	0	0	0	0	0	0	137	88	40	1	1
4	0	0	0	0	0	0	0	233	118	54	1	1
5	0	0	0	0	0	0	0	202	93	78	1	1
6	0	0	0	0	0	0	0	110	70	27	0	0
7	1	0	1	0	0	1	0	130	98	37	1	0
8	0	0	0	0	0	0	0	132	82	64	0	0
9	0	0	0	0	0	0	0	169	104	51	1	1
10	0	0	0	0	0	0	0	143	86	54	1	1
11	0	0	0	0	0	0	0	163	89	54	1	1
12	0	1	0	1	1	1	0	140	80	59	0	0
13	0	0	0	0	0	0	0	135	88	22	1	1
14	0	0	0	1	0	0	0	141	90	55	1	1
15	0	0	0	0	0	1	0	121	78	35	0	0
16	0	1	1	1	0	0	0	150	90	68	1	1
17	0	0	0	0	0	1	0	150	99	73	1	1
18	0	0	0	0	1	0	0	110	70	31	0	0
19	0	0	0	0	1	1	0	150	89	60	1	1
20	0	1	0	0	0	1	0	113	85	40	0	0
21	0	1	0	0	0	0	0	190	99	65	1	1
22	0	0	0	0	0	0	0	96	74	38	0	0
23	0	0	0	0	0	1	0	155	52	67	0	0
24	0	0	0	0	0	0	0	159	89	74	1	1
25	0	0	0	0	0	0	0	130	90	35	1	1
26	0	1	0	0	0	0	0	130	80	29	0	0
27	0	0	1	0	0	0	0	133	70	39	0	0
28	0	1	0	0	0	0	0	126	91	29	0	0
29	0	0	0	0	0	0	0	175	78	77	0	0
30	0	0	0	0	0	1	0	133	90	59	1	1
31	0	0	0	0	0	0	0	160	99	52	1	1
32	0	0	0	0	0	0	0	144	86	42	1	1
33	0	1	0	0	0	1	0	100	71	43	0	0
34	0	0	0	0	0	0	0	111	74	64	0	0
35	0	1	0	0	0	1	0	106	84	40	0	0
36	0	0	0	0	0	0	0	140	89	54	1	1
37	0	0	0	0	0	0	0	94	66	41	0	0
38	0	0	0	0	0	0	0	167	96	56	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
39	0	0	0	1	0	0	0	114	65	28	0	0
40	0	0	0	0	0	1	0	135	85	41	1	1
41	0	0	0	0	0	0	0	136	94	52	1	1
42	0	0	0	0	0	0	0	117	67	55	0	0
43	0	0	0	0	0	0	0	188	64	68	0	0
44	0	1	0	0	0	0	0	142	85	47	1	1
45	0	0	0	0	0	0	0	108	59	86	0	0
46	0	0	0	0	0	0	0	130	79	44	0	0
47	0	1	0	0	0	1	0	115	78	25	0	0
48	0	0	0	0	0	0	0	120	85	23	0	0
49	0	0	0	0	0	0	0	182	100	68	1	1
50	0	1	0	0	0	0	0	186	90	61	1	1
51	0	0	0	0	0	1	0	134	85	41	1	1
52	0	0	0	0	1	0	0	140	90	47	1	1
53	0	0	1	0	0	0	0	145	88	54	1	1
54	0	0	0	0	0	0	0	130	78	33	0	0
55	0	0	0	0	0	1	0	142	88	43	1	1
56	0	0	0	0	1	1	0	160	80	70	0	0
57	0	0	0	0	0	0	0	196	100	48	1	1
58	0	0	0	0	0	0	0	110	70	55	0	0
59	0	0	0	0	0	0	0	134	80	62	0	0
60	0	0	0	0	0	1	0	162	90	55	1	1
61	0	1	1	0	0	1	0	120	78	43	0	0
62	0	1	0	1	1	0	0	142	96	55	1	1
63	0	0	0	0	0	0	0	100	66	30	0	0
64	0	0	0	0	0	0	0	130	80	52	0	0
65	0	0	1	1	1	1	0	147	86	37	1	0
66	0	0	0	0	0	0	0	146	86	69	1	1
67	0	0	0	0	0	0	0	129	82	35	0	0
68	0	0	0	0	0	0	0	127	79	35	0	0
69	0	1	0	0	0	0	0	176	121	84	1	1
70	0	0	0	0	0	0	0	142	96	55	1	1
71	0	0	0	0	0	1	0	150	85	62	1	1
72	0	0	0	0	0	0	0	134	88	22	1	1
73	0	1	1	0	1	1	0	150	90	67	1	1
74	0	0	0	1	0	1	0	120	80	27	0	0
75	0	0	1	0	0	0	0	197	100	48	1	1
76	0	0	0	1	0	0	0	91	68	26	0	0
77	0	0	0	0	0	0	0	139	87	61	1	1
78	0	0	0	0	1	0	0	180	100	68	1	1
79	0	1	0	0	0	1	0	94	74	26	0	0
80	0	0	0	0	0	0	0	125	80	65	0	0
81	0	0	0	1	1	0	0	149	91	53	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
82	0	0	0	0	0	1	0	138	86	65	1	1
83	0	1	0	0	0	1	0	120	80	38	0	0
84	0	0	0	0	1	0	0	128	73	30	0	0
85	0	0	0	0	0	0	0	101	64	22	0	0
86	0	1	0	0	1	0	0	180	60	44	0	0
87	0	0	0	0	0	0	0	124	60	42	0	0
88	0	0	0	0	0	0	0	132	90	59	1	1
89	0	0	0	0	0	0	0	156	83	61	0	0
90	0	1	1	0	1	1	0	119	75	53	0	0
91	0	0	0	0	0	0	0	108	81	44	0	0
92	0	0	1	0	1	0	0	135	80	57	0	0
93	0	0	1	1	0	0	0	140	90	70	1	1
94	0	0	0	0	0	0	0	193	106	56	1	1
95	0	1	0	1	1	1	0	120	84	43	0	0
96	0	0	0	0	0	1	0	118	73	55	0	0
97	0	0	0	0	0	0	0	193	106	56	1	1
98	0	0	0	0	1	0	0	174	114	54	1	1
99	0	0	0	0	0	0	0	140	89	55	1	1
100	0	0	0	0	1	1	0	140	90	40	1	1
101	0	1	0	0	0	1	0	101	59	34	0	0
102	0	0	0	0	1	0	0	120	60	72	0	0
103	0	1	0	0	0	0	0	130	90	35	1	1
104	0	0	0	0	0	0	0	120	75	29	0	0
105	0	0	0	0	0	0	0	130	80	54	0	0
106	0	0	0	0	0	0	0	100	80	23	0	0
107	0	0	0	1	0	0	0	135	80	43	0	0
108	0	0	0	1	0	0	0	145	88	53	1	1
109	0	1	0	0	0	0	0	114	81	30	0	0
110	0	1	0	0	0	1	0	120	60	74	0	0
111	0	1	0	0	0	1	0	93	50	27	0	0
112	0	1	0	0	0	1	0	140	90	45	1	1
113	0	0	0	0	0	0	0	143	91	62	1	1
114	0	0	0	0	0	0	0	100	70	36	0	0
115	0	0	0	0	0	0	0	130	80	54	0	0
116	0	1	0	0	0	0	0	157	101	54	1	1
117	0	0	1	0	1	0	0	171	106	57	1	1
118	0	0	0	0	0	0	0	108	71	69	0	0
119	0	1	0	1	0	0	0	157	88	71	1	1
120	0	0	1	0	1	1	0	140	70	103	0	0
121	0	0	0	0	0	0	0	171	77	73	0	0
122	0	0	0	0	0	0	0	130	93	34	1	1
123	0	0	0	0	0	0	0	104	54	69	0	0
124	0	0	0	0	0	0	0	151	85	41	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
125	0	0	1	0	1	0	0	119	68	24	0	0
126	0	0	0	0	0	0	0	143	86	38	1	1
127	0	0	0	0	0	0	0	128	90	32	0	0
128	0	1	0	0	1	1	0	158	93	37	1	1
129	0	0	0	0	0	0	0	150	85	45	1	1
130	0	0	0	0	0	0	0	131	89	49	1	1
131	0	0	0	0	0	0	0	145	85	44	1	1
132	0	0	0	0	0	0	0	104	69	51	0	0
133	0	0	0	0	0	0	0	105	68	35	0	0
134	0	0	0	0	0	0	0	191	107	59	1	1
135	0	0	0	0	0	0	0	122	83	64	0	0
136	0	0	0	0	1	1	0	148	91	54	1	1
137	0	0	0	1	0	0	0	141	91	55	1	1
138	0	0	0	0	0	0	0	175	115	55	1	1
139	0	0	0	0	0	0	0	147	88	40	1	1
140	0	0	0	0	0	0	0	224	114	52	1	1
141	0	0	0	0	0	0	0	142	98	39	1	1
142	0	0	1	0	1	0	0	117	71	24	0	0
143	0	1	0	0	0	0	0	130	90	35	1	1
144	0	0	0	0	0	0	0	120	70	31	0	0
145	0	0	0	0	1	0	0	120	80	56	0	0
146	0	1	0	0	0	0	0	180	80	77	0	0
147	0	0	1	1	1	0	0	140	80	48	0	0
148	0	0	0	0	0	0	0	100	60	72	0	0
149	0	1	0	1	1	0	0	143	94	57	1	1
150	0	1	0	1	0	1	0	147	66	84	0	0
151	0	0	0	0	1	0	0	128	55	45	0	0
152	0	1	0	1	0	1	0	214	96	51	1	1
153	0	0	0	0	0	1	0	130	90	58	1	1
154	0	0	0	0	0	0	0	133	84	53	0	0
155	0	1	0	0	0	0	0	110	70	30	0	0
156	0	0	0	0	0	0	0	129	70	42	0	0
157	0	1	0	0	1	0	0	159	90	43	1	1
158	0	0	0	0	0	0	0	150	88	40	1	1
159	0	0	0	0	1	1	0	100	70	24	0	0
160	0	0	0	0	0	0	0	135	80	38	0	0
161	0	0	0	0	0	0	0	143	100	36	1	1
162	0	0	0	0	0	0	0	135	89	45	1	1
163	0	0	0	0	0	0	0	130	80	53	0	0
164	0	0	0	0	0	0	0	166	93	49	1	1
165	0	0	0	0	0	0	0	128	86	48	0	0
166	0	1	1	1	0	1	0	155	90	60	1	1
167	0	0	0	0	0	0	0	179	100	68	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
168	0	0	0	0	0	0	0	120	83	26	0	0
169	0	0	0	1	0	1	0	130	80	45	0	0
170	0	0	0	0	0	0	0	140	69	69	0	0
171	0	0	0	0	0	0	0	130	90	64	1	1
172	0	0	0	1	0	1	0	140	72	63	0	0
173	0	0	0	0	0	0	0	130	80	33	0	0
174	0	0	0	0	1	0	0	133	89	28	1	1
175	0	0	0	0	0	0	0	92	75	26	0	0
176	0	0	0	0	0	0	0	100	60	36	0	0
177	0	1	0	0	0	1	0	137	70	67	0	0
178	1	0	0	0	0	0	0	98	54	30	0	0
179	0	0	0	0	0	0	0	140	78	55	0	0
180	0	0	0	0	0	0	0	133	78	49	0	0
181	0	0	1	1	0	0	0	150	90	75	1	1
182	0	0	0	0	0	0	0	118	76	35	0	0
183	0	0	0	0	0	1	0	120	80	31	0	0
184	0	0	0	0	0	0	0	130	90	34	1	1
185	1	0	1	0	0	1	0	176	84	74	0	0
186	0	0	0	0	1	0	0	192	101	58	1	1
187	0	0	1	0	1	1	0	150	89	50	1	1
188	0	0	0	0	0	0	0	100	80	37	0	0
189	0	0	0	0	0	0	0	147	111	53	1	1
190	0	1	0	0	0	1	0	100	70	31	0	0
191	0	0	1	0	0	0	0	148	99	53	1	1
192	0	0	0	0	0	1	0	180	101	59	1	1
193	0	0	0	0	0	1	0	135	81	62	0	0
194	0	1	0	0	0	1	0	117	73	48	0	0
195	0	0	0	0	0	0	0	145	88	55	1	1
196	0	1	0	0	0	0	0	150	80	53	0	0
197	0	0	0	0	0	0	0	142	89	62	1	1
198	0	1	0	1	0	1	0	140	80	36	0	0
199	0	0	0	0	0	0	0	106	58	35	0	0
200	0	1	1	0	0	0	0	121	59	38	0	0
201	0	1	1	1	0	0	0	160	82	66	0	1
202	0	0	0	0	0	0	0	100	60	66	0	0
203	0	1	0	0	0	1	0	111	63	54	0	0
204	0	0	0	1	0	0	0	116	73	33	0	0
205	0	0	0	0	0	0	0	155	59	45	0	0
206	0	0	0	0	0	1	0	150	98	73	1	1
207	0	0	1	0	1	0	0	141	74	63	0	0
208	0	0	0	0	0	0	0	140	90	56	1	1
209	0	1	0	0	0	0	0	148	79	69	0	0
210	0	0	0	0	1	0	0	160	67	71	0	0

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
211	0	0	0	0	1	0	0	120	73	29	0	0
212	0	0	0	0	0	0	0	110	65	54	0	0
213	0	0	0	0	0	1	0	160	90	60	1	1
214	0	1	0	0	0	1	0	120	80	64	0	0
215	0	0	0	0	0	0	0	111	73	50	0	0
216	0	0	1	0	0	0	0	140	89	48	1	1
217	0	1	0	0	0	0	0	161	95	76	1	1
218	0	1	0	1	0	1	0	140	73	70	0	0
219	0	0	1	1	0	1	0	165	93	66	1	1
220	0	0	0	0	0	0	0	141	90	56	1	1
221	0	1	0	0	0	1	0	202	121	62	1	1
222	0	0	0	0	1	0	0	158	94	63	1	1
223	0	0	0	0	0	0	0	110	60	34	0	0
224	0	0	0	0	0	0	0	120	80	35	0	0
225	0	0	0	0	0	0	0	160	94	61	1	1
226	0	1	0	0	0	1	0	148	82	64	0	0
227	0	1	0	1	1	1	0	104	82	32	0	0
228	0	0	0	0	0	0	0	168	96	56	1	1
229	0	1	0	0	1	0	0	165	100	64	1	1
230	0	0	0	0	0	0	0	136	79	34	0	0
231	0	1	0	0	0	0	0	147	82	53	0	0
232	0	1	0	0	0	0	0	131	90	49	1	1
233	0	0	1	0	0	0	0	140	90	41	1	1
234	0	0	0	0	0	0	0	136	87	30	1	1
235	0	0	0	0	0	0	0	119	78	27	0	0
236	0	0	0	0	0	0	0	127	74	22	0	0
237	0	1	0	1	0	0	0	116	77	33	0	0
238	0	0	0	0	0	0	0	130	90	35	1	1
239	0	1	1	0	0	1	0	100	80	23	0	0
240	0	0	0	0	0	0	0	160	89	71	1	1
241	0	0	0	0	0	0	0	130	90	62	1	1
242	0	0	1	0	0	0	0	156	79	66	0	0
243	0	1	0	0	0	1	0	160	80	65	0	0
244	0	0	1	0	0	1	0	118	71	57	0	0
245	0	0	0	0	0	0	0	121	82	65	0	0
246	0	0	1	0	0	1	0	161	90	48	1	1
247	0	0	0	0	0	0	0	135	90	45	1	1
248	0	1	0	0	0	1	0	132	98	42	1	1
249	0	0	0	0	0	0	0	131	89	49	1	1
250	0	1	0	0	0	0	0	172	89	67	1	1
251	0	0	0	0	0	0	0	112	81	38	0	0
252	0	1	0	0	0	1	0	118	71	56	0	0
253	0	0	0	0	0	0	0	140	90	56	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
254	0	0	0	0	0	1	0	130	80	42	0	0
255	0	1	0	0	1	1	0	149	73	69	0	0
256	0	0	0	0	0	0	0	110	80	48	0	0
257	0	0	0	0	0	0	0	134	87	22	1	1
258	0	0	0	0	0	0	0	128	75	34	0	0
259	0	1	0	1	1	0	0	144	87	70	1	1
260	0	0	0	0	0	0	0	130	70	68	0	0
261	0	0	1	1	0	0	0	162	103	59	1	1
262	0	0	0	0	0	0	0	130	90	35	1	1
263	0	0	0	0	0	0	0	158	86	70	1	1
264	0	0	0	0	0	0	0	144	87	70	1	1
265	0	0	0	1	0	0	0	164	110	38	1	1
266	0	0	0	1	0	0	0	158	87	75	1	1
267	1	1	0	0	0	1	0	130	70	41	0	0
268	1	0	0	0	1	0	0	183	88	72	1	1
269	0	1	0	0	1	1	0	159	90	41	1	1
270	0	0	0	0	0	0	0	157	99	61	1	1
271	0	0	0	1	0	0	0	146	88	79	1	1
272	0	0	0	0	0	0	0	90	60	26	0	0
273	0	0	0	0	1	1	0	160	69	70	0	0
274	0	0	0	0	0	0	0	130	70	42	0	0
275	0	0	0	0	0	0	0	161	98	52	1	1
276	0	1	0	0	0	1	0	97	63	45	0	0
277	0	0	0	0	0	0	0	133	86	42	1	1
278	0	0	1	1	1	0	0	158	79	61	0	0
279	0	0	0	0	0	0	0	121	65	42	0	0
280	0	0	0	0	0	0	0	157	96	62	1	1
281	0	0	0	0	0	0	0	140	90	47	1	1
282	0	1	0	0	0	0	0	164	90	79	1	1
283	0	1	0	0	0	1	0	150	100	53	1	1
284	0	0	0	0	0	0	0	142	98	40	1	1
285	0	0	0	0	0	1	0	190	110	66	1	1
286	0	0	0	0	0	0	0	140	90	40	1	1
287	0	0	0	0	0	1	0	124	78	30	0	0
288	0	0	0	0	0	0	0	162	99	79	1	1
289	0	0	0	0	1	0	0	135	80	45	0	0
290	0	0	0	0	0	0	0	167	92	48	1	1
291	0	1	0	0	0	1	0	160	85	53	1	1
292	0	1	1	0	0	1	0	125	72	56	0	0
293	0	0	0	0	0	0	0	148	111	52	1	1
294	0	0	0	0	0	1	0	140	80	56	0	0
295	0	0	0	0	0	0	0	93	65	32	0	0
296	0	0	1	0	1	1	0	147	86	37	1	1

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
297	0	0	0	0	0	0	0	131	77	65	0	0
298	0	0	0	0	0	0	0	151	91	61	1	1
299	0	0	0	0	0	0	0	166	94	50	1	1
300	0	0	0	0	0	1	0	120	79	38	0	0
301	0	0	0	1	1	1	0	148	92	55	1	1
302	0	0	0	0	0	0	0	134	87	22	1	1
303	0	0	0	0	0	0	0	145	85	44	1	1
304	0	0	0	0	0	1	0	110	79	23	0	0
305	0	0	0	1	0	0	0	114	66	71	0	0
306	0	1	0	1	0	0	0	97	60	21	0	0
307	0	0	0	0	0	1	0	95	80	55	0	0
308	0	1	0	0	0	0	0	130	80	58	0	0
309	0	0	0	0	0	0	0	119	81	37	0	0
310	0	1	0	1	0	1	0	148	90	82	1	1
311	0	0	0	1	0	0	0	142	100	56	1	1
312	0	1	0	0	1	0	0	160	70	76	0	0
313	0	0	0	0	0	0	0	112	76	22	0	0
314	0	1	1	0	0	0	0	160	88	59	1	1
315	0	0	0	0	0	0	0	130	80	43	0	0
316	0	0	0	0	0	0	0	107	76	24	0	0
317	0	1	0	0	0	0	0	151	99	56	1	1
318	0	1	0	1	0	0	0	158	83	66	0	0
319	0	0	1	0	1	0	0	150	89	52	1	1
320	0	1	0	0	0	1	0	171	87	65	1	1
321	0	0	1	0	0	1	0	156	75	56	0	0
322	0	1	0	1	0	1	0	158	80	57	0	0
323	0	0	0	0	0	0	0	126	79	77	0	0
324	0	0	0	0	0	0	0	140	90	46	1	1
325	0	1	0	1	0	0	0	162	90	75	1	1
326	0	1	1	1	1	1	0	184	109	67	1	1
327	0	0	0	0	0	0	0	145	89	48	1	1
328	0	0	0	0	0	0	0	124	98	57	0	0
329	0	1	0	0	1	0	0	177	70	83	0	0
330	0	0	0	0	0	0	0	109	107	60	0	0
331	0	0	1	0	0	0	0	118	77	35	0	0
332	0	1	0	0	1	1	0	130	80	24	0	0
333	0	0	1	0	0	1	0	156	52	67	0	0
334	0	0	0	1	0	0	0	154	90	57	1	1
335	0	0	0	0	1	0	0	153	90	67	1	1
336	0	0	1	0	1	1	0	156	103	61	1	1
337	0	1	0	0	0	1	0	139	83	37	0	1
338	0	0	0	1	0	1	0	146	88	79	1	1
339	0	1	0	0	0	1	0	106	155	57	0	0

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
340	0	0	0	0	0	0	0	116	80	48	0	0
341	0	0	0	0	0	1	0	150	98	73	1	1
342	0	0	0	0	1	1	0	130	80	42	0	0
343	0	0	0	0	0	0	0	158	92	46	1	1
344	0	1	0	0	0	1	0	129	81	45	0	0
345	0	1	0	1	0	0	0	146	89	68	1	1
346	0	1	1	0	1	0	0	158	95	65	1	1
347	0	0	0	0	0	0	0	130	80	50	0	0
348	0	0	0	1	0	0	0	157	89	48	1	1
349	0	0	0	0	0	0	0	158	87	74	1	1
350	0	0	0	1	0	0	0	140	70	54	0	0
351	0	0	0	1	0	1	0	161	73	60	0	0
352	0	0	0	0	0	0	0	136	90	56	1	1
353	0	0	1	0	0	0	0	160	87	63	1	1
354	0	0	0	0	0	0	0	123	76	43	0	0
355	0	0	0	0	0	0	0	143	92	59	1	1
356	0	0	0	0	0	0	0	142	91	62	1	1
357	0	1	0	0	0	0	0	130	80	54	0	0
358	0	0	1	0	1	0	0	149	90	52	1	1
359	0	0	0	0	0	0	0	132	90	44	1	1
360	0	0	0	0	0	0	0	143	85	54	1	1
361	0	0	1	1	0	0	0	135	75	65	0	0
362	0	0	0	0	0	0	0	136	85	30	1	1
363	0	0	0	0	0	0	0	156	100	61	1	1
364	0	0	0	0	0	0	0	156	85	70	1	1
365	0	0	0	0	0	1	0	100	80	35	0	0
366	0	0	1	0	1	0	0	150	89	52	1	1
367	0	0	0	0	0	0	0	123	67	58	0	0
368	0	0	1	0	0	1	0	129	60	80	0	0
369	0	1	1	0	0	0	0	140	70	65	0	0
370	0	0	0	0	0	0	0	172	88	65	1	1
371	0	0	0	0	0	0	0	120	80	23	0	0
372	0	0	0	0	0	0	0	168	89	68	1	1
373	0	0	0	1	0	0	0	142	88	33	1	1
374	0	0	0	0	0	0	0	137	90	33	1	1
375	0	0	0	0	0	0	0	140	88	59	1	1
376	0	0	0	0	0	0	0	141	90	56	1	1
377	0	0	0	0	0	0	0	130	80	44	0	0
378	0	1	0	0	0	0	0	105	81	57	0	0
379	0	0	0	0	0	0	0	171	110	56	1	1
380	0	1	0	0	0	0	0	150	74	66	0	0
381	0	0	0	0	0	0	0	149	75	55	0	0
382	0	0	1	1	1	1	0	125	83	29	0	0

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	Prediksi
383	0	0	0	0	0	0	0	132	68	41	0	0
384	0	0	0	0	0	0	0	135	80	56	0	0
385	0	0	0	0	0	0	0	150	89	60	1	1
386	0	0	0	1	1	0	0	170	88	58	1	1
387	0	0	0	0	1	0	0	147	91	52	1	1
388	0	1	1	0	0	0	0	148	98	56	1	1
389	0	1	0	0	0	0	0	100	70	53	0	0
390	0	0	0	0	1	0	0	130	90	22	1	1
391	0	0	0	0	0	0	0	145	85	52	1	1
392	0	0	0	0	0	0	0	143	86	38	1	1
393	0	0	0	0	1	0	0	150	90	49	1	1
394	0	0	0	0	0	0	0	132	88	41	1	1
395	0	1	0	0	0	0	0	130	93	56	1	1
396	0	0	0	0	0	0	0	135	93	59	1	1
397	0	1	0	0	0	1	0	157	96	63	1	1
398	0	0	0	0	0	0	0	135	85	41	1	1
399	0	1	0	0	0	0	0	124	92	35	0	0

Keterangan :

- X1 : Merokok
- X2 : Kurang Aktivitas Fisik
- X3 : Gula Berlebihan
- X4 : Garam Berlebihan
- X5 : Lemak Berlebihan
- X6 : Kurang Makan Buah dan Sayur
- X7 : Konsumsi Alkohol
- X8 : Tekanan Sistolik
- X9 : Tekanan Diastolik
- X10 : Usia
- Y : Diagnosis Hipertensi

## Lampiran 2 Script Code Python

```

#Import Library
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import OneHotEncoder
from sklearn import preprocessing
from sklearn.svm import SVC
from sklearn import metrics
from sklearn import svm, datasets
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.metrics import accuracy_score, recall_score,
precision_score, confusion_matrix, roc_auc_score, f1_score
from sklearn.metrics import precision_recall_fscore_support as
score
import tensorflow as tf

#Import Data
dataset = pd.read_excel("skrining _skripsi hilda.xlsx")
dataset.info()
dataset.describe() #menampilkan statistik deskriptif data numerik
#menampilkan pie chart
plt.figure(figsize = (10, 8))
plt.pie(pd.value_counts(dataset['Diagnosis']),
        labels = ['Tidak Hipertensi', 'Hipertensi'],
        autopct = '%.2f%%',
        textprops = {'size' : 'x-large',
                    'fontweight' : 'bold',
                    'rotation' : '0',
                    'color' : 'w'})

plt.legend()
# plt.title('Persentase hipertensi', fontsize = 15, fontweight =
'bold')
plt.show()

# Categorical Bivariate Analysis
cat = ['MEROKOK', 'KURANG AKTIFITAS FISIK', 'GULA BERLEBIHAN',
'GARAM BERLEBIHAN', 'LEMAK BERLEBIHAN', 'KURANG MAKAN BUAH DAN
SAYUR', 'KONSUMSI ALKOHOL']
plt.figure(figsize=(25, 10)) # Adjust the figure size as needed

```

```

for i, attribute in enumerate(cat, 1):
    plt.subplot(2, 4, i) # 2 rows, 4 columns, adjust as needed
    sns.countplot(data=dataset, x=attribute, hue='Diagnosis')
    plt.title(f'Countplot for {attribute}')

plt.tight_layout()
plt.show()

#Data Preprocessing
#Data Transformation
dataset['MEROKOK'] = dataset['MEROKOK'].map({'YA': 1, 'TIDAK':
0})
dataset['KURANG AKTIFITAS FISIK'] = dataset['KURANG AKTIFITAS
FISIK'].map({'YA': 1, 'TIDAK': 0})
dataset['GULA BERLEBIHAN'] = dataset['GULA
BERLEBIHAN'].map({'YA': 1, 'TIDAK': 0})
dataset['LEMAK BERLEBIHAN'] = dataset['LEMAK
BERLEBIHAN'].map({'YA': 1, 'TIDAK': 0})
dataset['KURANG MAKAN BUAH DAN SAYUR'] = dataset['KURANG MAKAN
BUAH DAN SAYUR'].map({'YA': 1, 'TIDAK': 0})
dataset['KONSUMSI ALKOHOL'] = dataset['KONSUMSI
ALKOHOL'].map({'YA': 1, 'TIDAK': 0})
dataset['GARAM BERLEBIHAN'] = dataset['GARAM
BERLEBIHAN'].map({'YA': 1, 'TIDAK': 0})

#Mengubah Data Tanggal Lahir Menjadi Usia
dataset["TANGGAL LAHIR *"] = pd.to_datetime(dataset["TANGGAL
LAHIR *"], format="%d-%m-%Y",errors="coerce")
today_date = datetime.today().replace(day=1, month=1).date() #
Konversi ke datetime.date
dataset["Usia"] = (today_date - dataset["TANGGAL LAHIR
*"].dt.date).apply(lambda x: x.days) // 365.25

#Splitting Data
X = dataset.drop(["Diagnosis"], axis = 1)
Y = dataset["Diagnosis"]

# Oversampling Data
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split

sm = SMOTE(random_state=42)
X, Y = sm.fit_resample(X, Y)

X_train, X_test, Y_train, Y_test = train_test_split(X, Y,

```

```

# Percobaan C = 0,01 Menggunakan Kernel Linear
svm_linear_1 = svm.SVC(kernel='linear', C=0.01).fit(X_train,
Y_train)
svm_linear_1_pred = svm_linear_1.predict(X_test)
num_support_vectors = svm_linear_1.support_vectors_
print("Banyak Support Vector:")
print(len(num_support_vectors))
print("=====")
print(classification_report(Y_test,svm_linear_1_pred))
print("Akurasi =",accuracy_score(svm_linear_1_pred, Y_test))
cm = tf.math.confusion_matrix(labels=Y_test,predictions =
svm_linear_1_pred)

print(cm)
plt.figure(figsize = (10,7))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Nilai Kebenaran')
plt.ylabel('Prediksi Diagnosis Hipertensi, 0: TIdak Hipertensi/1:
Hipertensi')

# K-Fold Cross Val
def get_metrics(estimator, X, y):
    precision = cross_val_score(estimator, X, y,
cv=KFold(n_splits=5), scoring='precision')
    recall = cross_val_score(estimator, X, y,
cv=KFold(n_splits=5), scoring='recall')
    accuracy = cross_val_score(estimator, X, y,
cv=KFold(n_splits=5), scoring='accuracy')
    return precision, recall, accuracy

precision, recall, accuracy = get_metrics(svm_rbf_4_1, X_train,
Y_train)
df = pd.DataFrame({
    'Precision': precision,
    'Recall': recall,
    'Accuracy': accuracy
})
print(df)

```

## RIWAYAT HIDUP



Hilda Zaqya Elnaz Putri atau lebih dikenal Hilda, lahir di Kota Bekasi pada 20 Februari 2003. Penulis merupakan anak kedua dari tiga bersaudara dari pasangan Bapak Nazri dan Ibu Elvasari. Penulis telah menempuh pendidikan mulai dari TK Kamila yang lulus pada tahun 2008, dilanjutkan menempuh pendidikan sekolah dasar di MIT Al-Hamid Jakarta dan lulus pada tahun 2014. Kemudian penulis melanjutkan pendidikan sekolah menengah pertama di MTS Al-Hamid Jakarta dan lulus pada tahun 2017. Selanjutnya menempuh pendidikan sekolah menengah atas di MAS Al-Hamid Jakarta dan lulus pada tahun 2020. Pada tahun yang sama, penulis melanjutkan pendidikan di Universitas Islam Negeri Maulana Malik Ibrahim Malang pada Program Studi Matematika, Fakultas Sains dan Teknologi. Selama menempuh pendidikan tinggi, penulis aktif mengikuti beberapa kegiatan. Penulis bergabung dalam organisasi HMPS “Integral” Matematika selama dua periode sebagai anggota Divisi Pengembangan Minat dan Bakat dan Sekretaris Umum. Penulis juga pernah menjadi asisten praktikum Pemrograman Komputer selama dua semester. Selain itu, penulis juga mengikuti kegiatan di luar kampus seperti pelatihan dan seminar.



**BUKTI KONSULTASI SKRIPSI**

Nama : Hilda Zaqya Elnaz Putri  
NIM : 200601110112  
Fakultas/Jurusan : Sains dan Teknologi/Matematika  
Judul Skripsi : Implementasi Metode *Support Vector Machine* pada  
Klasifikasi Diagnosis Penyakit Hipertensi  
Pembimbing I : Hisyam Fahmi, M. Kom.  
Pembimbing II : Juhari, M. Si.

No	Tanggal	Hal	Tanda Tangan
1.	12 Oktober 2023	Konsultasi Bab I, II, dan III	1.
2.	20 Oktober 2023	Konsultasi Kajian Agama	2.
3.	23 Oktober 2023	Konsultasi Revisi Kajian Agama	3.
4.	25 Oktober 2023	ACC Kajian Agama Bab I dan II	4.
5.	26 Oktober 2023	Konsultasi Revisi Bab III	5.
6.	02 November 2023	ACC Bab I, II, dan III	6.
7.	21 November 2023	ACC Seminar Proposal	7.
8.	04 Desember 2023	Konsultasi Revisi Seminar Proposal	8.
9.	05 Februari 2024	Konsultasi Bab IV dan V	9.
10.	12 Februari 2024	Konsultasi Bab IV dan V	10.
11.	16 Februari 2024	Konsultasi Kajian Agama Bab IV	11.
12.	20 Februari 2024	ACC Kajian Agama Bab IV	12.
13.	04 Maret 2024	Konsultasi Bab IV dan V	13.
14.	07 Maret 2024	ACC Bab IV dan V	14.
15.	14 Maret 2024	ACC Seminar Hasil	15.



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

16.	25 Maret 2024	Konsultasi Revisi Seminar Hasil	16.
17.	26 April 2024	ACC Matriks Revisi Seminar Hasil	17.
18.	15 Mei 2024	ACC Sidang Skripsi	18.
19.	30 Mei 2024	ACC Revisi Akhir	19.

Malang, 30 Mei 2024

Mengetahui,

Ketua Program Studi Matematika

Dr. Elly Susanti, M.Sc.

NIP. 19741129 200012 2 005