

**KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM) BERBASIS FITUR
KOMBINASI HASIL SELEKSI *RECURSIVE FEATURE
ELIMINATION* (RFE) DAN *CHI-SQUARE***

SKRIPSI

Oleh:
YUWATSIQUL AQWAM
NIM. 210605110160



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM) BERBASIS FITUR
KOMBINASI HASIL *SELEKSI RECURSIVE FEATURE
ELIMINATION (RFE)* DAN *CHI-SQUARE***

SKRIPSI

Diajukan kepada:

Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :

YUWATSIQUL AQWAM
NIM. 210605110160

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN


**KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN
SUPPORT VECTOR MACHINE (SVM) BERBASIS FITUR
KOMBINASI HASIL SELEKSI *RECURSIVE FEATURE
ELIMINATION* (RFE) DAN *CHI-SQUARE***

SKRIPSI


Oleh:
YUWATSIQUL AQWAM
NIM. 210605110160

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 24 Oktober 2025

Pembimbing I,

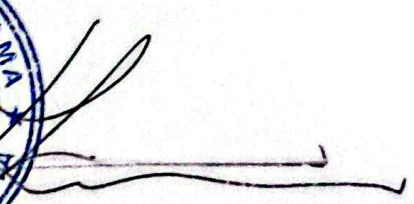

Dr. Irwan Budi Santoso, M.Kom
NIP. 19770103 201101 1 004

Pembimbing II,


Nur Fitriyah Ayu Tunjung Sari, M.Cs
NIP. 19911226 202012 2 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Nggeri Maulana Malik Ibrahim Malang




Supriyono, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI PENYAKIT KARDIOVASKULAR MENGGUNAKAN *SUPPORT VECTOR MACHINE (SVM)* BERBASIS FITUR KOMBINASI HASIL SELEKSI *RECURSIVE FEATURE ELIMINATION (RFE)* DAN *CHI-SQUARE*

SKRIPSI

Oleh:

YUWATSIQUL AQWAM
NIM. 210605110160

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 5 November 2025

Susunan Dewan Penguji

Ketua Penguji	: <u>Dr. M. Amin Hariyadi, M.T</u> NIP. 19670018 200501 1 001
Anggota Penguji I	: <u>A'la Syauqi, M.Kom</u> NIP. 19771201 200801 1 007
Anggota Penguji II	: <u>Dr. Irwan Budi Santoso, M.Kom</u> NIP. 19770103 201101 1 004
Anggota Penguji III	: <u>Nur Fitriyah Ayu Tunjung Sari, M.Cs</u> NIP. 19911226 202012 2 001

()
()
()
()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Supriyono, M. Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Yuwatsiqul Aqwam
NIM : 210605110160
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Klasifikasi Penyakit Kardiovaskular Menggunakan
*Support Vector Machine (SVM) Berbasis Fitur
Kombinasi Hasil Seleksi Recursive Feature
Elimination (RFE) Dan Chi-Square*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang,
Yang membuat pernyataan,



Yuwatsiqul Aqwam
NIM.210605110160

MOTTO

*"Apa yang melewatkan ku tidak akan menjadi takdirku dan apa yang
ditakdirkan untukku tidak akan pernah melewatkan ku."*

(Umar bin Khattab)

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, nikmat, dan kekuatan yang diberikan sehingga skripsi ini dapat terselesaikan. Shalawat serta salam kepada Rasulullah Shallallahu ‘alaihi wasallam, yang telah membawa kita dari zaman jahiliyah menuju addinul Islam. Karya ini penulis persembahkan dengan penuh cinta dan rasa hormat kepada:

Keluarga tercinta,

Bapak Rudy dan Ibu Triyana serta Saudara-saudara saya yang selalu menyertai langkah penulis dengan doa, kasih sayang, motivasi, dan keteladanan yang tiada ternilai.

Dosen pembimbing serta para pendidik,

yang dengan tulus membimbing, memberi ilmu, serta menjadi cahaya penuntun dalam proses akademik ini.

Sahabat dan rekan seperjuangan,

yang menjadi tempat berbagi cerita, tawa, serta penguat dalam perjalanan penuh tantangan.

Untuk diri sendiri,

sebagai pengingat bahwa perjuangan, kesabaran, dan usaha tidak pernah sia-sia.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Alhamdulillahirabbilalamin, segala puji dan syukur senantiasa penulis panjatkan ke hadirat Allah SWT subhanahu wa ta'ala atas berkat Rahmat, serta hidayahNya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Penyakit Kardiovaskular Menggunakan *Support Vector Machine* (SVM) Berbasis Fitur Kombinasi Hasil Seleksi *Recursive Feature Elimination* (RFE) Dan *Chi-Square*” dengan baik dan lancar. Shalawat dan salam senantiasa penulis haturkan kepada Rasulullah Muhammad SAW., sosok mulia yang menjadi panutan, yang telah menuntun umat manusia keluar dari masa kejahilan menuju era cahaya kebenaran, ajaran Islam, dan kemajuan ilmu pengetahuan yang masih dapat kita nikmati hingga saat ini.

Skripsi ini disusun sebagai salah satu syarat akademik untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Dalam perjalanan penyusunan karya tulis ini, penulis memperoleh begitu banyak dukungan, bantuan, doa, serta motivasi dari berbagai pihak, baik yang terlibat langsung maupun secara tidak langsung. Dengan penuh ketulusan, penulis ingin menyampaikan penghargaan dan rasa terima kasih yang mendalam kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang, atas kebijakan visioner dalam mengembangkan sarana dan prasarana pembelajaran yang mendukung terciptanya kualitas pendidikan yang unggul.

2. Dr. H. Agus Mulyono, S.Pd., M.Kes., selaku Dekan Fakultas Sains dan Teknologi, atas kepemimpinan yang penuh inspirasi serta dedikasi dalam meningkatkan mutu pendidikan di lingkungan fakultas.
3. Supriyono, M. Kom., selaku Ketua Program Studi Teknik Informatika, atas segala motivasi, dorongan, serta arahan yang senantiasa diberikan selama masa studi hingga penyelesaian karya ilmiah ini.
4. Dr. Irwan Budi Santoso, M.Kom., selaku Dosen Pembimbing I, atas kesabaran, bimbingan, arahan, serta nasihat yang begitu berarti dan berharga dalam proses penelitian hingga penyusunan skripsi ini.
5. Nur Fitriyah Ayu Tunjung Sari, M.Cs., selaku Dosen Pembimbing II, atas ketelatenan, saran, serta bimbingan yang sangat membantu dalam menyempurnakan penelitian ini.
6. Dr. Ir. M. Amin Hariyadi, M.T., selaku penguji I, dan A'la Syauqi, M.Kom., selaku penguji II, atas saran, kritik, masukan, serta bimbingan yang berharga demi kesempurnaan karya tulis ini.
7. Prof. Dr. H. Muhammad Faisal, S.Kom., M.T., selaku Dosen Wali, atas perhatian, arahan, serta bimbingan yang senantiasa diberikan kepada penulis sepanjang masa perkuliahan.
8. Nia Faricha, S.Si selaku admin Program Studi Teknik Informatika, atas kesabaran dalam memberikan pelayanan, informasi, dan bantuan administratif baik dalam perkuliahan maupun proses penyusunan skripsi.
9. Seluruh Dosen, Admin, Laboran, serta Civitas Akademika Program Studi Teknik Informatika yang telah berkontribusi besar dengan memberikan

ilmu pengetahuan, pengalaman, serta arahan selama penulis menempuh pendidikan.

10. Kedua orang tua tercinta, Bapak Rudy Purwadi dan Ibu Triyana Noor Hidayati, yang dengan penuh kasih sayang, doa yang tak pernah putus, serta dukungan baik moral maupun material, menjadi sumber kekuatan terbesar penulis. Ucapan terima kasih juga penulis sampaikan kepada saudara penulis, I'tishomul Hanif dan Yutsabitul Aqdam, atas dorongan, semangat, dan motivasi yang diberikan sehingga penulis mampu melewati berbagai rintangan.
11. Seluruh sahabat dan keluarga besar Teknik Informatika, khususnya Angkatan 2021 “ASTER,” atas kebersamaan, ilmu, pengalaman, serta semangat yang selalu menjadi penyemangat dalam menuntaskan perjalanan ini. Semoga ikatan persaudaraan ini senantiasa terjaga dan kita semua dapat menggapai cita-cita yang diharapkan.
12. Semua pihak yang tidak dapat disebutkan satu per satu, namun telah memberikan dukungan, doa, serta bantuan dalam berbagai bentuk yang sangat berarti bagi penulis.
13. Dan yang terakhir, kepada diri sendiri, yang tetap bertahan dalam perjuangan, tidak menyerah di tengah keterbatasan, serta terus berusaha untuk melangkah maju. Terima kasih telah menunjukkan bahwa batasan hanyalah ada dalam pikiran, dan bahwa kemampuan diri jauh lebih besar dari yang pernah dibayangkan.

Akhir kata, penulis sepenuhnya menyadari bahwa skripsi ini masih jauh dari kata sempurna. Namun, besar harapan penulis semoga karya ini dapat memberikan manfaat, menjadi amal jariyah di sisi Allah Subhanahu wa Ta'ala, serta memberi kontribusi nyata bagi pengembangan ilmu pengetahuan dan praktik di bidang yang relevan.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 20 November 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
BAB II STUDI PUSTAKA	8
2.1 Penelitian Klasifikasi	8
2.2 Penyakit Kardiovaskular	14
2.3 <i>Support Vector Machine</i>	15
2.4 Seleksi Fitur <i>Chi-Square</i>	17
2.5 Seleksi Fitur <i>Recursive Feature Elimination (RFE)</i>	18
BAB III DESAIN DAN IMPLEMENTASI	21
3.1 Desain Penelitian	21
3.2 Pengumpulan data	22
3.3 Desain Sistem	24
3.3.1 <i>Preprocessing</i>	25
3.3.2 <i>Chi-Square</i>	27

3.3.3 <i>Recursive Feature Elimination</i>	29
3.4 <i>Support Vector Machine</i>	32
3.5 Skenario Uji Coba	39
3.6 Evaluasi	41
BAB IV HASIL DAN PEMBAHASAN	44
4.1 Hasil <i>Preprocessing Data</i>	44
4.1.1 Hasil <i>Data Cleaning</i>	44
4.1.2 Hasil Normalisasi Data	45
4.2 Hasil <i>Chi-square</i>	47
4.3 Hasil <i>Recursive Feature Elimination (RFE)</i>	48
4.4 Hasil Uji Coba	51
4.4.1 Hasil Uji Coba Tanpa Seleksi Fitur	51
4.4.2 Hasil Uji Coba Dengan Seleksi Fitur	59
4.5 Pembahasan	67
4.6 Integrasi Islam	75
BAB V KESIMPULAN DAN SARAN	79
5.1 Kesimpulan	79
5.2 Saran	80
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>support vector machine</i>	16
Gambar 3.1 Prosedur penelitian	22
Gambar 3.2 Desain sistem	24
Gambar 3.3 Alur Preprocessing	25
Gambar 3.4 Flowchart Seleksi Fitur <i>Chi-square</i>	28
Gambar 3.5 Flowchart Seleksi Fitur <i>Recursive Feature Elimination</i>	30
Gambar 3.6 Flowchart <i>Sequential Minimal Optimization</i>	36
Gambar 3.7 <i>5-fold cross-validation visualization</i>	40
Gambar 4.1 Kurva RFECV - Akurasi vs Jumlah Fitur yang Dipilih	49
Gambar 4.2 Confusion Matrix Kernel Linear - Semua Fitur	53
Gambar 4.3 Confusion Matrix Kernel RBF - Semua Fitur	54
Gambar 4.4 Confusion Matrix Kernel Polinomial Degree 3 - Semua Fitur	55
Gambar 4.5 Confusion Matrix Kernel Polinomial Degree 5 - Semua Fitur	57
Gambar 4.6 Confusion Matrix Kernel Polinomial Degree 7 - Semua Fitur	58
Gambar 4.7 Confusion Matrix Kernel Linear - Seleksi Fitur	60
Gambar 4.8 Confusion Matrix Kernel RBF - Seleksi Fitur	61
Gambar 4.9 Confusion Matrix Kernel Polinomial Degree 3 - Seleksi Fitur	63
Gambar 4.10 Confusion Matrix Kernel Polinomial Degree 5 - Seleksi Fitur	64
Gambar 4.11 Confusion Matrix Kernel Polinomial Degree 7 - Seleksi Fitur	66
Gambar 4.12 Perbandingan Metrik Evaluasi Tanpa Seleksi Fitur	68
Gambar 4.13 Perbandingan Metrik Evaluasi dengan Seleksi Fitur	69
Gambar 4.14 Perbandingan Akurasi pada Setiap Kernel	71
Gambar 4.15 Perbandingan Standart Deviasi Akurasi antar Kernel	73

DAFTAR TABEL

Tabel 2.1 Penelitian terkait	13
Tabel 3.1 Contoh sampel dataset	23
Tabel 3.2 Deskripsi Fitur.....	23
Tabel 3.3 <i>Confusion Matrix</i>	42
Tabel 4.1 Sampel Data sebelum normalisasi	45
Tabel 4.2 Sampel Data Sesudah Normalisasi	46
Tabel 4.3 Hasil uji <i>Chi-square</i>	47
Tabel 4.4 Akurasi Cross-Validation terhadap Jumlah Fitur yang Dipilih	48
Tabel 4.5 Fitur Terpilih dari RFECV	50
Tabel 4.6 Hasil Interseksi Fitur dari Metode Chi-Square dan RFE	50
Tabel 4.7 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Linear	52
Tabel 4.8 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel RBF	53
Tabel 4.9 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial 3	55
Tabel 4.10 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial 5	56
Tabel 4.11 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial 7	57
Tabel 4.12 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Linear.....	59
Tabel 4.13 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel RBF	60
Tabel 4.14 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial 3	62
Tabel 4.15 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial 5	63
Tabel 4.16 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial 7	65
Tabel 4.17 Performa Model SVM Tanpa Seleksi Fitur	67
Tabel 4.18 Performa Model SVM dengan Seleksi Fitur.....	69
Tabel 4.19 Standar Deviasi Akurasi antar Kernel.....	72

ABSTRAK

Aqwam, Yuwatsiqul. 2025. **Klasifikasi Penyakit Kardiovaskular Menggunakan *Support Vector Machine* (SVM) Berbasis Fitur Kombinasi Hasil Seleksi *Recursive Feature Elimination* (RFE) dan *Chi-Square***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Irwan Budi Santoso, M.Kom (II) Nur Fitriyah Ayu Tunjung Sari, M.Cs.

Kata kunci: Penyakit Kardiovaskular, Klasifikasi, *Support Vector Mechine*, *Chi-Square*, *Recursive Feature Elimination*.

Penyakit kardiovaskular merupakan salah satu masalah kesehatan paling serius yang masih menjadi penyebab kematian tertinggi di dunia setiap tahunnya. Kondisi ini menunjukkan bahwa diagnosis dini dan klasifikasi penyakit kardiovaskular sangat penting untuk meningkatkan peluang kesembuhan serta memperpanjang harapan hidup pasien. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi penyakit kardiovaskular menggunakan metode *Support Vector Machine* (SVM) serta menganalisis pengaruh seleksi fitur *Recursive Feature Elimination* (RFE) dan *Chi-Square* terhadap performa komputasi model. Dataset yang digunakan adalah *Heart Disease Dataset* dari IEEE Dataport yang terdiri atas 1.190 data. Metode RFE dan *Chi-Square* digunakan untuk memilih fitur paling relevan, yang menghasilkan tujuh fitur utama yang digunakan dalam proses klasifikasi dengan berbagai kernel SVM. Hasil penelitian menunjukkan bahwa penerapan seleksi fitur secara signifikan meningkatkan stabilitas model, serta mempercepat proses komputasi, terutama pada kernel dengan kompleksitas tinggi. Kernel linear memberikan akurasi tertinggi sebesar 85,84%, sedangkan kernel polinomial derajat tujuh meningkat dari 77,23% menjadi 84,31%, dan kernel RBF meningkat dari 83,66% menjadi 84,86% setelah seleksi fitur. Kombinasi metode RFE dan *Chi-Square* terbukti efektif dalam meningkatkan stabilitas, dan efisiensi komputasi, sehingga berpotensi diterapkan sebagai alternatif dalam mendukung proses klasifikasi penyakit kardiovaskular.

ABSTRACT

Aqwam, Yuwatsiqul. 2025. **Classification of Cardiovascular Diseases Using *Support Vector Machine* (SVM) Based on a Combination of Features Selected by *Recursive Feature Elimination* (RFE) and *Chi-Square***. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Advisors: (I) Dr. Irwan Budi Santoso, M.Kom (II) Nur Fitriyah Ayu Tunjung Sari, M.Cs.

Cardiovascular disease is one of the most serious health problems and remains the leading cause of death worldwide each year. This condition shows that early diagnosis and classification of cardiovascular disease are very important to increase the chances of recovery and prolong the life expectancy of patients. This study aims to develop a cardiovascular disease classification system using the *Support Vector Machine* (SVM) method and analyze the effect of *Recursive Feature Elimination* (RFE) and *Chi-Square* feature selection on model computational performance. The dataset used is the *Heart Disease Dataset* from IEEE Dataport, which consists of 1,190 data points. The RFE and *Chi-Square* methods were used to select the most relevant features, resulting in seven main features used in the classification process with various SVM kernels. The results showed that the application of feature selection significantly improved model stability and accelerated the computational process, especially for kernels with high complexity. The linear kernel provided the highest accuracy of 85.84%, while the seventh-degree polynomial kernel increased from 77.23% to 84.31%, and the RBF kernel increased from 83.66% to 84.86% after feature selection. The combination of the RFE and *Chi-Square* methods proved to be effective in improving stability and computational efficiency, making it a potential alternative to support the cardiovascular disease classification process.

Keywords: Cardiovascular Disease, Classification, *Support Vector Machine*, *Chi-Square*, *Recursive Feature Elimination*.

مستخلص البحث

أقوام، يواتسيفل. 2025. تصنيف أمراض القلب والأوعية الدموية باستخدام آلة الدعم المتجه (SVM) بناءً على نتائج الجمع بين ميزات التخلص التكراري من الميزات (RFE) واختيار كاي-مربع. أطروحة. برنامج دراسة هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية، مالانج. المشرفان: (I) الدكتور إروان بودي سانتوسو، ماجستير في علوم الحاسوب (II) نور فيترياه أيو تونجونغ ساري، ماجستير في علوم الحاسوب

الكلمات المفتاحية: أمراض القلب والأوعية الدموية، التصنيف، آلة الدعم المتجه، كاي مربع، إزالة السمات التكرارية.

تعد أمراض القلب والأوعية الدموية من أخطر المشكلات الصحية، ولا تزال السبب الرئيسي للوفاة في جميع أنحاء العالم كل عام. وتبين هذه الحالة أن التشخيص المبكر وتصنيف أمراض القلب والأوعية الدموية أمران في غاية الأهمية لزيادة فرص الشفاء وإطالة العمر المتوقع للمرضى. يهدف هذا البحث إلى تطوير نظام تصنيف أمراض القلب والأوعية الدموية باستخدام طريقة آلة المتجهات الداعمة (SVM) وتحليل تأثير اختيار ميزة حذف الميزة التكراري (RFE) و *Chi-Square* على أداء حسابات النموذج. المجموعة البياناتية المستخدمة هي مجموعة بيانات أمراض القلب من بوابة بيانات IEEE Dataport والتي تتكون من 1,190 بياضة. تم استخدام طريقي RFE و *Chi-Square* لاختيار الميزات الأكثر صلة، مما أدى إلى استخدام سبع ميزات رئيسية في عملية التصنيف مع نوى SVM مختلفة. تظهر نتائج الدراسة أن تطبيق اختيار الميزات يحسن بشكل كبير من استقرار النموذج ويسرع عملية الحساب، خاصة بالنسبة للنوى ذات التعقيد العالي. قدمت النواة الخطية أعلى دقة بنسبة 85.84٪، بينما زادت النواة متعددة الحدود من الدرجة السابعة من 77.23٪ إلى 84.31٪، وزادت النواة RBF من 83.66٪ إلى 84.86٪ بعد اختيار الميزات. أثبتت فعالية الجمع بين طريقي RFE و *Chi-Square* في تحسين الاستقرار والكفاءة الحسابية، مما يجعلهما قابلتين للتطبيق كبديل في دعم عملية تصنيف أمراض القلب والأوعية الدموية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit kardiovaskular adalah gangguan jantung dan pembuluh darah yang umumnya disebabkan oleh gaya hidup tidak sehat dan faktor risiko seperti hipertensi, kolesterol tinggi, dan diabetes, serta sering terlambat terdeteksi karena berkembang secara perlahan. (Natsir, 2024). Menurut data *World Health Organization* (WHO), penyakit kardiovaskular menjadi penyebab kematian nomor satu secara global. Pada tahun 2019, tercatat sekitar 17,9 juta orang meninggal dunia akibat penyakit ini, mencakup sekitar 32% dari total kematian global. Dari angka tersebut, 85% di antaranya disebabkan oleh serangan jantung dan stroke (Rachmawati et al., 2021.).

Organisasi *World Heart Federation* (WHF) memperkirakan bahwa pada tahun 2025, penyakit jantung akan menjadi penyebab utama kematian di negara-negara Asia. Lebih lanjut, disebutkan bahwa 78% kematian akibat penyakit jantung terjadi di negara-negara berpenghasilan rendah dan menengah, yang sebagian besar belum memiliki sistem diagnosis yang cepat dan akurat. Sementara itu, di negara berkembang, angka kematian akibat penyakit jantung diperkirakan akan meningkat sebesar 120% pada wanita dan 137% pada pria antara tahun 1990 hingga 2020 (Farida & Bahri, 2024). Fakta-fakta ini menunjukkan bahwa penyakit kardiovaskular tidak hanya berbahaya, tetapi juga menjadi beban besar bagi sistem kesehatan global.

Diagnosis dini dan klasifikasi terhadap penyakit kardiovaskular menjadi tantangan utama dalam dunia medis, mengingat gejala dan indikator klinisnya sering kali tidak spesifik dan bervariasi pada setiap individu. Kondisi ini menuntut tenaga medis untuk melakukan analisis data klinis yang komprehensif agar dapat menilai tingkat risiko secara tepat. Tantangan tersebut semakin besar di tengah keterbatasan tenaga medis dan fasilitas kesehatan yang belum sepenuhnya mampu melakukan diagnosis cepat dan akurat. Dalam beberapa tahun terakhir, kemajuan di bidang kecerdasan buatan, khususnya pada teknik klasifikasi berbasis *machine learning*, telah membuka peluang baru untuk membantu tenaga medis dalam mengidentifikasi pola data dan mengklasifikasikan kondisi pasien berdasarkan hasil pemeriksaan klinis. Dengan pendekatan ini, proses analisis dan prediksi risiko penyakit kardiovaskular dapat dilakukan dengan lebih efisien dan objektif untuk mendukung pengambilan keputusan medis yang tepat. (Ortega-Martorell et al., 2022).

Upaya ini sejalan dengan nilai-nilai dalam ajaran Islam yang mendorong umatnya untuk menjaga kesehatan dan berikhtiar dalam menghadapi penyakit. Hal ini sebagaimana tercermin dalam firman Allah SWT dalam Al-Qur'an surat *Al-Isra'* ayat 82:

وَنُزِّلُ مِنَ الْقُرْآنِ مَا هُوَ شِفَاءٌ وَرَحْمَةٌ لِّلْمُؤْمِنِينَ وَلَا يَرْيَدُ الظَّالِمِينَ إِلَّا خَسَارًا

“Dan Kami turunkan dari Al-Quran (sesuatu) yang menjadi penawar dan rahmat bagi orang yang beriman, sedangkan bagi orang yang zalim (Al-Qur'an itu) hanya akan menambah kerugian.” (Q.S. al-Isra' [17]: 82)

Selain itu, pentingnya klasifikasi dan pengelompokan dalam memahami perbedaan serta mengambil keputusan yang adil juga ditegaskan dalam Al-Qur'an surat Al-Hujurat ayat 13:

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا ۚ إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ أَتْقَىٰكُمْ ۚ إِنَّ اللَّهَ عَلِيمٌ

خَبِيرٌ

“Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Maha Teliti.” (Q.S. al-Hujurat [49]: 13)

Kedua ayat ini mencerminkan nilai-nilai dalam Islam tentang kesehatan dan klasifikasi. Dalam QS. Al-Isra' ayat 82 menekankan bahwa Al-Qur'an adalah sumber penyembuh (*asy-syifā*) dan rahmat, bukan hanya untuk penyakit batin, tetapi juga dapat memberikan ketenangan dan motivasi spiritual dalam menghadapi penyakit fisik, seperti penyakit kardiovaskular. Sementara itu, QS. Al-Hujurat ayat 13 menggambarkan bahwa Allah SWT menciptakan manusia dalam kelompok-kelompok atau klasifikasi sosial agar saling mengenal dan memahami. Nilai ini sejajar dengan prinsip klasifikasi dalam dunia medis yang bertujuan mengenali dan membedakan kondisi pasien berdasarkan data klinis, guna memberikan penanganan yang tepat. Oleh karena itu, upaya diagnosis dini dan klasifikasi penyakit menggunakan pendekatan teknologi seperti pembelajaran mesin tidak hanya merupakan bentuk ikhtiar ilmiah, tetapi juga selaras dengan ajaran Islam dalam menjaga amanah tubuh dan membangun pemahaman berdasarkan perbedaan yang ada.

Dalam konteks teknologi, salah satu pendekatan yang banyak dikembangkan untuk mendukung diagnosis medis adalah penerapan *machine learning*, khususnya dalam klasifikasi penyakit berdasarkan data klinis pasien (Carolina Wibowo et al., 2024). Salah satu pendekatan yang penulis gunakan dalam penelitian ini adalah algoritma *Support Vector Machine* (SVM). SVM memiliki kemampuan dalam menangani data berdimensi tinggi dan memberikan hasil klasifikasi yang akurat. SVM bekerja dengan membentuk *hyperplane* optimal yang memisahkan kelas data dengan margin maksimum, serta dapat menangani kasus non-linear melalui penerapan kernel. Berbagai penelitian terdahulu membuktikan efektivitas metode ini, khususnya dengan algoritma SVM. Penelitian yang membandingkan performa SVM dengan *Neural Network* dalam klasifikasi penyakit jantung dan menemukan bahwa SVM memiliki performa yang lebih baik dengan tingkat akurasi yang lebih tinggi (Tino et al., 2023).

Namun demikian, performa algoritma klasifikasi sangat dipengaruhi oleh kualitas fitur yang digunakan. Kehadiran fitur yang tidak relevan atau berlebihan dapat menyebabkan model menjadi kurang optimal, meningkatkan risiko *overfitting*, dan memperlambat proses komputasi. Oleh karena itu, proses seleksi fitur menjadi tahap penting untuk menyederhanakan representasi data. Dua metode seleksi fitur yang banyak digunakan adalah *Chi-Square* dan *Recursive Feature Elimination* (RFE). *Chi-Square* mengevaluasi hubungan antara fitur dan target dengan uji statistik, sementara RFE menghapus fitur secara iteratif berdasarkan kontribusinya terhadap model. Penggunaan kedua metode ini secara terpisah telah banyak dibuktikan mampu meningkatkan akurasi model, namun penerapan

kombinasi keduanya masih jarang ditemukan dalam studi klasifikasi penyakit kardiovaskular.

Beberapa penelitian sebelumnya telah menunjukkan hasil yang menjanjikan. Penelitian oleh Nidhi Sinha menunjukkan bahwa penggunaan *Chi-Square* dan RFE mampu meningkatkan akurasi SVM dalam klasifikasi penyakit jantung pada *dataset* Z-Alizadeh Sani (Sinha et al., 2022). Penelitian oleh Fahrudin & Suroso (2024) juga membuktikan bahwa seleksi fitur yang tepat mampu meningkatkan akurasi model SVM dari 87,65% menjadi 96,56% (Fahrudin et al., 2024). Sementara itu, penelitian oleh Sarra dan rekan-rekannya menunjukkan bahwa integrasi *Chi-Square* dalam sistem *ensemble learning* mampu meningkatkan akurasi klasifikasi penyakit jantung dari 85,5% menjadi 90,8% (Sarra et al., n.d.).

Berdasarkan studi-studi tersebut, dapat disimpulkan bahwa kombinasi metode seleksi fitur *Chi-Square* dan RFE berpotensi memberikan peningkatan kinerja klasifikasi yang signifikan. Oleh karena itu, penelitian ini akan membahas tentang klasifikasi penyakit kardiovaskular menggunakan algoritma SVM berbasis fitur kombinasi hasil seleksi *Chi-Square* dan RFE.

Penelitian ini diharapkan dapat menghasilkan model klasifikasi penyakit kardiovaskular yang lebih efisien, dan stabil, serta memberikan kontribusi dalam pengembangan sistem pendukung keputusan medis berbasis *machine learning*. Sistem yang dikembangkan tidak dimaksudkan untuk menggantikan peran tenaga medis profesional, melainkan sebagai sarana penelitian yang dapat mendukung proses analisis dan pengambilan keputusan di bidang kesehatan. Selain itu, pendekatan ini juga mencerminkan integrasi antara nilai spiritual dan kemajuan

teknologi modern dalam upaya menjaga dan meningkatkan kualitas hidup manusia secara berkelanjutan.

1.2 Rumusan Masalah

Bagaimana pengaruh kombinasi metode seleksi fitur *Recursive Feature Elimination* (RFE) dan *Chi-Square* terhadap performa komputasi algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan penyakit kardiovaskular?

1.3 Batasan Masalah

1. Batasan masalah dalam penelitian ini adalah data *Heart Disease Dataset (Comprehensive)* yang diambil dari *IEEE Dataport*. *Dataset* ini berisi data medis terkait pasien dengan kemungkinan penyakit kardiovaskular, meliputi variabel seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan beberapa faktor gaya hidup yang berhubungan dengan risiko penyakit tersebut (Siddhartha, 2020).
2. Penelitian ini memanfaatkan subset fitur yang dihasilkan melalui proses seleksi fitur menggunakan metode RFE dan *Chi-Square* untuk menilai pengaruh kombinasi kedua metode tersebut terhadap performa komputasi algoritma SVM. Proses seleksi fitur dilakukan semata-mata untuk tujuan analisis komputasi dan peningkatan efisiensi model, bukan untuk menentukan atau menghapus relevansi klinis dari setiap atribut medis yang terdapat dalam dataset.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis pengaruh kombinasi metode seleksi fitur RFE dan *Chi-Square* terhadap performa komputasi algoritma SVM dalam mengklasifikasikan penyakit kardiovaskular.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat menghasilkan model klasifikasi penyakit kardiovaskular yang lebih akurat dan efisien dengan menggunakan kombinasi metode seleksi fitur yang optimal. Model yang dikembangkan berpotensi digunakan sebagai alternatif dalam pengembangan sistem pendukung keputusan berbasis *machine learning* untuk membantu tenaga medis dalam mendeteksi penyakit kardiovaskular secara lebih dini dan akurat. Selain itu, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan kajian akademik terkait penerapan teknik seleksi fitur pada algoritma SVM di bidang klasifikasi medis, serta berperan dalam mendukung upaya global dalam mengurangi angka kematian dan dampak yang ditimbulkan oleh penyakit kardiovaskular.

BAB II STUDI PUSTAKA

2.1 Penelitian Klasifikasi

Penelitian yang dilakukan oleh Tino et al., (2023) membandingkan performa algoritma *Support Vector Machines* (SVM) dengan *Neural Network* dalam klasifikasi penyakit jantung. Dengan menggunakan *dataset* dari *UCI Machine Learning Repository* yang terdiri atas 416 *instance*, penelitian ini mengevaluasi model menggunakan metrik akurasi, AUC, presisi, *recall*, dan *F1-score*. Hasilnya menunjukkan bahwa algoritma SVM memiliki akurasi lebih tinggi (83%) dibandingkan *Neural Network* (82%). Penelitian ini menyoroti keunggulan SVM dalam menangani klasifikasi dengan dimensi fitur yang tinggi, meskipun *Neural Network* tetap relevan untuk menyelesaikan masalah kompleks lainnya (Tino et al., 2023).

Penelitian yang berjudul “*Enhanced Stacked Ensemble-Based Heart Disease Prediction with Chi-Square Feature Selection Method*” dengan mengembangkan sistem prediksi penyakit jantung berbasis ensemble learning yang menggabungkan prediksi dari algoritma *Decision Tree*, SVM, dan *Multilayer Perceptron* (MLP). Metode ini dipadukan dengan seleksi fitur *Chi-Square* untuk meningkatkan efisiensi komputasi dan akurasi model. Menggunakan *dataset Cleveland* dari *UCI Machine Learning Repository*, hasil penelitian menunjukkan bahwa akurasi meningkat dari 85.5% menjadi 90.8% setelah penerapan seleksi fitur. Penelitian ini membuktikan efektivitas pendekatan ensemble dan seleksi fitur dalam menciptakan sistem prediksi yang akurat dan efisien (Sarraf et al., n.d.).

Penelitian oleh Fahrudin et al., (2024) yang berjudul “Pengembangan Model *Support Vector Machine* untuk Meningkatkan Akurasi Klasifikasi Diagnosis Penyakit Jantung” mengusulkan pengembangan model klasifikasi berbasis algoritma SVM dengan memanfaatkan teknik seleksi fitur dan *Grid Search* untuk meningkatkan performa diagnosis penyakit jantung. Dengan menggunakan dataset dari *UCI Machine Learning Repository*, model yang diusulkan berhasil meningkatkan akurasi dari 87.65% menjadi 96.56%, serta mengurangi jumlah fitur dari 14 menjadi 8. Penelitian ini menunjukkan pentingnya pemilihan fitur yang optimal dalam meningkatkan efisiensi dan akurasi klasifikasi (Fahrudin et al., 2024).

Dalam penelitian Sinha et al., (2022) yang membahas kerangka kerja berbasis *machine learning* untuk prediksi penyakit kardiovaskular dalam konteks smart healthcare. *Dataset* yang digunakan mencakup 56 atribut terkait demografi, gejala, dan hasil pemeriksaan. Empat metode seleksi fitur yaitu LASSO, *Tree-based algorithms*, *Chi-Square*, dan *Recursive Feature Elimination* (RFE) diterapkan untuk menghasilkan 16 set fitur dengan ukuran berbeda (5, 10, 15, dan 20 fitur). Dari total 112 model yang diuji, SVM dengan 15 fitur hasil seleksi menggunakan *Chi-Square* memberikan akurasi terbaik sebesar 92,31%. RFE juga menunjukkan performa yang kompetitif dengan akurasi 92,21% pada subset fitur yang sama. Penelitian ini menonjolkan keunggulan *Chi-Square* dan RFE dalam meningkatkan akurasi dan efisiensi klasifikasi dibandingkan metode seleksi fitur lainnya (Sinha et al., 2022).

Dalam penelitian yang dilakukan oleh Alifah et al., (2020) tentang penerapan metode seleksi fitur *Recursive Feature Elimination* (RFE) dan *Chi-Square* untuk mendeteksi Retinopati Diabetik, mereka menggunakan data medis pasien diabetes tipe 2 dari Rumah Sakit Cipto Mangunkusumo, Jakarta. Penelitian ini bertujuan untuk mengidentifikasi fitur paling signifikan yang berkaitan dengan Retinopati Diabetik menggunakan dua metode seleksi fitur. Data yang telah diproses kemudian diklasifikasikan menggunakan algoritma SVM. Hasil penelitian menunjukkan tingkat akurasi 97% hingga 100% untuk kedua metode seleksi fitur, dengan RFE memiliki waktu eksekusi yang lebih efisien dibandingkan dengan *Chi-Square* (Alifah et al., 2020).

Penelitian yang membandingkan algoritma *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM) dalam klasifikasi penyakit serangan jantung oleh Arif et al., (2024). *Dataset* penelitian diperoleh dari Kaggle dengan total 1025 data pasien, termasuk 14 fitur seperti kadar kolesterol, tekanan darah, dan gula darah puasa. Penelitian ini menggunakan *Principal Component Analysis* (PCA) untuk seleksi fitur, yang berhasil mengurangi dimensi data menjadi 8 fitur. Evaluasi menggunakan *confusion matrix* dan ROC menunjukkan bahwa akurasi KNN meningkat dari 83,6% menjadi 90,16%, sedangkan SVM meningkat dari 85,7% menjadi 86,88%. Model KNN dengan akurasi lebih tinggi dianggap lebih efektif untuk klasifikasi serangan jantung dibandingkan dengan SVM dalam penelitian ini (Arif et al., 2024).

Dalam Penelitian yang dilakukan oleh Carolina Wibowo et al., (2024) menganalisis tiga algoritma *machine learning*, yaitu *K-Nearest Neighbor* (KNN),

Logistic Regression, dan *Decision Tree*, dalam klasifikasi penyakit jantung menggunakan *Heart Disease Dataset* dari Kaggle yang terdiri dari 1025 baris dengan 14 atribut. Penelitian ini mencakup tahapan *preprocessing*, pembagian *dataset* menjadi data latih dan uji, serta evaluasi performa menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Hasilnya menunjukkan bahwa algoritma KNN memiliki akurasi tertinggi sebesar 94%, diikuti *Decision Tree* dengan 93%, dan *Logistic Regression* sebesar 86%. Penelitian ini menyimpulkan bahwa algoritma KNN memberikan kinerja terbaik untuk klasifikasi data klinis penyakit jantung dibandingkan dengan dua algoritma lainnya (Carolina Wibowo et al., 2024).

Penelitian oleh Farida & Bahri (2024) dengan menggunakan metode SVM untuk klasifikasi penyakit gagal jantung. *Dataset* yang digunakan adalah *Heart Failure Prediction* dari Kaggle, terdiri atas 918 data dengan 11 atribut prediktor seperti tekanan darah, kolesterol, dan denyut jantung maksimum. Pengujian menggunakan kernel linear, RBF, dan polinomial, menunjukkan akurasi tertinggi pada kernel linear sebesar 86,92% dengan cost 1, 10, dan 100. Model terbaik menghasilkan presisi 88,68%, *recall* 87,85%, *F1-score* 88,26%, dan akurasi 86,41%. Penelitian ini menyoroti efektivitas kernel linear dalam menangani data linier dengan risiko *overfitting* yang rendah. Penulis menyarankan pengembangan model lebih lanjut dengan menambahkan kernel sigmoid untuk meningkatkan generalisasi klasifikasi penyakit gagal jantung (Farida & Bahri, 2024).

Penelitian yang dilakukan oleh Natsir (2024) dengan judul “Analisis Deteksi Dini Penyakit Jantung dengan Pendekatan *Support Vector Machine* pada

Data Pasien” yang mengevaluasi deteksi dini risiko penyakit jantung menggunakan algoritma SVM pada data pasien RSUD Haji Makassar. *Dataset* terdiri dari 640 data pasien yang telah diproses, dibagi ke dalam tiga klaster risiko: rendah, tinggi, dan sangat tinggi menggunakan metode *Elbow*. Model SVM yang diterapkan menunjukkan kinerja unggul dengan akurasi sebesar 98,44%, presisi 98%, *recall* 99%, dan *F1-score* 98%. Penelitian ini menyoroti pentingnya *preprocessing* data dan seleksi fitur dalam meningkatkan akurasi prediksi. Penulis menyimpulkan bahwa SVM adalah alat yang efektif untuk mendeteksi risiko penyakit jantung secara dini, memungkinkan intervensi medis yang lebih cepat dan akurat (Natsir, 2024).

Penelitian dengan judul “*Comparative Evaluation of Feature Selection Methods for Heart Disease Classification with Support Vector Machine*” oleh Bidul et al., (2024) yang membandingkan efektivitas berbagai metode seleksi fitur untuk meningkatkan kinerja algoritma SVM dalam klasifikasi penyakit jantung menggunakan *dataset* besar. Metode seleksi fitur yang diuji meliputi *Logistic Regression-Recursive Feature Elimination* (LR-RFE), *Logistic Regression-Sequential Forward Selection* (LR-SFS), *Correlation-Based Feature Selection* (CFS), dan *Variance Threshold*. Hasil penelitian menunjukkan bahwa kombinasi SVM dengan LR-RFE, LR-SFS, dan *Variance Threshold* memberikan akurasi tertinggi sebesar 89%, sementara metode CFS menunjukkan akurasi lebih rendah. Selain akurasi, penelitian ini mengevaluasi presisi, *recall*, dan *F1-score* serta waktu komputasi, dengan *Variance Threshold* menghasilkan waktu komputasi tercepat (118,15 detik) untuk 23 fitur penting. Penelitian ini menyoroti pentingnya memilih

metode seleksi fitur yang sesuai untuk menangani tantangan data besar dalam klasifikasi penyakit jantung (Bidul et al., 2024).

Tabel 2.1 Penelitian terkait

Penulis	Metode	Dataset	Hasil
(Tino et al., 2023)	SVM dan Neural Network	<i>UCI Machine Learning Repository: Heart Disease Dataset</i>	Akurasi SVM 83%, Neural Network 82%
(Sarraf et al., 2024)	Ensemble Learning (DT, SVM, MLP) dan <i>Chi-Square</i>	<i>UCI Machine Learning Repository: Cleveland Heart Disease Dataset</i>	Akurasi meningkat dari 85.5% menjadi 90.8%
(Fahrudin et al., 2024)	SVM, Seleksi Fitur, dan Grid Search	<i>UCI Machine Learning Repository: Cleveland Heart Disease Dataset</i>	Akurasi meningkat dari 87.65% menjadi 96.56%, fitur berkurang dari 14 ke 8
(Sinha et al., 2022)	SVM, <i>Chi-Square</i> , RFE, LASSO, Tree-based feature selection	<i>UCI Machine Learning Repository: Z Alizadeh Sani dataset</i>	Akurasi SVM terbaik 92,31% dengan 15 fitur hasil seleksi <i>Chi-Square</i>
(Alifah et al., 2020)	RFE, <i>Chi-Square</i> , SVM	Dataset rekam medis pasien diabetes tipe 2 di Rumah Sakit Cipto Mangunkusumo	Akurasi 97%-100%; RFE lebih efisien dalam waktu eksekusi
(Arif et al., 2024)	KNN, SVM, PCA	<i>Kaggle: Heart Attack Analysis & Prediction</i>	KNN: 90.16% akurasi, SVM: 86.88% akurasi
(Carolina Wibowo et al., 2024)	KNN, Logistic Regression, Decision Tree	<i>Kaggle: Heart Disease Dataset</i>	akurasi KNN: 94%, Decision Tree: 93%, Logistic Regression: 86%
(Farida & Bahri, 2024)	SVM (Kernel Linear, RBF, Polinomial)	<i>Kaggle: Heart Failure Prediction</i>	Kernel Linear SVM: 86.92% akurasi, F1-score: 88.26%
(Natsir, 2024)	SVM (Cluster Elbow)	RSUD Haji Makassar, 640 data pasien	SVM: 98.44% akurasi, presisi: 98%, Recall: 99%
(Bidul et al., 2024)	SVM dengan LR-RFE, LR-SFS, Variance Threshold, CFS.	<i>Kaggle: Indicators of Heart Disease</i>	LR-RFE, LR-SFS, Variance Threshold: 89% akurasi, Variance Threshold tercepat (118.15 detik)
Usulan Penelitian			

Penulis	Metode	Dataset	Hasil
Yuwatsiqul Aqwam	SVM, <i>Chi-square</i> , RFE	<i>IEEE Dataport: Heart Disease Dataset (Comprehensive)</i>	-

2.2 Penyakit Kardiovaskular

Penyakit Kardiovaskular (PKV) adalah gangguan yang memengaruhi jantung dan pembuluh darah, seperti penyakit jantung koroner, tekanan darah tinggi (hipertensi), gagal jantung, aritmia, dan penyakit arteri perifer. Penyakit ini biasanya terjadi ketika pembuluh darah menyempit atau tersumbat, yang dapat mengganggu aliran darah ke jantung atau otak. Misalnya, penyakit jantung koroner disebabkan oleh penumpukan plak di arteri yang menyuplai darah ke jantung, sedangkan penyakit arteri perifer terjadi karena penyempitan pembuluh darah yang sering menyerang area kaki (Farida & Bahri, 2024).

PKV merupakan salah satu penyebab kematian tertinggi di dunia. Berdasarkan data dari *World Health Organization* (WHO), setiap tahunnya sekitar 17,9 juta orang meninggal akibat penyakit ini, yang mencakup hampir sepertiga dari seluruh kematian global. *World Heart Federation* (WHF) juga memprediksi bahwa pada tahun 2025, penyakit jantung akan menjadi penyebab utama kematian di wilayah Asia. Selain itu, sekitar 78% kematian akibat penyakit jantung terjadi di negara-negara berpendapatan rendah dan menengah, yang umumnya masih kekurangan sistem diagnosis yang cepat dan akurat. Di negara-negara berkembang, peningkatan kematian akibat penyakit ini juga diproyeksikan cukup signifikan, yaitu sebesar 120% pada perempuan dan 137% pada laki-laki dalam rentang waktu 1990 hingga 2020 (Rachmawati et al., n.d.).

Ada beberapa faktor yang dapat meningkatkan risiko seseorang terkena serangan jantung, di antaranya adalah pola makan tinggi lemak jenuh dan garam, kurangnya aktivitas fisik, kebiasaan merokok, serta konsumsi alkohol berlebihan. Faktor kesehatan seperti tekanan darah tinggi, diabetes, dan kolesterol tinggi juga berperan besar. Selain itu, faktor usia dan riwayat keluarga dengan penyakit jantung turut meningkatkan kemungkinan terkena serangan jantung, terutama pada pria di atas usia 45 tahun dan wanita pascamenopause (Pane & Simorangkir, 2022).

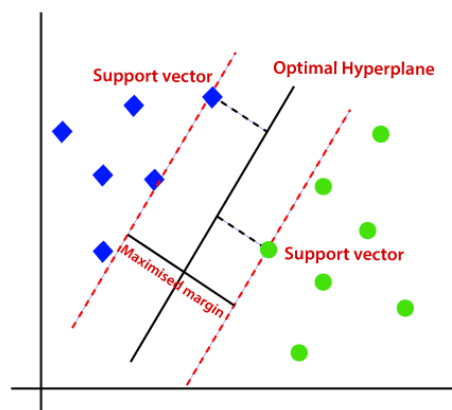
Pencegahan penyakit kardiovaskular dapat dilakukan dengan menjalani gaya hidup sehat, seperti mengonsumsi makanan bergizi seimbang, rutin berolahraga, menghindari rokok, dan mengelola stres dengan baik. Selain itu, memeriksakan kesehatan secara rutin, seperti pengecekan tekanan darah, kadar gula, dan kolesterol, sangat penting untuk mendeteksi risiko sejak dini. Edukasi kesehatan dan penyuluhan kepada masyarakat juga dapat membantu meningkatkan kesadaran akan pentingnya menjaga kesehatan jantung agar terhindar dari penyakit ini.

2.3 Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran *supervised learning* yang dikembangkan oleh Boser, Guyon dan Vapnik yang pertama kali dipresentasikan pada tahun 1992 dalam acara *Annual Workshop on Computational Learning Theory* (Amalia et al., 2024). SVM merupakan salah satu algoritma *machine learning* dengan model *supervised learning* yang berguna sebagai algoritma klasifikasi dan regresi. Algoritma SVM termasuk metode klasifikasi linear dengan memaksimal margin dan menggunakan *hyperplane* dalam melakukan

klasifikasi data secara biner. Memaksimalkan nilai margin merupakan metode yang digunakan untuk mendapatkan nilai *hyperplane* yang optimal. Margin merupakan jarak antara posisi terdekat di setiap kelas ke *hyperplane*. *Hyperplane* merupakan batas antara dua kelas yang telah ditentukan (Farida & Bahri, 2024).

Metode SVM bertujuan untuk menciptakan *hyperplane* optimal yang mampu memisahkan data menjadi dua kelas. *Hyperplane* ini adalah bidang yang berfungsi sebagai pembatas antara kelas-kelas dalam data, yang posisinya tegak lurus terhadap pola-pola data terdekat. Pola-pola terdekat tersebut dikenal sebagai *support vector* dan merupakan titik-titik yang merepresentasikan himpunan data dengan pengaruh signifikan terhadap posisi *hyperplane* (Desiani et al., 2025).



Gambar 2.1 Ilustrasi *support vector machine*

sumber : <https://pemrogramanmatlab.com/2023/08/08/penerapan-algoritma-support-vector-machine-dalam-klasifikasi-citra-menggunakan-matlab/>

Untuk menentukan *hyperplane* yang optimal, langkah pentingnya adalah menemukan margin maksimum, yaitu jarak terlebar antara *hyperplane* dan SVM dari masing-masing kelas. Dengan demikian, SVM mampu memaksimalkan pemisahan antar kelas, memastikan model memiliki performa yang baik pada data baru (Maulina Putri et al., 2023).

Untuk mengoptimalkan perhitungan dalam SVM, diperlukan algoritma optimasi untuk menghasilkan nilai parameter α dan bias yang konvergen. Salah satu algoritma yang populer dan efisien untuk optimasi adalah *Sequential Minimal Optimization* (SMO), yang bertujuan memperbarui nilai α secara bertahap melalui iterasi. SMO dirancang untuk menyelesaikan masalah optimisasi kuadratik dalam pembelajaran SVM. Dengan SMO, SVM dapat menghitung parameter-parameter model seperti bobot (w), bias (b), dan koefisien *Lagrange* (α_i) secara lebih cepat dan akurat (Sirait et al., 2019). Algoritma SMO bekerja dengan memilih dua variabel α dalam setiap iterasi, menghitung batas bawah dan atas (L dan H), serta memperbarui nilai-nilai ini hingga mencapai kondisi konvergen sesuai batas toleransi yang diinginkan. (Kocaoğlu, 2024).

Karena kesederhanaan dan fleksibilitas dari SVM yang memungkinkan pemisahan data yang kompleks, metode ini telah terbukti andal dalam berbagai aplikasi, termasuk di bidang kesehatan. Dengan fungsi kernel, SVM dapat menangani data non-linear, seperti mendeteksi perubahan struktural pada otak yang menyebabkan epilepsi. Selain itu, asumsi minimal terhadap data membuat SVM adaptif untuk berbagai jenis *dataset*. Oleh karena itu, SVM menjadi pilihan populer dalam pengembangan model klasifikasi medis (Santoso et al., 2022).

2.4 Seleksi Fitur *Chi-Square*

Chi-Square adalah metode statistik yang sering digunakan untuk memilih fitur pada data yang berbentuk kategorikal, khususnya dalam masalah klasifikasi. Metode ini mengevaluasi hubungan antara setiap fitur dengan label target menggunakan tabel untuk melihat perbandingan. Hubungan ini diukur dengan

membandingkan "frekuensi yang diamati" (berapa sering sesuatu terjadi) dengan "frekuensi yang diharapkan" (berapa sering sesuatu seharusnya terjadi) untuk melihat apakah ada hubungan yang signifikan antara fitur dan target. Karena tidak membutuhkan data dengan distribusi tertentu, metode ini fleksibel dan dapat digunakan untuk berbagai jenis data. Dalam prosesnya, *Chi-Square* memberi skor pada setiap fitur berdasarkan seberapa besar hubungan tersebut, sehingga mempermudah menentukan fitur mana yang paling relevan (Almutiri & Saeed, 2019).

Dalam *machine learning*, *Chi-Square* adalah metode yang sederhana tetapi efektif untuk memilih fitur. Metode ini menghitung skor *Chi-Square* untuk setiap fitur terhadap label target, lalu mengurutkan fitur-fitur tersebut berdasarkan skor yang didapat. Fitur dengan skor tertinggi dianggap paling berpengaruh terhadap hasil klasifikasi. Keunggulan utama metode ini adalah kemampuannya untuk membuang fitur yang tidak relevan, sehingga membuat proses komputasi lebih cepat dan meningkatkan akurasi model (Franke et al., 2012). Namun, *Chi-Square* hanya menganalisis setiap fitur secara terpisah dan tidak mempertimbangkan hubungan antar fitur. Hal ini bisa menjadi kekurangan jika data memiliki fitur yang saling bergantung atau jika hubungan antar fitur penting untuk prediksi. Meskipun begitu, metode ini tetap banyak digunakan karena sederhana, cepat, dan cocok untuk data kategorikal (McHugh, 2013).

2.5 Seleksi Fitur *Recursive Feature Elimination* (RFE)

Recursive Feature Elimination (RFE) adalah metode pemilihan fitur yang dirancang untuk menangani masalah klasifikasi dengan jumlah sampel kecil, yang

awalnya diterapkan pada klasifikasi berbasis *microarray* dalam penelitian kanker, di mana jumlah sampel pelatihan kurang dari 100 tetapi memiliki ribuan fitur. RFE bekerja secara iteratif untuk memilih subset fitur yang paling relevan dalam membangun model prediktif, dengan tujuan meningkatkan kinerja model pembelajaran mesin melalui penghapusan fitur-fitur yang kurang signifikan (Almutiri & Saeed, 2019). Proses RFE dimulai dengan melatih model menggunakan seluruh set fitur dalam *dataset*. Model yang digunakan biasanya adalah model berbasis algoritma yang mampu memberikan informasi tentang bobot atau kontribusi fitur, seperti SVM, *Logistic Regression*, atau *Random Forest*. Setelah model dilatih, fitur dievaluasi berdasarkan pentingnya kontribusi mereka terhadap kinerja model. Fitur dengan bobot atau skor pentingnya paling rendah kemudian dihapus dari *dataset*. Proses ini diulangi secara bertahap, mengurangi jumlah fitur pada setiap iterasi, hingga mencapai jumlah fitur yang diinginkan atau kriteria penghentian tertentu.

Keunggulan utama RFE terletak pada kemampuannya untuk secara efektif menyaring fitur yang tidak relevan, sehingga meningkatkan efisiensi dan akurasi model. Dengan menggunakan hanya fitur-fitur yang relevan, RFE membantu mengurangi risiko *overfitting* dan membuat model lebih mampu melakukan generalisasi pada data baru. Selain itu, metode ini fleksibel karena dapat diterapkan pada berbagai algoritma pembelajaran mesin, dan sangat berguna untuk menyederhanakan analisis data dengan mengurangi jumlah dimensi *dataset* (Alifah et al., 2020). Namun, RFE juga memiliki beberapa kekurangan. Salah satunya adalah biaya komputasi yang tinggi, terutama ketika diterapkan pada *dataset*

dengan jumlah fitur yang sangat besar, karena metode ini memerlukan pelatihan model secara berulang-ulang. Selain itu, hasil seleksi fitur RFE sangat bergantung pada algoritma model yang digunakan. Jika model awal tidak akurat, fitur yang dipilih mungkin tidak optimal. RFE juga memiliki keterbatasan dalam menangani fitur yang saling berkorelasi, di mana fitur relevan mungkin dihapus karena kontribusinya tumpang tindih dengan fitur lain (Lin et al., 2017).

BAB III

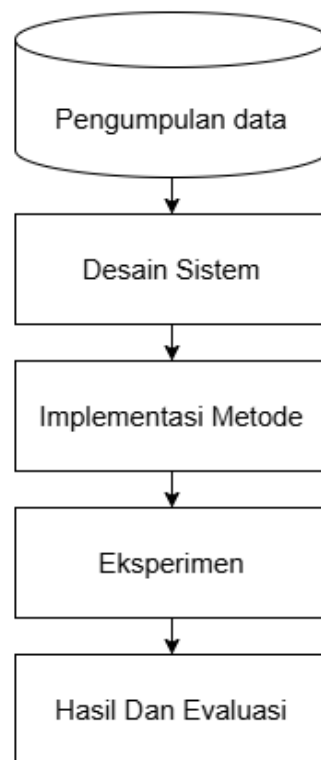
DESAIN DAN IMPLEMENTASI

Bab ini akan menjelaskan tentang perancangan sistem dan implementasi pada penelitian ini. Dalam penelitian ini memiliki beberapa tahapan, termasuk tahapan kebutuhan sistem yang akan dibuat dan solusi untuk masalah klasifikasi penyakit kardiovaskular menggunakan SVM berbasis fitur hasil seleksi RFE dan *Chi-square*.

3.1 Desain Penelitian

Desain penelitian disusun secara sistematis dengan tahapan berurutan untuk mencapai tujuan penelitian. Proses ini diawali dengan pengumpulan data, yang selanjutnya digunakan dalam tahap desain sistem. Pada tahap ini, dilakukan perancangan alur kerja, pemrosesan data, serta pemilihan metode yang akan diterapkan. Setelah itu, masuk ke tahap implementasi metode, yaitu penerapan teknik atau algoritma yang telah direncanakan.

Langkah berikutnya adalah eksperimen, di mana sistem diuji untuk mengevaluasi kinerja metode yang telah diimplementasikan dan memastikan bahwa sistem berjalan sesuai harapan. Tahap akhir adalah hasil dan evaluasi, yang berfokus pada analisis efektivitas metode yang diterapkan. Evaluasi ini bertujuan untuk menilai apakah hasil eksperimen sesuai dengan tujuan awal penelitian serta mengidentifikasi kemungkinan perbaikan atau pengembangan lebih lanjut. Diagram alur pada Gambar 3.1 menggambarkan setiap tahap penelitian, mulai dari pengumpulan data hingga evaluasi hasil.



Gambar 3.1 Prosedur penelitian

3.2 Pengumpulan data

Penelitian ini menggunakan data sekunder yang berasal dari dataset "*Heart Disease Dataset (Comprehensive)*", yang dapat diakses secara bebas melalui situs *IEEE Dataport* di "<https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive>" Data ini mencakup 1.190 data individu dengan dengan 629 individu terdiagnosis (positif) dan 561 individu tidak terdiagnosis (negatif) penyakit kardiovaskular, serta memiliki 11 atribut sebagai fitur dan 1 atribut sebagai variabel output. Data ini merupakan hasil penggabungan lima data populer yang sebelumnya tersedia secara terpisah, yaitu *Cleveland, Hungarian, Switzerland, Long Beach VA, dan Statlog (Heart) Data Set*. Dataset ini dikompilasi oleh Manu Siddhartha dan terakhir diperbarui pada 11 Juni 2020 (Alizadehsani et al., 2019). Berikut 10 contoh

sampel data yang terdiri dari individu yang terdiagnosis maupun tidak terdiagnosis penyakit kardiovaskular dapat dilihat pada Tabel 3.1.

Tabel 3.1 Contoh sampel *dataset*

age	sex	chest pain type	resting bp s	cholesterol	fasting blood sugar	resting ecg	max heart rate	exercise angina	oldpeak	ST slope	Target
40	1	2	140	289	0	0	172	0	0	1	0
49	0	3	160	180	0	0	156	0	1	2	1
37	1	2	130	283	0	1	98	0	0	1	0
48	0	4	138	214	0	0	108	1	1.5	2	1
54	1	3	150	195	0	0	122	0	0	1	0
...
45	1	1	110	264	0	0	132	0	1.2	2	1
68	1	4	144	193	1	0	141	0	3.4	2	1
57	1	4	130	131	0	0	115	1	1.2	2	1
57	0	2	130	236	0	2	174	0	0	2	1
38	1	3	138	175	0	0	173	0	0	1	0

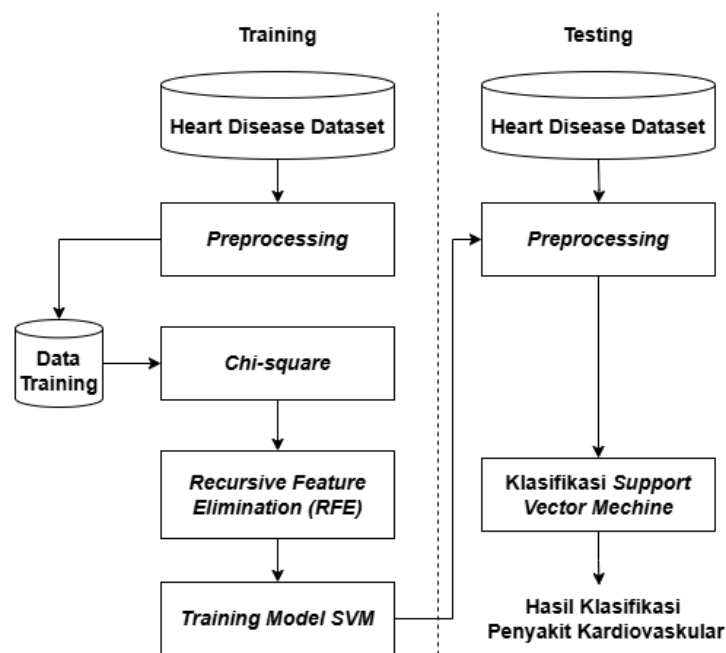
Untuk penjelasan lengkap mengenai atribut-atribut pada *dataset* ini dapat dilihat pada Tabel 3.2, yang memuat fitur *dataset* beserta deskripsi singkatnya.

Tabel 3.2 Deskripsi Fitur

No.	Nama Fitur	Deskripsi
1	Age	Usia pasien dalam tahun.
2	Sex	Jenis kelamin pasien (1: Laki-laki, 0: Perempuan).
3	Chest Pain Type	Jenis nyeri dada (1: Typical angina, 2: Atypical angina, 3: Non-anginal pain, 4: Asymptomatic).
4	Resting BP S	Tekanan darah saat istirahat dalam mmHg.
5	Cholesterol	Kadar kolesterol serum dalam mg/dl.
6	Fasting Blood Sugar	Kadar gula darah puasa > 120 mg/dl (1: Ya, 0: Tidak).
7	Resting ECG	Hasil elektrokardiogram saat istirahat (0: Normal, 1: ST-T wave abnormality, 2: Left ventricular hypertrophy).
8	Max Heart Rate	Denyut jantung maksimum yang dicapai selama latihan (dalam rentang 71-202 bpm).
9	Exercise Angina	Angina yang disebabkan oleh latihan (1: Ya, 0: Tidak).
10	Oldpeak	Depresi ST dibandingkan dengan saat istirahat.
11	ST Slope	Kemiringan segmen ST selama latihan (0: Up-sloping, 1: Flat, 2: Down-sloping).
12	Target	Klasifikasi penyakit (1: Penyakit jantung, 0: Tidak ada penyakit jantung).

3.3 Desain Sistem

Tahap desain sistem adalah proses yang dilakukan peneliti untuk menyusun alur penelitian secara terstruktur. Desain sistem memberikan gambaran menyeluruh mengenai sistem yang akan dikembangkan. Pada penelitian ini, desain sistem mencakup beberapa tahapan yang telah digambarkan pada Gambar 3.2.



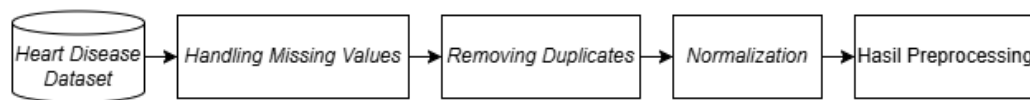
Gambar 3.2 Desain sistem

Desain sistem penelitian ini dibagi menjadi dua tahap utama, yaitu training dan testing. Pada tahap training, *Heart Disease Dataset (Comprehensive)* terlebih dahulu melalui proses *preprocessing* yang mencakup penanganan *missing value*, penghapusan data duplikat, serta normalisasi. Setelah itu, dilakukan seleksi fitur menggunakan metode *Chi-Square* dan *Support Vector Machine Recursive Feature Elimination (SVM-RFE)* untuk memilih atribut yang paling berpengaruh terhadap klasifikasi penyakit kardiovaskular. Model SVM kemudian dilatih menggunakan data yang telah diproses. Pada tahap testing, data diuji menggunakan model SVM

yang telah dilatih sebelumnya. Hasil klasifikasi kemudian dievaluasi berdasarkan metrik performa seperti *confusion matrix*, akurasi, presisi, *recall*, dan *F1-score* untuk menilai efektivitas model dalam memprediksi penyakit kardiovaskular.

3.3.1 Preprocessing

Proses *preprocessing* data bertujuan untuk memastikan bahwa data yang digunakan dalam algoritma memiliki kualitas yang baik dan siap untuk diproses lebih lanjut. Setelah data terkumpul, tahap berikutnya adalah melakukan *preprocessing* yang meliputi beberapa langkah penting. Alur *preprocessing* dapat dilihat pada Gambar 3.3 berikut:



Gambar 3.3 Alur *Preprocessing*

1. Missing Value

Pada tahap ini, data diperiksa untuk mendeteksi nilai yang hilang (*missing values*) yang dapat memengaruhi kinerja evaluasi model pada klasifikasi. Nilai yang hilang pada atribut penting diidentifikasi dan ditangani agar tidak menyebabkan ketidakseimbangan dalam model. Dalam penelitian ini, metode penanganan *missing value* dilakukan dengan menghapus baris data yang memiliki nilai kosong pada atribut krusial. Langkah ini dilakukan untuk memastikan bahwa hanya data yang lengkap dan relevan yang digunakan dalam proses pelatihan model.

2. Duplikasi Data

Selain menangani *missing values*, tahap pembersihan data juga mencakup penghapusan duplikasi data. Data duplikat dapat menyebabkan bias dalam analisis dan mengurangi keakuratan model. Oleh karena itu, dalam penelitian ini, setiap baris data yang teridentifikasi sebagai duplikat berdasarkan atribut tertentu dihapus. Dengan menghilangkan duplikasi, *dataset* menjadi lebih bersih dan setiap sampel yang digunakan dalam proses pelatihan model memiliki informasi yang unik, sehingga meningkatkan kualitas hasil prediksi. Setelah proses penghapusan duplikasi, jumlah data dalam *dataset* berkurang sesuai dengan jumlah data yang dihapus, sehingga *dataset* yang digunakan menjadi lebih bersih dan siap untuk tahap pemrosesan selanjutnya.

3. Normalisasi Data

Normalisasi data adalah proses pengubahan skala data atau rentang nilai dalam *dataset* agar semua atribut atau fitur data memiliki skala yang sebanding. Tujuan dari adanya normalisasi ini adalah untuk menghilangkan perbedaan skala antara atribut-atribut dalam *dataset* sehingga setiap atribut memiliki pengaruh yang seimbang dalam analisis pemodelan data. Peneliti akan menggunakan metode *Min-Max Scaling* untuk mengubah nilai atribut ke dalam rentang $[0, 1]$ dengan Persamaan (3.1) berikut:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

Keterangan:

X_{scaled} : Nilai baru setelah normalisasi.

X : Nilai asli yang akan dinormalisasi.

X_{min} : Nilai minimum dari keseluruhan fitur X .

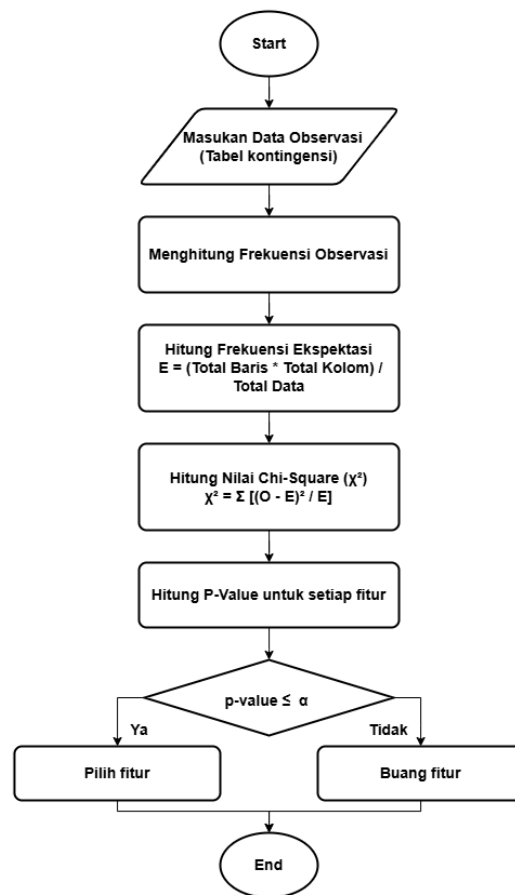
X_{max} : Nilai maksimum dari keseluruhan fitur X .

3.3.2 *Chi-Square*

Metode *Chi-Square* (χ^2) merupakan salah satu teknik statistik yang umum digunakan dalam proses seleksi fitur, khususnya untuk data yang bersifat kategorikal. Tujuan utama dari metode ini adalah untuk menguji tingkat ketergantungan antara setiap fitur independen dengan variabel target (kelas). Dalam konteks klasifikasi, *Chi-Square* digunakan untuk mengevaluasi sejauh mana suatu fitur memberikan informasi yang berguna terhadap proses prediksi kelas, yaitu dengan melihat apakah terdapat hubungan statistik yang signifikan antara nilai-nilai fitur tersebut dengan label target.

Secara prinsip, metode *Chi-Square* bekerja dengan membandingkan distribusi observasi aktual (*observed frequency*) dari setiap kategori dalam fitur terhadap distribusi harapan (*expected frequency*) yang akan terjadi apabila tidak ada keterkaitan sama sekali antara fitur dan kelas target. Jika hasil perbandingan menunjukkan bahwa distribusi aktual sangat berbeda secara signifikan dari distribusi yang diharapkan, maka dapat disimpulkan bahwa fitur tersebut memiliki kontribusi yang bermakna dalam proses klasifikasi. Sebaliknya, jika tidak terdapat perbedaan yang berarti, maka fitur tersebut cenderung tidak relevan dan dapat dipertimbangkan untuk dieliminasi dari pemodelan.

Gambar 3.4 menunjukkan alur seleksi fitur menggunakan metode *Chi-Square* dalam penelitian ini.

Gambar 3.4 Flowchart Seleksi Fitur *Chi-square*

Proses seleksi fitur pada Gambar 3.4 Flowchart Seleksi Fitur *Chi-square* dengan metode *Chi-Square* terdiri dari beberapa tahapan sistematis. Tahap pertama adalah membentuk tabel kontingensi, yaitu tabel dua dimensi yang menunjukkan frekuensi gabungan antara kategori dari fitur dan kelas target. Tabel ini memungkinkan analisis lebih lanjut terhadap penyebaran data dan identifikasi pola keterkaitan antar variabel. Setelah itu, dilakukan perhitungan terhadap frekuensi observasi (O), yaitu jumlah kemunculan masing-masing kategori fitur pada setiap kelas target. Kemudian, dilakukan perhitungan frekuensi ekspektasi (E) untuk setiap pasangan nilai fitur dan target dengan menggunakan rumus sebagai berikut:

$$E = \frac{(Total\ baris \times Total\ kolom)}{Total\ seluruh\ data} \quad (3.2)$$

Frekuensi ekspektasi ini menunjukkan jumlah kemunculan yang diharapkan jika fitur tidak memiliki hubungan dengan variabel target. Jika nilai aktual sangat berbeda dengan nilai ekspektasi, maka kemungkinan besar fitur tersebut memiliki hubungan dengan variabel target. Selanjutnya, dilakukan perhitungan nilai *Chi-Square* untuk setiap kategori fitur menggunakan rumus:

$$X^2 = \sum \frac{(O-E)^2}{E} \quad (3.3)$$

Keterangan:

O = frekuensi yang diamati pada kategori tertentu

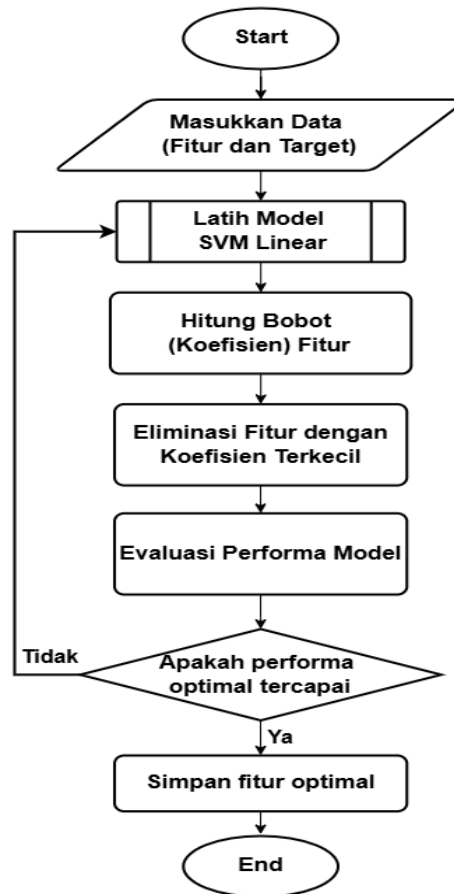
E = frekuensi yang diharapkan jika tidak ada hubungan antara variable

Dimana *O* adalah frekuensi aktual yang diamati, dan *E* adalah frekuensi ekspektasi. Setelah semua nilai *Chi-Square* dihitung, dilakukan perhitungan *p-value* untuk setiap fitur. *P-value* ini digunakan untuk mengukur tingkat signifikansi hubungan antara fitur dan variabel target. Tahap berikutnya adalah membandingkan *p-value* dengan nilai ambang (α) yang telah ditentukan. Kriteria pemilihan fitur adalah sebagai berikut jika $p \leq \alpha$, fitur dianggap memiliki hubungan yang signifikan dengan kelas maka dipilih; jika $p > \alpha$, fitur dianggap tidak signifikan maka dibuang. Dengan pendekatan ini, fitur yang tidak berkontribusi secara signifikan terhadap prediksi dapat dieliminasi, sehingga meningkatkan efisiensi dan akurasi model klasifikasi.

3.3.3 Recursive Feature Elimination

Metode *Recursive Feature Elimination* (RFE) bekerja dengan melatih model klasifikasi pada seluruh fitur, mengevaluasi kontribusi masing-masing fitur, lalu menghapus fitur dengan bobot paling kecil dalam proses iteratif hingga jumlah fitur yang optimal tercapai. Dalam penelitian ini, RFE digunakan untuk mengurangi

jumlah fitur yang digunakan dalam model. Gambar 3.5 menunjukkan alur seleksi fitur menggunakan metode RFE dalam penelitian ini.



Gambar 3.5 Flowchart Seleksi Fitur Recursive Feature Elimination (sumber: Atif & Salmi, 2022)

Dari Gambar 3.5 Flowchart Seleksi Fitur Recursive Feature Elimination (sumber: Atif & Salmi, 2022) Langkah pertama dalam penerapan RFE adalah menyiapkan *dataset* yang terdiri dari fitur-fitur terbaik yang dipisahkan menjadi fitur independen (X) dan variabel target (y). Selanjutnya, SVM dengan kernel linear digunakan sebagai estimator dalam proses seleksi fitur. SVM dipilih karena kemampuannya untuk menghasilkan *hyperplane* yang optimal dalam memisahkan kelas, yang membantu dalam menentukan fitur-fitur mana yang paling berpengaruh dalam klasifikasi.

Pada tahap pelatihan model, algoritma SVM dengan kernel linear digunakan sebagai estimator dalam metode RFE. SVM linear sangat sesuai digunakan dalam proses seleksi fitur karena menghasilkan bobot (*koefisien*) untuk setiap fitur setelah proses pelatihan. Bobot tersebut menunjukkan seberapa besar kontribusi masing-masing fitur terhadap pemisahan kelas dalam ruang vektor.

Berikut adalah langkah-langkah proses pelatihan RFE menggunakan estimator SVM kernel linear yang terdiri dari tiga tahap (Sundari et al., 2024):

1. Tahap pertama adalah melatih *dataset* untuk menghitung bobot setiap fitur.

Bobot fitur ini diperoleh dari nilai α yang dihasilkan oleh model SVM.

Persamaan untuk menghitung bobot fitur didefinisikan oleh:

$$w = \sum_1^i a_i y_i x_i \quad (3.4)$$

Dimana:

w : vektor bobot akhir setelah semua alpha di perbarui

a_i : pengali *lagrange* yang dioptimalkan untuk data latih ke i

x_i : data pelatihan dari fitur ke- k

y_i : label kelas dari fitur ke- k

2. Kemudian tahap kedua yaitu menghitung kriteria peringkat untuk mengurutkan fitur berdasarkan bobotnya. Fungsi peringkat kriteria didefinisikan dengan persamaan:

$$C_i = w_i^2, \quad i = 1, 2, \dots, |S| \quad (3.5)$$

Dimana:

c : nilai peringkat kriteri

w : bobot yang dikuadratkan untuk setiap fitur ke- i

$|S|$: jumlah seluruh fitur

3. Tahap terakhir yaitu mengurutkan fitur berdasarkan skor kontribusinya, dan fitur dengan skor terkecil dieliminasi terlebih dahulu. *Dataset* kemudian dilatih ulang menggunakan fitur yang tersisa. Proses ini diulang secara rekursif hingga jumlah fitur optimal tercapai atau hingga kriteria tertentu terpenuhi.

Proses ini disebut eliminasi rekursif karena dilakukan secara bertahap dan berulang (*recursive*), memastikan bahwa fitur yang tersisa adalah yang paling signifikan bagi model. Setelah jumlah fitur yang optimal tercapai, proses seleksi fitur dengan RFE selesai. Fitur-fitur ini akan disilangkan (*interception*) dengan fitur-fitur hasil seleksi menggunakan metode *Chi-Square*, kemudian hasil kombinasi tersebut akan digunakan pada tahap implementasi model klasifikasi menggunakan SVM.

3.4 Support Vector Machine

SVM adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan dua kelas dalam suatu *dataset* dengan margin maksimum. Konsep utama dalam SVM adalah menemukan garis (untuk data dua dimensi) atau bidang (untuk data tiga dimensi atau lebih) yang dapat memisahkan data secara optimal sehingga memberikan performa klasifikasi yang baik pada data baru yang belum pernah dilihat sebelumnya.

Dalam kasus di mana data dapat dipisahkan secara linear (*linear separable*), SVM akan mencari *hyperplane* dengan margin terbesar di antara dua kelas. Namun, jika data tidak dapat dipisahkan secara linear (*non-linear separable*), SVM menggunakan kernel trick untuk memetakan data ke dimensi yang lebih tinggi sehingga data tersebut dapat dipisahkan dengan *hyperplane* di ruang baru tersebut. SVM banyak digunakan dalam berbagai bidang seperti pengenalan wajah, deteksi spam, klasifikasi teks, dan diagnosis penyakit. Dalam penelitian ini, SVM diterapkan untuk mengklasifikasikan penyakit kardiovaskular berdasarkan fitur-

fitur yang telah dipilih menggunakan metode *Chi-Square* dan *Recursive Feature Elimination (RFE)*.

SVM beroperasi dengan mencari *hyperplane* optimal yang memisahkan dua kelas dengan *margin* maksimum. Secara matematis, *hyperplane* dinyatakan dalam bentuk persamaan

$$f(x) = w^T x + b = 0 \quad (3.6)$$

Keterangan:

$f(x)$ = fungsi keputusan yang digunakan untuk klasifikasi.

w = vektor bobot.

x = vektor fitur dari sampel data.

b = bias.

Hyperplane ini digunakan sebagai batas klasifikasi, di mana data dengan $f(x) > 0$ diklasifikasikan sebagai kelas positif dan data dengan $f(x) < 0$ sebagai kelas negatif. Untuk menemukan *hyperplane* optimal, SVM memaksimalkan *margin*, yaitu jarak antara *hyperplane* dan *support vectors*. Jarak antara kedua *hyperplane* ini adalah $2/\|w\|$, di mana $\|w\|$ adalah norma dari vektor bobot w . Tujuan kita adalah untuk memaksimalkan *margin*, yang sama dengan meminimalkan $\|w\|$, dengan memecahkan masalah optimasi yang dapat dituliskan sebagai:

$$\min_{w, b} \frac{1}{2} \|w\|^2 \quad (3.7)$$

dengan kendala:

$$y_i(w^T x_i + b) \geq 1, \quad i = 1, \dots, N$$

Dalam beberapa kasus, data tidak sepenuhnya dapat dipisahkan secara linear oleh *hyperplane*, sehingga SVM memperkenalkan slack variable ξ_i untuk mengizinkan beberapa data melanggar batas *margin*.

$$\min_{w, b, \xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (3.8)$$

dengan kendala:

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \forall i \quad (3.9)$$

Dengan adanya slack variable, SVM memberikan toleransi terhadap kesalahan klasifikasi dengan tetap mempertahankan margin maksimum. Untuk mengontrol trade-off antara margin yang lebih besar dan jumlah pelanggaran batas margin, digunakan parameter C. Nilai C yang besar cenderung menghasilkan model dengan kesalahan klasifikasi yang lebih kecil tetapi berisiko overfitting, sedangkan nilai C yang kecil memungkinkan margin lebih besar dengan toleransi kesalahan yang lebih tinggi (Santoso et al., 2022).

Dalam penelitian ini, untuk menyelesaikan masalah optimasi dalam SVM digunakan algoritma Sequential Minimal Optimization (SMO). SMO merupakan metode yang secara efisien menyelesaikan *Dual Optimization Problem* dari SVM dengan memecah masalah optimasi yang sangat besar menjadi sub-masalah yang terkecil yang mungkin. Sub-masalah ini hanya melibatkan dua Lagrange multipliers (α_i, α_j) pada setiap iterasi karena batasan linier dari dual problem mengharuskan minimal dua *multipliers* diperbarui secara bersamaan. Keunggulan utama SMO terletak pada fakta bahwa sub-masalah ini dapat diselesaikan secara analitis, yang menghindari penggunaan *numerical quadratic programming (QP) optimizer* yang memakan waktu dan rumit (Platt, n.d.).

Dalam pendekatan dual Lagrange, masalah optimasi SVM dapat diformulasikan sebagai berikut:

$$\min \Psi(\alpha) = \min \left(\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j K(x_i, x_j) \alpha_i \alpha_j - \sum_{i=1}^N \alpha_i \right) \quad (3.10)$$

Keterangan:

α_i = *Lagrange multiplier*.

y_i, y_j = label kelas dari sampel data.

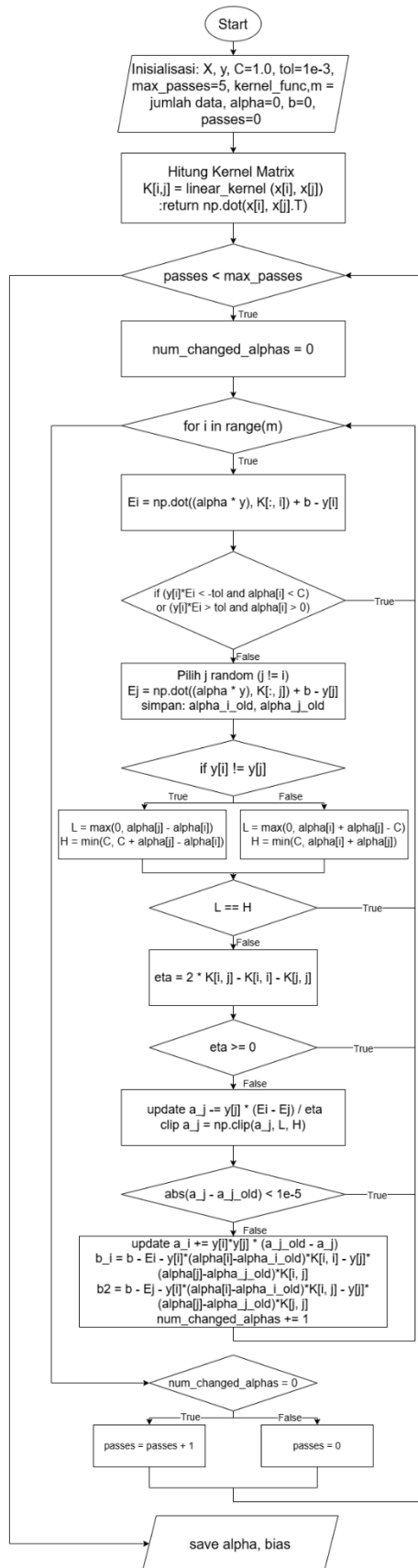
$K(x_i, x_j)$ = fungsi kernel yang mengukur kesamaan antara dua titik data x_i dan x_j .

N = jumlah data latih

Dengan batasan:

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

Algoritma SMO bertujuan untuk menyelesaikan optimasi SVM dengan cara membagi masalah utama menjadi sub-masalah yang lebih kecil, yaitu dengan memilih dua variabel *Lagrange multiplier* (α_i dan α_j) pada setiap iterasi. Adapun *flowchart* yang menggambarkan proses metode SVM dengan algoritma SMO dapat dilihat pada Gambar 3.6 *Flowchart Sequential Minimal Optimization*



Gambar 3.6 Flowchart Sequential Minimal Optimization

Pada Gambar 3.6 *Flowchart Sequential Minimal Optimization*, langkah pertama adalah inisialisasi parameter, yaitu semua *multiplier Lagrange* (α) diatur ke nol, begitu pula dengan parameter bias (b). Pada tahap ini, ditetapkan pula hyperparameter seperti C (parameter regulasi), toleransi (tol), jumlah iterasi maksimum (max_iter), serta parameter kernel seperti koefisien ($coef$) dan γ sesuai kebutuhan model.

Setelah inisialisasi, algoritma memasuki tahap pemilihan pasangan variabel. SMO menggunakan heuristik untuk memilih sepasang *multiplier Lagrange* (α_i dan α_j) yang akan dioptimalkan secara bersama. Pada pemilihan pertama (*outer loop*), algoritma menelusuri seluruh dataset untuk menemukan contoh yang melanggar kondisi Karush-Kuhn-Tucker (KKT). Jika ditemukan, contoh tersebut dipilih sebagai kandidat pertama α_i . Setelah satu kali iterasi penuh, algoritma berfokus pada subset contoh yang tidak berada pada batas (*non-bound*), yaitu dengan nilai $0 < \alpha_i < C$, karena contoh semacam ini lebih berpotensi melanggar kondisi KKT dan lebih relevan untuk mempercepat konvergensi. Setelah α_i dipilih, dilakukan pemilihan kedua (*inner loop*) untuk menentukan α_j . Tujuan pemilihan ini adalah memaksimalkan ukuran langkah yang dapat diambil selama optimisasi bersama. Besarnya langkah diaproksimasi oleh nilai absolut selisih error prediksi $|E_1 - E_2|$. Untuk mempercepat pemilihan, SMO menyimpan nilai error ($E_i = u_i - y_i$) dalam sebuah cache, kemudian memilih contoh kedua dengan error E_2 yang paling berbeda dari E_1 .

Setelah memilih pasangan variabel α_i dan α_j , langkah berikutnya dalam algoritma SMO adalah menghitung batas optimasi untuk memastikan bahwa nilai

α_j tetap berada dalam rentang yang diperbolehkan. Rentang ini ditentukan oleh dua nilai batas, yaitu L (batas bawah) dan H (batas atas). Nilai L dan H ini ditentukan berdasarkan apakah label kelas y_1 dan y_2 sama atau berbeda, yang dihitung menggunakan persamaan:

- Jika label berbeda ($y_1 \neq y_2$):

$$L = \max(0, \alpha_j - \alpha_i), H = \min(C, C + \alpha_j - \alpha_i) \quad (3.11)$$

- Jika label sama ($y_1 = y_2$):

$$L = \max(0, \alpha_j + \alpha_i - C), H = \min(C, \alpha_j + \alpha_i) \quad (3.12)$$

langkah berikutnya adalah melakukan optimisasi analitis terhadap pasangan variabel α_i dan α_j . Pertama-tama, dihitung nilai η (*Learning Rate*) yang merepresentasikan kelengkungan fungsi objektif di sepanjang garis kendala kesetaraan linear:

$$\eta = K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j) \quad (3.11)$$

Apabila nilai $\eta > 0$, maka nilai baru α_j dapat diperoleh dari formula berikut:

$$\alpha_j^{baru} = \alpha_j + \frac{y_j(E_i - E_j)}{\eta} \quad (3.12)$$

Setelah α_j diperbarui, nilai α_i dihitung kembali dengan memanfaatkan hubungan linear antara keduanya:

$$\alpha_i^{baru} = \alpha_i + y_i y_j (\alpha_j^{lama} - \alpha_j^{baru}) \quad (3.13)$$

Setelah nilai α_i dan α_j diperbarui, maka bias (b) juga perlu dihitung ulang untuk memastikan fungsi keputusan tetap valid. Terdapat dua formula pembaruan bias berdasarkan titik i dan j :

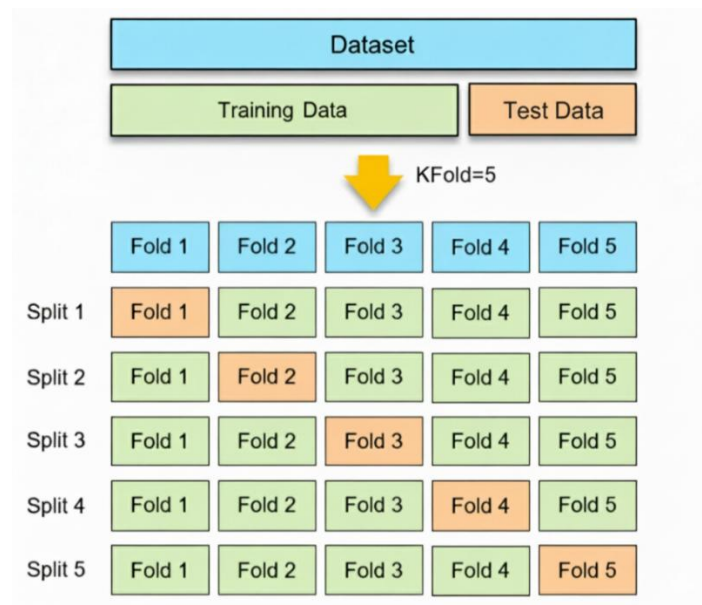
$$b_1 = b - E_i - y_i(\alpha_i^{\text{baru}} - \alpha_i^{\text{lama}})K(x_i, x_i) - y_j(\alpha_j^{\text{baru}} - \alpha_j^{\text{lama}})K(x_i, x_j) \quad (3.14)$$

$$b_2 = b - E_j - y_i(\alpha_i^{\text{baru}} - \alpha_i^{\text{lama}})K(x_i, x_j) - y_j(\alpha_j^{\text{baru}} - \alpha_j^{\text{lama}})K(x_j, x_j) \quad (3.15)$$

Proses ini berlanjut secara iteratif, bergantian antara seluruh dataset dan subset non-bound, hingga semua contoh pelatihan memenuhi kondisi KKT dalam batas toleransi atau hingga tidak ada lagi pasangan variabel yang dapat dioptimalkan. Pada titik tersebut, proses optimisasi dihentikan, dan parameter model (α, b) yang optimal disimpan sebagai hasil akhir.

3.5 Skenario Uji Coba

Skenario pengujian dalam penelitian ini dirancang untuk mengevaluasi performa algoritma SVM dalam mengklasifikasikan penyakit kardiovaskular. Evaluasi dilakukan berdasarkan dua faktor utama, yaitu penggunaan kernel dalam algoritma SVM dan penerapan seleksi fitur. Setiap kombinasi pengujian dilakukan menggunakan pendekatan *5-Fold cross validation*, sehingga seluruh data digunakan secara bergantian sebagai data pelatihan dan data pengujian untuk menghasilkan evaluasi performa model yang stabil dan tidak bias. Berikut ilustrasi dari proses *5-Fold cross validation* pada Gambar 3.7 *5-fold cross-validation visualization*



Gambar 3.7 5-fold cross-validation visualization

Pada Gambar 3.7 5-fold cross-validation visualization ditunjukkan ilustrasi proses 5-Fold Cross Validation yang digunakan dalam penelitian ini. Dataset awal dibagi menjadi lima bagian yang kurang lebih seimbang (Fold 1 hingga Fold 5). Pada setiap iterasi atau *split*, satu fold berperan sebagai data uji (*test set*), sementara empat fold lainnya digunakan sebagai data latih (*training set*). Proses ini diulang sebanyak lima kali sehingga setiap fold memiliki kesempatan untuk menjadi data uji sekali dan menjadi data latih sebanyak empat kali. Sebagai contoh, pada Split 1, Fold 1 digunakan sebagai data uji sementara Fold 2 sampai Fold 5 digunakan sebagai data latih, proses ini terus berlanjut hingga Split 5, di mana Fold 5 berperan sebagai data uji dan Fold 1 sampai Fold 4 digunakan untuk pelatihan. Dengan melakukan rata-rata hasil dari kelima percobaan, diperoleh estimasi performa model yang lebih stabil dan mengurangi risiko bias akibat pemilihan subset data tertentu.

Pada aspek kernel, model diuji dengan lima jenis kernel, yaitu kernel linear, *radial basis function* (RBF), dan polinomial dengan tiga variasi derajat (*degree*) yaitu 3, 5, dan 7 (Abdurrohman, 2024). Sementara itu, pada aspek seleksi fitur, penelitian ini membandingkan dua kondisi tanpa seleksi fitur dan dengan seleksi fitur, di mana proses seleksi dilakukan menggunakan kombinasi metode *Chi-Square* dan RFE untuk memperoleh fitur-fitur yang paling relevan terhadap proses klasifikasi.

Dengan menggabungkan lima jenis kernel dan dua kondisi seleksi fitur, terdapat total 10 skenario pengujian. Masing-masing skenario diuji menggunakan metrik evaluasi berupa akurasi, presisi, *recall*, dan *F1-score*, yang kemudian dirata-ratakan dari hasil *5-Fold Cross Validation* untuk memperoleh nilai performa akhir. Parameter-parameter yang digunakan pada model SVM ditetapkan secara konstan untuk menjaga keadilan antar skenario, yaitu $C = 1$, $\gamma = 0.1$, dan $\text{tolerance} = 0.01$. Untuk kernel polinomial, digunakan derajat 3, 5, dan 7 guna mengamati pengaruh kompleksitas model terhadap performa klasifikasi.

3.6 Evaluasi

Evaluasi model merupakan tahap penting dalam mengukur kinerja algoritma klasifikasi yang digunakan dalam penelitian ini. Salah satu metode yang digunakan untuk menilai performa model adalah *Confusion Matrix*, yang berperan sebagai alat untuk menganalisis keefektifan suatu pengklasifikasi dalam mengenali daftar kelas yang berbeda. Nilai *True-Positive* dan *True-Negative* menunjukkan bahwa pengklasifikasi melakukan klasifikasi dengan benar, sedangkan nilai *False-Positive* dan *False-Negative* menandakan bahwa pengklasifikasi melakukan

klasifikasi yang salah. Pada penelitian ini, model diklasifikasikan menggunakan *Confusion Matrix* untuk mengidentifikasi empat kategori utama hasil prediksi, yang ditunjukkan pada Tabel 3.3 *Confusion Matrix* berikut:

Tabel 3.3 *Confusion Matrix*

		Hasil Prediksi oleh Algoritma	
		Positif	Negatif
Kondisi Pasien Sebenarnya	Positif	<i>True Positive</i>	<i>False Negative</i>
	Negatif	<i>False Positive</i>	<i>True Negative</i>

Dalam kasus ini, *True Positive* (TP) mengacu pada data pasien yang benar-benar menderita penyakit kardiovaskular dan berhasil terdeteksi dengan benar oleh algoritma deteksi. *True Negative* (TN) mengacu pada data pasien yang tidak menderita penyakit kardiovaskular dan diidentifikasi dengan benar sebagai negatif oleh algoritma. Sementara itu, *False Positive* (FP) terjadi ketika pasien yang sebenarnya tidak menderita penyakit kardiovaskular salah diklasifikasikan sebagai positif oleh algoritma. Sebaliknya, *False Negative* (FN) terjadi ketika pasien yang seharusnya terdeteksi memiliki penyakit kardiovaskular justru salah diklasifikasikan sebagai negatif. Dalam konteks ini, TP dan TN mencerminkan hasil klasifikasi yang benar, sedangkan FP dan FN menunjukkan kesalahan identifikasi yang dapat mempengaruhi efektivitas model deteksi penyakit kardiovaskular.

Berdasarkan *Confusion Matrix*, kinerja model dievaluasi menggunakan beberapa metrik utama, yaitu Akurasi, Presisi, *Recall*, dan *F1-Score*. Akurasi mencerminkan tingkat ketepatan model dalam mengklasifikasikan data secara

keseluruhan. Presisi menunjukkan proporsi prediksi positif yang benar. *Recall* menilai kemampuan model dalam mendeteksi seluruh kasus positif. *F1-Score* merupakan rata-rata harmonis antara presisi dan *recall* untuk menyeimbangkan keduanya, terutama pada *dataset* yang tidak seimbang.

1. Akurasi adalah metrik yang mengukur seberapa sering model membuat prediksi yang benar secara keseluruhan.

$$Akurasi(\%) = \frac{TP+TN}{TP + TN + FP + FN} \times 100\% \quad (3.15)$$

2. Presisi mengukur seberapa tepat model dalam memprediksi kelas positif.

$$Presisi(\%) = \frac{TP}{TP+FP} \times 100\% \quad (3.16)$$

3. *Recall* mengukur kemampuan model untuk mengidentifikasi semua kasus positif yang sebenarnya.

$$Recall(\%) = \frac{TP}{TP+FN} \times 100\% \quad (3.17)$$

4. *F1-Score* adalah metrik gabungan antara Presisi dan *Recall*. *F1-Score* adalah rata-rata harmonik dari kedua metrik tersebut.

$$F1\ Score(\%) = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (3.18)$$

BAB IV

HASIL DAN PEMBAHASAN

Dalam bab ini akan menjelaskan tentang uji coba dan pembahasan Klasifikasi Penyakit Kardiovaskular Menggunakan SVM Berbasis Fitur Kombinasi Hasil Seleksi RFE Dan *Chi-Square*.

4.1 Hasil *Preprocessing* Data

Tahap preprocessing data dilakukan untuk memastikan bahwa dataset yang digunakan dalam penelitian ini memiliki kualitas yang baik sebelum dilakukan pemodelan. Proses preprocessing mencakup beberapa langkah, mulai dari pembersihan data (data cleaning), normalisasi, hingga seleksi fitur. Tahapan ini sangat penting karena kualitas data secara langsung memengaruhi kinerja model klasifikasi yang dibangun. Oleh karena itu, hasil dari setiap tahap preprocessing akan dijelaskan secara rinci pada subbab berikut.

4.1.1 Hasil *Data Cleaning*

Proses *data cleaning* merupakan tahapan penting untuk memastikan bahwa data yang digunakan bersih, relevan, dan layak untuk dianalisis lebih lanjut. Pada penelitian ini, pembersihan data dilakukan terhadap data “*Heart Disease Dataset (Comprehensive)*” dengan ukuran awal sebanyak 1190 baris dan 12 kolom. Salah satu masalah utama yang ditemukan adalah adanya nilai nol (0) pada fitur ‘*cholesterol.*’ Setelah dilakukan pemeriksaan, ditemukan sebanyak 172 nilai nol pada kolom ini. Dalam konteks medis, kadar kolesterol bernilai nol tidak mungkin terjadi, sehingga nilai-nilai tersebut diperlakukan sebagai data yang hilang (*missing values*). Untuk menangani hal ini, dilakukan metode imputasi dengan mengganti

nilai nol menggunakan nilai median dari kolom tersebut. Dengan langkah ini, kolom *cholesterol* tidak lagi memiliki nilai yang hilang dan dapat digunakan pada tahap analisis berikutnya.

Selanjutnya, dilakukan pemeriksaan terhadap duplikasi data. Dari hasil pemeriksaan ditemukan sebanyak 272 data duplikat. Setelah dilakukan penghapusan, ukuran dataset yang semula berjumlah 1190 baris berkurang menjadi 918 baris dengan tetap mempertahankan 12 kolom fitur. Dengan demikian, dataset yang digunakan pada tahap selanjutnya terdiri dari data unik dan lengkap, sehingga lebih representatif serta siap diproses pada tahapan berikutnya.

4.1.2 Hasil Normalisasi Data

Beberapa variabel numerik dalam *dataset* memiliki rentang nilai yang berbeda-beda. Oleh karena itu, dilakukan proses normalisasi untuk menyamakan rentang antar variabel sehingga seluruh fitur dapat memberikan kontribusi yang seimbang pada proses pelatihan model. Peneliti menggunakan metode *Min-Max Normalization*, yang mengubah nilai pada masing-masing fitur ke dalam rentang antara 0 hingga 1.

Tabel 4.1 Sampel Data sebelum normalisasi

age	sex	chest pain type	resting bp s	cholesterol	fasting blood sugar	resting ecg	max heart rate	...	target
40	1	2	140	289	0	0	172		0
49	0	3	160	180	0	0	156		1
37	1	2	130	283	0	1	98		0
48	0	4	138	214	0	0	108		1
54	1	3	150	195	0	0	122		0

Tabel 4.1 Sampel Data sebelum normalisasi, nilai pada masing-masing fitur masih dalam skala aslinya. Sebagai contoh, fitur *age* memiliki nilai sebesar 40 hingga 54, *resting bp s* berkisar antara 130 hingga 160, dan *cholesterol* bernilai antara 180 hingga 289. Nilai-nilai ini menunjukkan adanya perbedaan skala yang signifikan antar fitur numerik.

Tabel 4.2 Sampel Data Sesudah Normalisasi

age	sex	chest pain type	resting bp s	cholesterol	fasting blood sugar	restin g ecg	max heart rate	...	target
0.245	1.0	0.333	0.70	0.394	0.0	0.0	0.789		0
0.429	0.0	0.667	0.80	0.183	0.0	0.0	0.676		1
0.184	1.0	0.333	0.65	0.382	0.0	0.5	0.268		0
0.408	0.0	1.000	0.69	0.249	0.0	0.0	0.338		1
0.531	1.0	0.667	0.75	0.212	0.0	0.0	0.437		0

Tabel 4.2 Sampel Data Sesudah Normalisasi menunjukkan data setelah proses normalisasi dilakukan, seluruh nilai fitur numerik telah disesuaikan ke dalam skala 0 hingga 1. Nilai yang sebelumnya besar seperti *cholesterol* dan *age* kini telah berada pada rentang yang sama dengan fitur lainnya, sehingga tidak lagi mendominasi dalam perhitungan model. Dengan normalisasi ini, setiap fitur numerik memiliki skala yang seragam, sehingga lebih siap digunakan dalam proses pemodelan. Skala yang setara memungkinkan algoritma pembelajaran mesin melakukan evaluasi dan pembelajaran secara optimal terhadap seluruh fitur tanpa terpengaruh oleh besarnya nilai numerik suatu fitur.

4.2 Hasil *Chi-square*

Uji *Chi-Square* digunakan untuk mengevaluasi hubungan antara masing-masing fitur independen dengan variabel target. Dalam penelitian ini, metode *Chi-Square* digunakan untuk seleksi fitur dengan mempertimbangkan signifikansi statistik berdasarkan nilai *p-value*. Fitur dengan *p-value* di bawah 0.05 dianggap memiliki hubungan yang signifikan dengan target dan layak dipertahankan dalam model.

Tabel 4.3 Hasil uji *Chi-square*

No.	Feature	<i>Chi-Square</i>	<i>p-value</i>	Signifikan (< 0.05)
1	exercise angina	133.6401	6.55×10^{-31}	Ya
2	Fasting blood sugar	50.2969	1.32×10^{-12}	Ya
3	chest pain type	26.1441	3.17×10^{-7}	Ya
4	ST slope	21.2479	4.04×10^{-6}	Ya
5	sex	18.0062	2.20×10^{-5}	Ya
6	max heart rate	8.7383	3.12×10^{-3}	Ya
7	oldpeak	5.5465	1.85×10^{-2}	Ya
8	age	5.1919	2.27×10^{-2}	Ya
9	resting ecg	1.8370	0.1753	Tidak
10	cholesterol	0.2196	0.6393	Tidak
11	resting bp s	0.1374	0.7109	Tidak

Tabel 4.3 menampilkan hasil uji *Chi-square* pada 11 fitur input. Delapan fitur menunjukkan signifikansi statistik (*p-value* < 0.05), yang menunjukkan korelasi tinggi dengan variabel target. Tiga fitur lainnya, yaitu *resting ecg*, *cholesterol*, dan *resting bp s*, memiliki nilai *p-value* lebih besar dari 0.05, sehingga dianggap tidak signifikan dan akan dihapus dari langkah pemodelan selanjutnya. Karakteristik

yang tidak penting secara statistik akan dihapus untuk meningkatkan efisiensi dan kinerja model pada tahap pemilihan fitur berikutnya.

4.3 Hasil *Recursive Feature Elimination* (RFE)

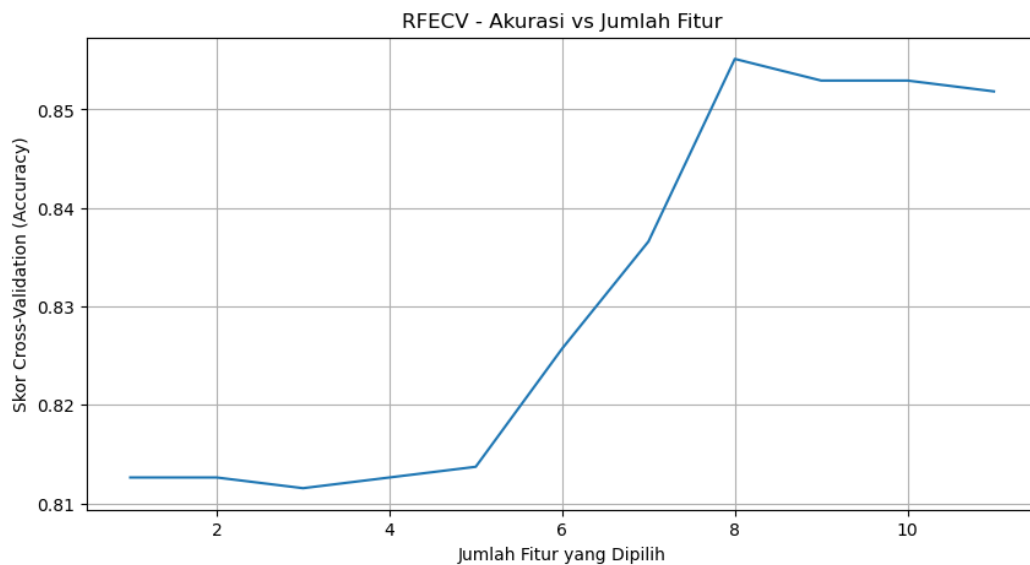
Setelah dilakukan seleksi fitur awal menggunakan metode *Chi-Square*, proses seleksi dilanjutkan dengan *Recursive Feature Elimination with Cross-Validation* (RFECV) menggunakan SVM dengan kernel linear sebagai estimator. RFECV bertujuan untuk mengidentifikasi jumlah fitur optimal yang memberikan performa terbaik berdasarkan akurasi validasi silang (*cross-validation*). Hasil akurasi *cross-validation* untuk setiap jumlah fitur ditunjukkan pada Tabel 4.4 berikut:

Tabel 4.4 Akurasi *Cross-Validation* terhadap Jumlah Fitur yang Dipilih

Jumlah Fitur	Akurasi <i>Cross-Validation</i>
1	0.8126
2	0.8126
3	0.8115
4	0.8126
5	0.8137
6	0.8257
7	0.8366
8	0.8551
9	0.8529
10	0.8529
11	0.8518

Dari Tabel 4.4 dapat diamati bahwa jumlah fitur terbaik berada pada angka delapan fitur, yang memberikan akurasi tertinggi. Terlihat bahwa akurasi cenderung meningkat seiring dengan bertambahnya jumlah fitur, hingga mencapai puncaknya pada saat delapan fitur digunakan, yaitu dengan akurasi sebesar 0.8551 (85.51%).

Setelah itu, penambahan fitur justru tidak memberikan peningkatan performa yang signifikan, bahkan sedikit menurun. Gambar 4.1 berikut menunjukkan visualisasi dari tren akurasi terhadap jumlah fitur yang dipilih.



Gambar 4.1 Kurva RFECV - Akurasi vs Jumlah Fitur yang Dipilih

Pada Gambar 4.1 menampilkan grafik hubungan antara jumlah fitur yang digunakan dengan akurasi *cross-validation* model. Sumbu x menunjukkan jumlah fitur (1–11) dan sumbu y menunjukkan skor akurasi. Terlihat bahwa penambahan fitur dari 1 hingga 5 tidak berdampak signifikan terhadap akurasi yang tetap stabil di kisaran 81.2%–81.3%. Peningkatan mulai tampak pada jumlah fitur ke-6 dengan akurasi 82.57%, dan terus naik hingga mencapai puncak pada fitur ke-8 dengan akurasi tertinggi sebesar 85.51%. Setelah itu, penambahan fitur justru menyebabkan penurunan akurasi secara perlahan. Delapan fitur terbaik yang dipilih berdasarkan RFECV dengan estimator SVM linear dapat dilihat pada Tabel 4.5 Fitur Terpilih dari RFECV.

Tabel 4.5 Fitur Terpilih dari RFECV

No.	Nama Fitur
1	sex
2	chest pain type
3	cholesterol
4	fasting blood sugar
5	max heart rate
6	exercise angina
7	oldpeak
8	ST slope

Selanjutnya, dilakukan proses interseksi untuk mengidentifikasi fitur-fitur yang secara konsisten terpilih oleh kedua metode seleksi. Tujuan interseksi ini adalah menentukan fitur yang paling relevan dan stabil dalam memengaruhi hasil klasifikasi penyakit kardiovaskular, sebagaimana ditampilkan pada Tabel 4.6 Hasil Interseksi Fitur dari Metode *Chi-Square* dan RFE.

Tabel 4.6 Hasil Interseksi Fitur dari Metode *Chi-Square* dan RFE

No.	Nama Fitur
1	sex
2	chest pain type
3	fasting blood sugar
4	max heart rate
5	exercise angina
6	oldpeak
7	ST slope

Tabel 4.6 menunjukkan fitur-fitur yang digunakan pada tahap pelatihan model SVM merupakan hasil dari proses seleksi fitur secara bertahap. Pertama, dilakukan

seleksi awal menggunakan metode *Chi-Square* untuk mengeliminasi fitur yang kurang relevan secara statistik terhadap label target. Selanjutnya, dilakukan proses seleksi lanjutan menggunakan metode RFECV dengan estimator SVM kernel linear untuk mengidentifikasi kombinasi fitur yang memberikan akurasi tertinggi pada validasi silang. Hasil akhir fitur terpilih merupakan *intersection* atau irisan dari dua metode seleksi tersebut, yaitu fitur-fitur yang secara konsisten dianggap penting oleh kedua metode.

4.4 Hasil Uji Coba

Pada subbab ini, penulis akan membahas hasil uji coba klasifikasi penyakit kardiovaskular menggunakan metode SVM. proses uji coba dilakukan berdasarkan dua faktor utama, yaitu jenis kernel pada algoritma SVM dan penerapan seleksi fitur. Setiap kombinasi pengujian dilakukan menggunakan pendekatan *5-Fold Cross Validation* untuk memperoleh hasil evaluasi model yang lebih stabil dan representatif. Jenis kernel yang digunakan meliputi linear, RBF, serta polinomial dengan derajat 3, 5, dan 7. Evaluasi performa model pada setiap kombinasi dilakukan dengan menggunakan *confusion matrix* yang menghasilkan metrik akurasi, presisi, *recall*, dan *f1-score* sebagai dasar penilaian kinerja model.

4.4.1 Hasil Uji Coba Tanpa Seleksi Fitur

Pada tahap ini, model SVM diuji tanpa menggunakan teknik seleksi fitur. Seluruh fitur pada *dataset* digunakan secara utuh dalam proses pelatihan dan pengujian. Evaluasi dilakukan menggunakan pendekatan *5-Fold Cross Validation*, di mana data dibagi menjadi lima bagian dan model dilatih serta diuji secara bergantian. Evaluasi dilakukan berdasarkan *confusion matrix* yang menghasilkan

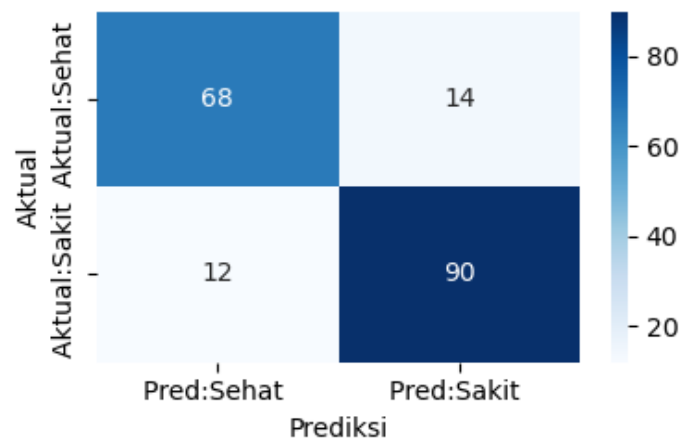
empat metrik utama: akurasi, presisi, *recall*, dan *f1-score*. Model diuji dengan lima jenis kernel yaitu linear, RBF, polinomial *degree* 3, 5, dan 7. Berikut adalah hasil rinci dari setiap kernel:

Pengujian menggunakan kernel linear tanpa seleksi fitur dilakukan dengan 5-*Fold Cross Validation*. Dari hasil evaluasi *confusion matrix* didapatkan nilai akurasi, presisi, *Recall*, dan *F-1 score* yang disajikan pada Tabel 4.7.

Tabel 4.7 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Linear

<i>Fold</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Fold 1	0.8533	0.9082	0.8318	0.8683
Fold 2	0.8859	0.9115	0.9035	0.9075
Fold 3	0.8696	0.8600	0.8958	0.8776
Fold 4	0.8306	0.8208	0.8788	0.8488
Fold 5	0.8525	0.8095	0.9239	0.8629
Rata-rata	0.8584	0.8620	0.8868	0.8730

Kernel linear menunjukkan performa yang konsisten di seluruh Fold. Akurasi tertinggi terdapat pada Fold 2 (88.59%) dan terendah pada Fold 4 (83.06%). *Recall* tertinggi dicapai pada Fold 5 (92.39%), menunjukkan kemampuan model dalam mengenali kelas positif secara baik. Evaluasi lebih lanjut ditunjukkan oleh *confusion matrix* rata-rata pada Gambar 4.2.

Gambar 4.2 *Confusion Matrix* Kernel Linear - Semua Fitur

Dari Gambar 4.2 Model mampu mengklasifikasikan 90 pasien sakit (*true positive*) dan 68 pasien sehat (*true negative*) dengan benar. Terdapat 12 pasien sakit yang diprediksi sehat (*false negative*) dan 14 pasien sehat yang diprediksi sakit (*false positive*). Secara keseluruhan, model dengan kernel linear tanpa seleksi fitur mampu membedakan kelas sehat dan sakit dengan baik, dengan rata-rata akurasi 85,84%, presisi 86,20%, *recall* 88,68%, dan *F1-score* 87,30%.

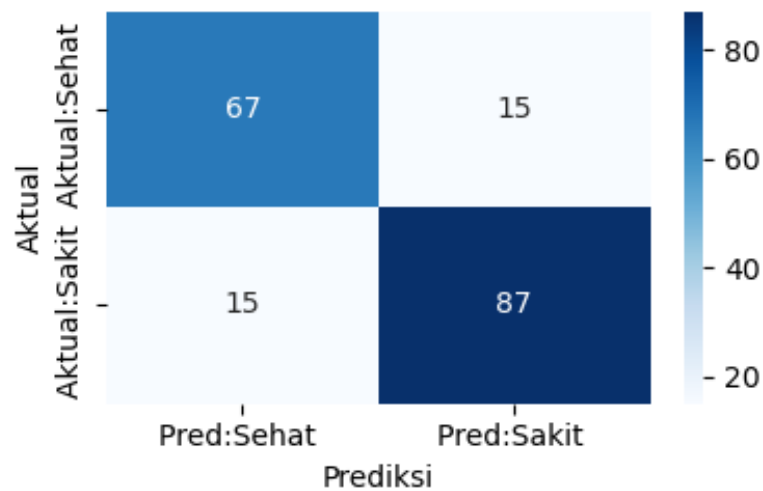
Kemudian, pengujian menggunakan kernel RBF tanpa seleksi fitur dilakukan dengan *5-Fold Cross Validation*. Dari hasil evaluasi *confusion matrix* didapatkan nilai akurasi, presisi, *Recall*, dan *F-1 score* yang disajikan pada Tabel 4.8.

Tabel 4.8 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel RBF

<i>Fold</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-Score</i>
Fold 1	0.8261	0.8866	0.8037	0.8431
Fold 2	0.8533	0.8991	0.8596	0.8789
Fold 3	0.8478	0.8400	0.8750	0.8571

<i>Fold</i>	<i>Akurasi</i>	<i>Presisi</i>	<i>Recall</i>	<i>F1-Score</i>
Fold 4	0.7978	0.8100	0.8182	0.8141
Fold 5	0.8579	0.8235	0.9130	0.8660
Rata-rata	0.8366	0.8518	0.8539	0.8519

Performa model dengan kernel RBF cukup stabil di seluruh Fold, dengan akurasi tertinggi 84.78% pada Fold 2 sedangkan akurasi terendah 79.78% pada Fold 4. *Recall* tertinggi dicapai pada Fold 5 sebesar 91.3%, menunjukkan kemampuan model dalam mengenali data sakit dengan baik. Rata-rata *confusion matrix* ditampilkan pada Gambar 4.3



Gambar 4.3 *Confusion Matrix* Kernel RBF - Semua Fitur

Berdasarkan Gambar 4.3, model berhasil mengklasifikasikan 87 pasien sakit (*true positive*) dan 67 pasien sehat (*true negative*) dengan benar. Namun, masih terdapat 15 pasien sakit yang salah diklasifikasikan sebagai sehat (*false negative*), serta 15 pasien sehat yang diprediksi sakit (*false positive*). Secara

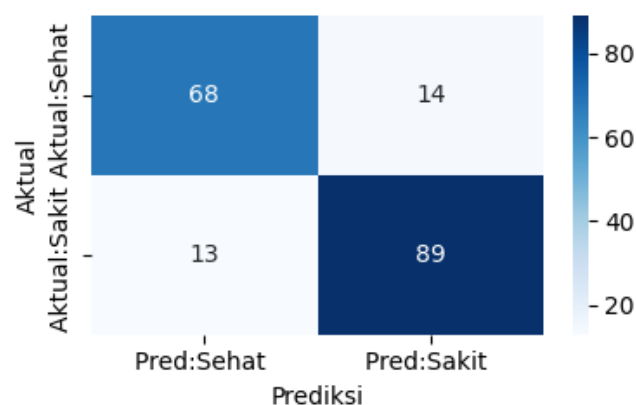
keseluruhan, kernel RBF menunjukkan performa yang cukup baik dengan rata-rata akurasi 83,66%, presisi 85,18%, *recall* 85,39%, dan *F1-score* 85,19%.

Pengujian menggunakan kernel polinomial dengan *degree* 3 tanpa seleksi fitur juga dilakukan dengan *5-Fold Cross Validation*. Hasil evaluasi *confusion matrix* ditampilkan pada Tabel 4.9.

Tabel 4.9 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial *Degree* 3

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.8478	0.8990	0.8318	0.8641
Fold 2	0.8804	0.9259	0.8772	0.9009
Fold 3	0.8859	0.8788	0.9062	0.8923
Fold 4	0.8197	0.8056	0.8788	0.8406
Fold 5	0.8361	0.8039	0.8913	0.8454
Rata-rata	0.8540	0.8626	0.8771	0.8686

Kernel polinomial derajat 3 memberikan performa yang cukup baik dan konsisten di seluruh Fold. Akurasi tertinggi dicapai pada Fold 3 sebesar 88.59%, sedangkan terendah pada Fold 4 dengan 81.97%. Nilai *recall* tertinggi tercatat di Fold 3 (90.62%), kemudian evaluasi dengan *confusion matrix* memiliki rata-rata yang ditunjukkan pada Gambar 4.4.



Gambar 4.4 *Confusion Matrix* Kernel Polinomial *Degree* 3 - Semua Fitur

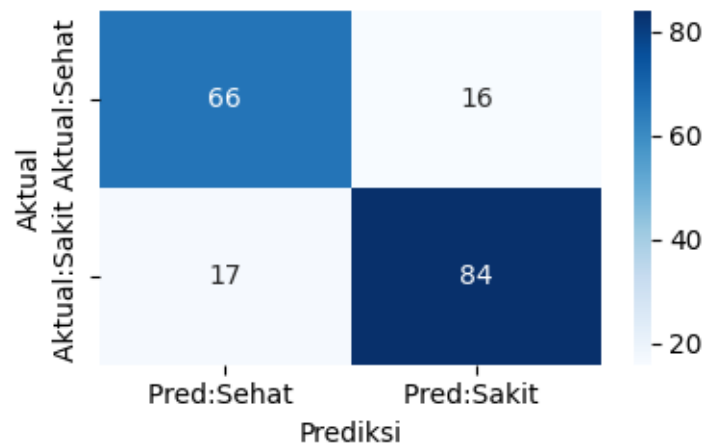
Berdasarkan Gambar 4.4, model berhasil mengklasifikasikan 89 pasien sakit (*true positive*) dan 68 pasien sehat (*true negative*) dengan benar. Kesalahan prediksi terdiri dari 13 pasien sakit yang diklasifikasikan sebagai sehat (*false negative*), dan 14 pasien sehat yang diklasifikasikan sebagai sakit (*false positive*). Secara keseluruhan, kernel ini menunjukkan performa yang cukup seimbang dengan rata-rata akurasi 85,40%, presisi 86,26%, *recall* 87,71%, dan *F1-score* 86,86%.

Selanjutnya, pengujian SVM dengan kernel polinomial derajat 5 dilakukan tanpa seleksi fitur menggunakan *5-Fold Cross Validation*. Hasil evaluasi tiap Fold disajikan pada Tabel 4.10.

Tabel 4.10 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial *Degree* 5

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.8207	0.9111	0.7664	0.8325
Fold 2	0.8315	0.9029	0.8158	0.8571
Fold 3	0.8424	0.8384	0.8646	0.8513
Fold 4	0.7760	0.7788	0.8182	0.7980
Fold 5	0.8251	0.7830	0.9022	0.8384
Rata-rata	0.8191	0.8429	0.8334	0.8355

Kernel polinomial *degree* 5 menunjukkan performa yang sedikit lebih rendah dibanding *degree* 3. Akurasi tertinggi diperoleh pada Fold 3 dengan 84.24% dan terendah di Fold 4 dengan 77.60%. Presisi tertinggi dicapai di Fold 1 dengan 91.11%, sementara *recall* tertinggi ada di Fold 5 dengan 90.22%. Rata-rata *confusion matrix* dari kelima Fold ditampilkan pada Gambar 4.5.



Gambar 4.5 *Confusion Matrix* Kernel Polinomial *Degree 5* - Semua Fitur

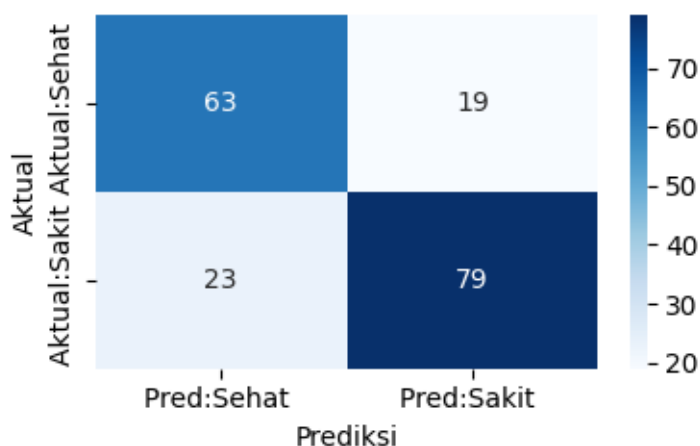
Dari Gambar 4.5, model mampu mengklasifikasikan 84 pasien sakit (*true positive*) dan 66 pasien sehat (*true negative*). Kesalahan klasifikasi terjadi pada 17 pasien sakit yang diprediksi sehat (*false negative*) dan 16 pasien sehat yang diprediksi sakit (*false positive*). Secara keseluruhan, kernel ini menghasilkan rata-rata akurasi 81,91%, presisi 84,29%, *recall* 83,34%, dan *F1-score* 83,55%. Meskipun performanya cukup baik, namun tidak lebih unggul dibanding kernel linear, RBF, maupun polinomial *degree 3*.

Pengujian kernel polinomial derajat 7 juga dilakukan tanpa seleksi fitur menggunakan *5-Fold Cross Validation*. Hasil pengujian disajikan pada Tabel 4.11.

Tabel 4.11 Hasil Evaluasi SVM Tanpa Seleksi Fitur Kernel Polinomial *Degree 7*

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.7283	0.8202	0.6822	0.7449
Fold 2	0.8152	0.9082	0.7807	0.8396
Fold 3	0.7935	0.8021	0.8021	0.8021
Fold 4	0.7541	0.7596	0.7980	0.7783
Fold 5	0.7705	0.7500	0.8152	0.7812
Rata-rata	0.7723	0.8080	0.7756	0.7892

Kernel polinomial *degree* 7 menunjukkan performa paling rendah di antara seluruh kernel yang diuji. Akurasi tertinggi tercatat pada Fold 2 sebesar 81.52%, sedangkan akurasi terendah terjadi pada Fold 1 dengan 72.83%. Presisi terbaik diperoleh pada Fold 2 sebesar 90.82%, namun Fold 5 mencatat presisi terendah 75%. *Confusion matrix* rata-rata untuk kernel ini ditunjukkan pada Gambar 4.6



Gambar 4.6 *Confusion Matrix* Kernel Polinomial *Degree* 7 - Semua Fitur

Berdasarkan Gambar 4.6, model berhasil mengklasifikasikan 79 pasien sakit (*true positive*) dan 63 pasien sehat (*true negative*). Namun, terdapat 23 pasien sakit yang salah diprediksi sebagai sehat (*false negative*), dan 19 pasien sehat salah diprediksi sebagai sakit (*false positive*).

Secara umum, kernel polinomial derajat 7 menghasilkan performa terendah, dengan rata-rata akurasi 77,23%, presisi 80,80%, *recall* 77,56%, dan *F1-score* 78,92%. Hasil ini menunjukkan bahwa kompleksitas model yang terlalu tinggi dapat menurunkan kinerja klasifikasi.

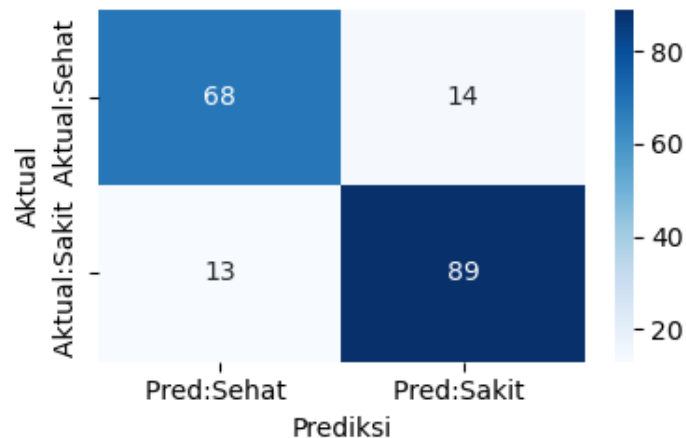
4.4.2 Hasil Uji Coba Dengan Seleksi Fitur

Setelah dilakukan seleksi fitur menggunakan metode *Chi-Square* dan RFE, pengujian kembali dilakukan terhadap model SVM dengan kernel linear. Evaluasi dilakukan menggunakan *5-Fold Cross Validation*. Hasil akurasi, presisi, *Recall*, dan *F-I score* disajikan pada Tabel 4.12.

Tabel 4.12 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Linear

Fold	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Fold 1	0.8423	0.8990	0.8318	0.8641
Fold 2	0.8804	0.9189	0.8947	0.9067
Fold 3	0.8695	0.8529	0.9062	0.8788
Fold 4	0.8087	0.8137	0.8384	0.8259
Fold 5	0.8525	0.8218	0.9022	0.8601
Rata-rata	0.8496	0.8613	0.8747	0.8671

Dari hasil di atas, dapat dilihat bahwa seleksi fitur memberikan dampak positif terhadap performa model. Dengan rata-rata akurasi 84.96%, dengan presisi 86.13%, *recall* 87.47%, dan *F1-score* 86.71%. Fold 2 menunjukkan performa terbaik dengan akurasi sebesar 88.59%, sementara Fold 4 memiliki akurasi terendah yakni 80.87%. *Confusion matrix* rata-rata dari model kernel linear setelah seleksi fitur ditampilkan pada Gambar 4.7.



Gambar 4.7 *Confusion Matrix* Kernel Linear - Seleksi Fitur

Dari *confusion matrix* pada Gambar 4.7, terlihat bahwa model berhasil mengklasifikasikan 89 pasien sakit dengan benar (*true positive*) dan 68 pasien sehat (*true negative*). Terdapat 14 kasus sehat tapi diprediksi sakit (*false positive*), dan 13 kasus sakit tapi diprediksi sehat (*false negative*). Perbandingan dengan kernel linear tanpa seleksi fitur menunjukkan perbedaan performa antara kedua model relatif kecil. Meski demikian, seleksi fitur tetap memberikan kontribusi positif dalam menjaga performa model sambil mengurangi kompleksitas data

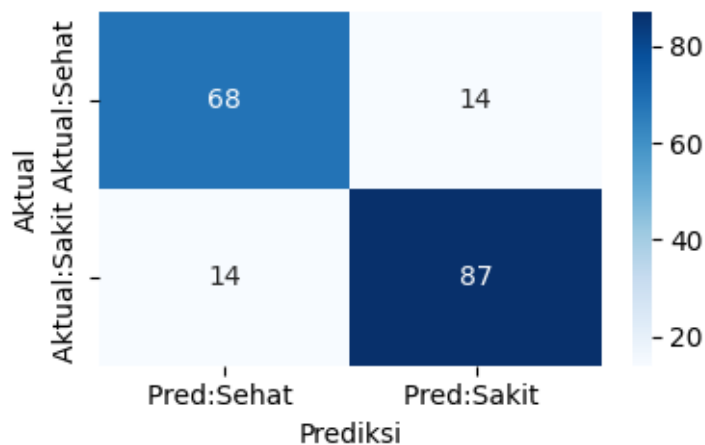
Pengujian pada model SVM menggunakan kernel RBF dengan seleksi fitur *Chi-Square* dan Recursive RFE. Dari hasil evaluasi *confusion matrix* didapatkan nilai akurasi, presisi, *Recall*, dan *F1 score* yang disajikan padat Tabel 4.13.

Tabel 4.13 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel RBF

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.8424	0.9062	0.8131	0.8571
Fold 2	0.8587	0.9074	0.8596	0.8829
Fold 3	0.8641	0.8586	0.8854	0.8718
Fold 4	0.8142	0.8218	0.8384	0.8300
Fold 5	0.8634	0.8317	0.9130	0.8705

Fold	Akurasi	Presisi	Recall	F1-Score
Rata-rata	0.8486	0.8651	0.8619	0.8625

Berdasarkan hasil evaluasi, model kernel RBF menunjukkan kinerja yang cukup stabil dengan rata-rata akurasi sebesar 84.86%, presisi 86.51%, *recall* 86.19%, dan *F1-score* 86.25%. Fold 3 memiliki akurasi tertinggi sebesar 86.41%, sedangkan Fold 4 memiliki akurasi terendah yakni 81.42%. *Confusion matrix* rata-rata hasil prediksi dengan kernel RBF setelah seleksi fitur ditampilkan pada Gambar 4.8.



Gambar 4.8 *Confusion Matrix* Kernel RBF - Seleksi Fitur

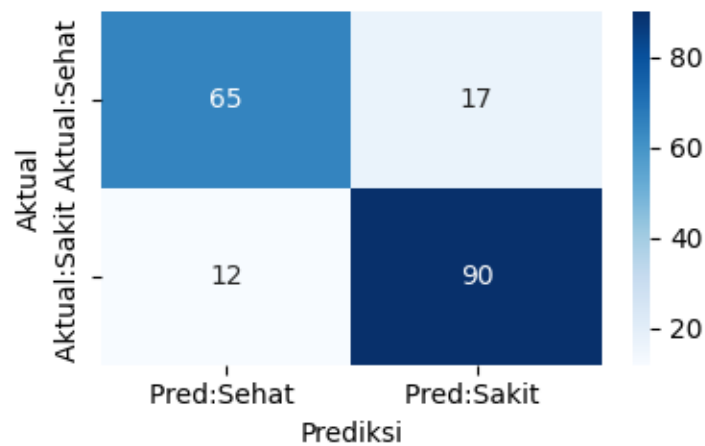
Berdasarkan *confusion matrix* pada Gambar 4.8, model berhasil mengklasifikasikan 87 pasien sakit secara benar dan 68 pasien sehat. Sementara itu, terdapat 14 kasus *false positive* (sehat diklasifikasikan sakit) dan 14 kasus *false negative* (sakit diklasifikasikan sehat). Jika dibandingkan dengan kernel RBF tanpa seleksi fitur, model dengan seleksi fitur menunjukkan peningkatan yang stabil pada seluruh metrik evaluasi, menunjukkan bahwa seleksi fitur efektif dalam meningkatkan performa klasifikasi.

Setelah dilakukan seleksi fitur menggunakan metode *Chi-Square* dan RFE, pengujian juga dilakukan terhadap model SVM dengan kernel Polinomial *degree* 3. Evaluasi dilakukan menggunakan skema *5-Fold Cross Validation*, sebagaimana ditampilkan pada Tabel 4.14.

Tabel 4.14 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial *Degree* 3

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.8207	0.8558	0.8318	0.8436
Fold 2	0.8696	0.9018	0.8860	0.8938
Fold 3	0.8750	0.8476	0.9271	0.8856
Fold 4	0.7978	0.7870	0.8586	0.8213
Fold 5	0.8579	0.8113	0.9348	0.8687
Rata-rata	0.8442	0.8407	0.8876	0.8626

Berdasarkan hasil evaluasi, kernel polinomial *degree* 3 menunjukkan performa yang baik dengan rata-rata akurasi sebesar 84.42%, presisi 84.07%, *recall* 88.76%, dan *F1-score* 86.26%. Fold 3 memberikan hasil akurasi terbaik dengan 87.50% dan Fold 2 dengan *F1-score* 89.38%, sedangkan Fold 4 memberikan performa paling rendah dengan akurasi 79.78%. Gambar 4.9 menunjukkan *confusion matrix* rata-rata dari hasil klasifikasi menggunakan kernel polinomial *degree* 3 setelah seleksi fitur.



Gambar 4.9 *Confusion Matrix* Kernel Polinomial *Degree 3* - Seleksi Fitur

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 90 pasien sakit secara benar dan 65 pasien sehat. Kesalahan klasifikasi yang terjadi adalah 17 *false positive* dan 12 *false negative*. Jika dibandingkan dengan model kernel polinomial *degree 3* tanpa seleksi fitur, penerapan seleksi fitur menghasilkan peningkatan pada nilai *recall*, dari 87.71% menjadi 88.76%, yang menunjukkan peningkatan kemampuan model dalam mendeteksi kasus positif (pasien sakit). Namun, rata-rata nilai akurasi, presisi, dan *F1-score* mengalami sedikit penurunan.

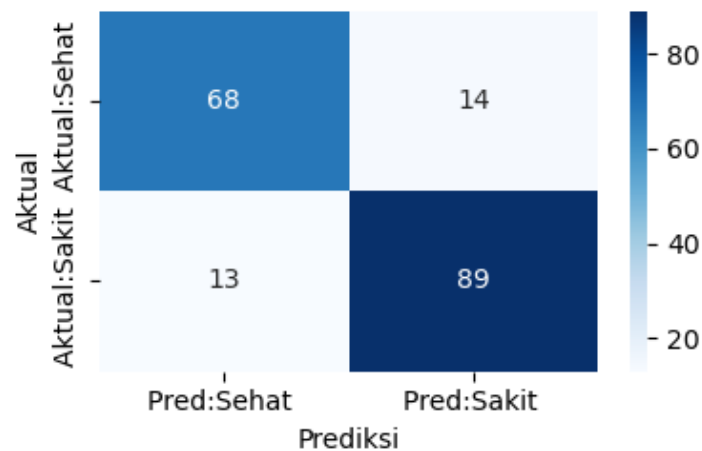
Pengujian selanjutnya dilakukan pada model SVM dengan kernel polinomial *degree 5*, yang dilatih menggunakan fitur-fitur hasil seleksi dari metode *Chi-Square* dan RFE. Evaluasi kinerja model menggunakan *5-Fold Cross Validation*, seperti yang ditunjukkan pada Tabel 4.15.

Tabel 4.15 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial *Degree 5*

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 1	0.8478	0.9247	0.8037	0.8600
Fold 2	0.8750	0.9333	0.8596	0.8950
Fold 3	0.8696	0.8396	0.9271	0.8812
Fold 4	0.8306	0.8148	0.8889	0.8502

Fold	Akurasi	Presisi	Recall	F1-Score
Fold 5	0.8415	0.8000	0.9130	0.8528
Rata-rata	0.8529	0.8625	0.8785	0.8678

Dari Tabel 4.15 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial Degree 5 diatas, dapat dilihat bahwa kernel polinomial *degree* 5 memberikan performa yang cukup stabil dengan rata-rata akurasi 85.29%, presisi 86.25%, *recall* 87.85%, dan *F1-score* sebesar 86.78%. Nilai tertinggi diperoleh pada Fold 2 dengan akurasi 87.5% dan *F1-score* 89.5%. Gambar 4.10 menampilkan *confusion matrix* rata-rata untuk kernel polinomial *degree* 5 setelah dilakukan seleksi fitur.



Gambar 4.10 *Confusion Matrix* Kernel Polinomial Degree 5 - Seleksi Fitur

Berdasarkan Gambar 4.10, model mampu mengklasifikasikan 89 pasien sakit secara benar (*true positive*) dan 68 pasien sehat (*true negative*). Kesalahan klasifikasi terdiri atas 15 *false positive* dan 14 *false negative*. Jika dibandingkan dengan model kernel polinomial *degree* 5 tanpa seleksi fitur, model dengan *degree* 5 menunjukkan peningkatan performa yang cukup konsisten.

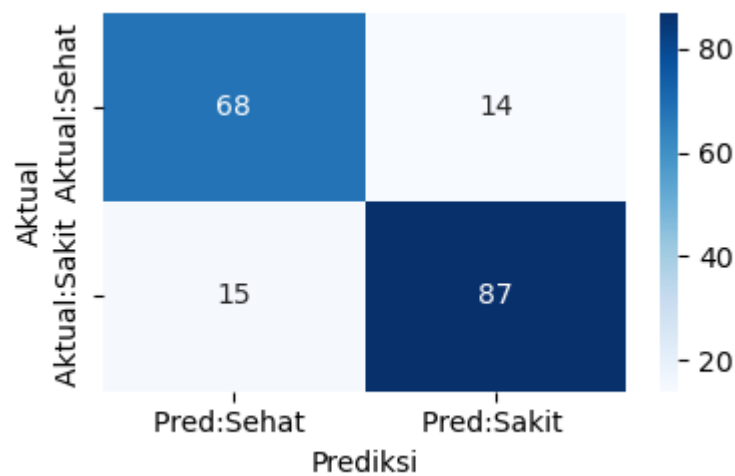
Rata-rata nilai akurasi meningkat dari 81.91% menjadi 85.29%, presisi dari 84.29% menjadi 86.25%, *recall* dari 83.34% menjadi 87.85%, dan *F1-score* dari 83.55% menjadi 86.78%.

Pengujian terakhir dilakukan menggunakan model SVM dengan kernel polinomial *degree* 7. Seperti sebelumnya, model dilatih menggunakan fitur-fitur hasil seleksi dari metode *Chi-Square* dan RFE. Evaluasi dilakukan dengan metode *5-Fold Cross Validation*, dan hasil lengkap ditampilkan pada Tabel 4.16.

Tabel 4.16 Hasil Evaluasi SVM dengan Seleksi Fitur Kernel Polinomial *Degree* 7

Fold	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Fold 1	0.8533	0.9255	0.8131	0.8657
Fold 2	0.8370	0.9118	0.8158	0.8611
Fold 3	0.8641	0.8515	0.8958	0.8731
Fold 4	0.8415	0.8365	0.8788	0.8571
Fold 5	0.8197	0.7810	0.8913	0.8325
Rata-rata	0.8431	0.8613	0.8590	0.8579

Dari hasil di atas, terlihat bahwa kernel polinomial *degree* 7 memberikan performa cukup baik dengan rata-rata akurasi 84.31%, presisi 86.13%, *recall* 85.9%, dan *F1-score* 85.79%. Fold 3 menjadi Fold dengan performa tertinggi, terutama pada aspek akurasi yang mencapai 86.41%. Gambar 4.11 berikut merupakan visualisasi *confusion matrix* rata-rata dari seluruh Fold untuk kernel ini.



Gambar 4.11 *Confusion Matrix* Kernel Polinomial *Degree 7* - Seleksi Fitur

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 87 pasien sakit secara benar (*true positive*) dan 68 pasien sehat (*true negative*). Namun, terdapat 14 pasien sehat yang salah diklasifikasikan sebagai sakit (*false positive*) dan 15 pasien sakit yang tidak terdeteksi (*false negative*). Jika dibandingkan antara model tanpa dan dengan seleksi fitur, terlihat bahwa penerapan seleksi fitur memberikan peningkatan performa yang cukup signifikan pada kernel polinomial *degree 7*. Tanpa seleksi fitur, model hanya mampu mencapai rata-rata akurasi sebesar 77.23%, presisi 80.80%, *recall* 77.56%, dan *F1-score* 78.92%. Setelah dilakukan seleksi fitur, nilai rata-rata akurasi meningkat menjadi 84.31%, presisi menjadi 86.13%, *recall* mencapai 85.90%, dan *F1-score* naik menjadi 85.79%. Peningkatan pada seluruh metrik evaluasi ini menunjukkan bahwa seleksi fitur berperan penting dalam menyaring atribut yang relevan dan mengurangi kompleksitas model.

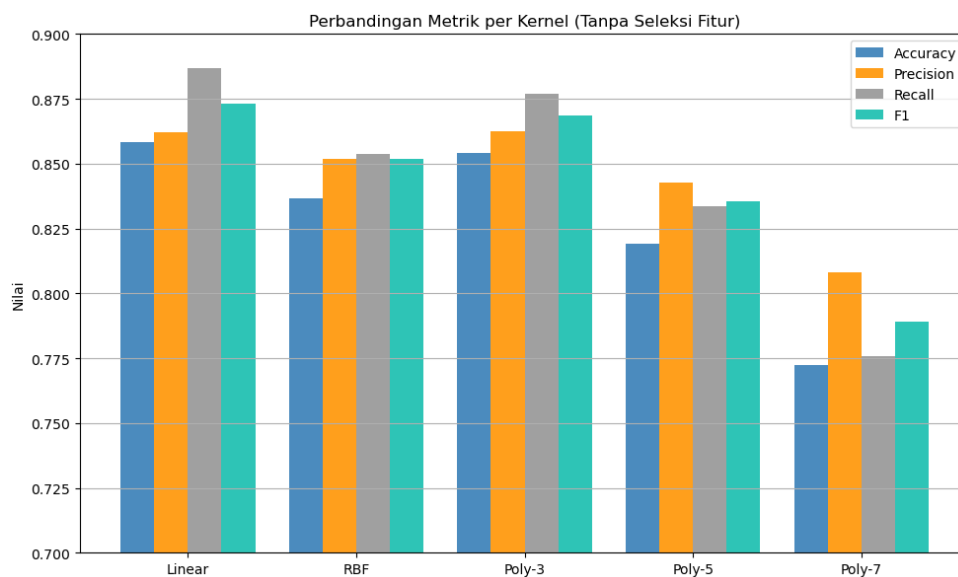
4.5 Pembahasan

Pada penelitian ini telah dilakukan klasifikasi penyakit kardiovaskular menggunakan metode SVM dengan membandingkan performa model tanpa seleksi fitur dan model dengan seleksi fitur. Data yang digunakan telah melalui tahap *preprocessing* seperti penghapusan duplikasi, penanganan *missing value*, dan normalisasi. Selanjutnya dilakukan proses seleksi fitur menggunakan dua metode yaitu *Chi-Square* dan RFE, dan fitur hasil irisan keduanya digunakan untuk pelatihan model pada pendekatan kedua. Evaluasi dilakukan menggunakan *5-Fold Cross Validation* dengan lima jenis kernel yaitu linear, RBF, polinomial *degree 3*, *degree 5*, dan *degree 7*. Metrik evaluasi yang digunakan meliputi akurasi, presisi, *recall*, *F1-score*, dan standar deviasi.

Hasil rata-rata pengujian performa model SVM tanpa seleksi fitur pada masing-masing kernel ditampilkan pada Tabel 4.17 dan Gambar 4.12.

Tabel 4.17 Performa Model SVM Tanpa Seleksi Fitur

Kernel	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
Linear	0.8584	0.8620	0.8868	0.8730
RBF	0.8366	0.8518	0.8539	0.8519
Polinomial d = 3	0.8540	0.8626	0.8771	0.8686
Polinomial d = 5	0.8191	0.8429	0.8334	0.8355
Polinomial d = 7	0.7723	0.8080	0.7756	0.7892



Gambar 4.12 Perbandingan Metrik Evaluasi Tanpa Seleksi Fitur

Berdasarkan hasil pengujian pada Tabel 4.17 dan Gambar 4.12 Perbandingan Metrik Evaluasi tanpa seleksi fitur, diketahui bahwa kernel linear menunjukkan performa terbaik tanpa seleksi fitur, dengan akurasi sebesar 85,84%, presisi 86,20%, *recall* 88,68%, dan *F1-score* 87,30%. Kernel polinomial derajat 3 menyusul dengan hasil yang kompetitif, yakni akurasi 85,40% dan *F1-score* 86,86%, menunjukkan bahwa kompleksitas menengah masih dapat mengimbangi kernel linear. Kernel RBF juga memberikan hasil stabil dengan akurasi 83,66%, *F1-score* 85,19%, namun cenderung lebih rendah. Hal ini menunjukkan bahwa kernel linear mampu membangun model klasifikasi yang cukup baik meskipun tanpa proses reduksi fitur, kemungkinan karena kompleksitas model yang relatif rendah dan ketahanannya terhadap fitur-fitur yang tidak terlalu relevan.

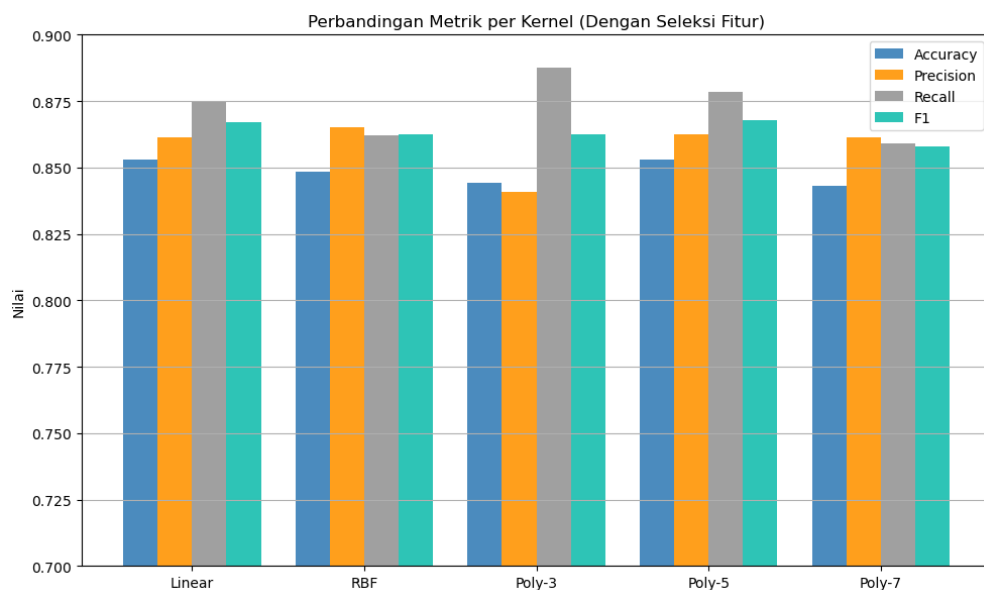
Namun demikian, beberapa kernel lain seperti polinomial derajat 5 dan 7 menunjukkan performa yang kurang optimal tanpa seleksi fitur, dengan akurasi masing-masing hanya sebesar 81,91% dan 77,23%. Hal ini mengindikasikan bahwa

model dengan kernel polinomial tingkat tinggi cenderung *overfitting* terhadap fitur-fitur yang tidak memberikan kontribusi signifikan dalam proses klasifikasi.

Selanjutnya yaitu pengujian performa model SVM dengan seleksi fitur pada masing-masing kernel ditampilkan pada Tabel 4.18 dan Gambar 4.13.

Tabel 4.18 Performa Model SVM dengan Seleksi Fitur

Kernel	Akurasi	Presisi	Recall	F1-Score
Linear	0.8529	0.8613	0.8747	0.8671
RBF	0.8486	0.8651	0.8619	0.8625
Polinomial d=3	0.8442	0.8407	0.8876	0.8626
Polinomial d=5	0.8529	0.8625	0.8785	0.8678
Polinomial d=7	0.8431	0.8613	0.8590	0.8579



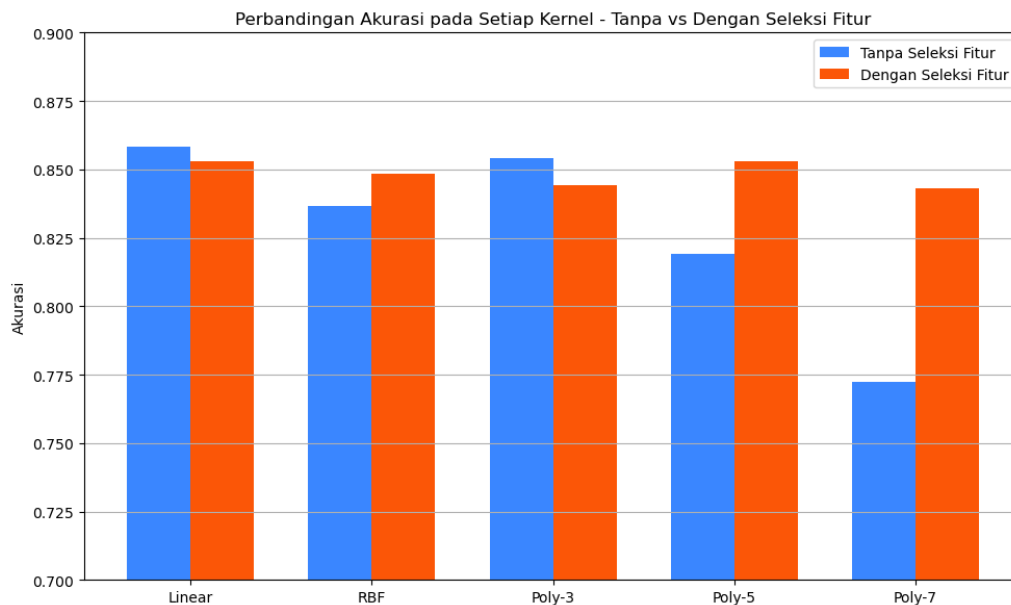
Gambar 4.13 Perbandingan Metrik Evaluasi dengan Seleksi Fitur

Berdasarkan hasil evaluasi yang disajikan pada Tabel 4.18 dan Gambar 4.13, penerapan seleksi fitur memberikan dampak yang bervariasi terhadap performa masing-masing kernel pada model SVM. Kernel polinomial derajat 5 mencatatkan performa terbaik secara keseluruhan dengan akurasi 85,29%, presisi 86,25%, *recall*

87,85%, dan *F1-score* 86,78%, menunjukkan bahwa pada derajat kompleksitas menengah, seleksi fitur mampu memperkuat kinerja model dalam mengidentifikasi kelas target secara akurat dan seimbang. Kernel linear, yang pada pengujian tanpa seleksi fitur menunjukkan performa tertinggi, kini memiliki akurasi yang sama yaitu 85,29%, namun dengan sedikit penurunan pada *recall* menjadi 87,47% dan *F1-score* 86,71%, mengindikasikan bahwa reduksi fitur berdampak kecil namun tetap menjaga kestabilan performa.

Kernel RBF menunjukkan peningkatan jika dibandingkan dengan hasil tanpa seleksi fitur, dengan akurasi mencapai 84,86%, presisi 86,51%, *recall* 86,19%, dan *F1-score* 86,25%, mencerminkan keseimbangan yang baik dalam menangani data non-linear pasca seleksi fitur. Sementara itu, kernel polinomial derajat 3 mencatat akurasi 84,42% dan *recall* tertinggi di antara semua kernel sebesar 88,76%, namun presisi-nya relatif lebih rendah yaitu 84,07%, sehingga menghasilkan *F1-score* 86,26%. Ini menunjukkan bahwa kernel ini sangat sensitif dalam mengenali kelas positif, meskipun dengan peningkatan risiko *false positive*.

Terakhir, kernel polinomial derajat 7 mencatat performa paling rendah dengan akurasi 84,31% dan *F1-score* 85,79%, meskipun presisi dan *recall*-nya masih kompetitif. Secara umum, hasil ini menunjukkan bahwa penerapan seleksi fitur mampu mempertahankan bahkan meningkatkan kinerja beberapa kernel, khususnya polinomial derajat 5 dan RBF. Hal ini memperkuat peran penting seleksi fitur dalam mereduksi kompleksitas data tanpa mengorbankan kualitas prediksi model, sekaligus menunjukkan bahwa pemilihan kernel yang tepat tetap menjadi faktor kunci dalam optimasi performa klasifikasi.



Gambar 4.14 Perbandingan Akurasi pada Setiap Kernel

Gambar 4.14 menunjukkan perbandingan akurasi algoritma SVM pada berbagai jenis kernel, baik sebelum maupun sesudah penerapan kombinasi metode seleksi fitur RFE dan Chi-Square. Secara umum, penerapan seleksi fitur memberikan pengaruh yang berbeda pada setiap kernel, tergantung tingkat kompleksitas model yang digunakan.

Pada kernel dengan kompleksitas tinggi, seperti *Polynomial derajat 5* dan *Polynomial derajat 7*, terjadi peningkatan akurasi yang signifikan setelah seleksi fitur. Hal ini terjadi karena kernel kompleks cenderung lebih sensitif terhadap fitur yang tidak relevan atau berlebihan. Dengan menghapus fitur-fitur tersebut, model menjadi lebih sederhana dan fokus, sehingga dapat mengenali pola data dengan lebih baik serta mengurangi risiko *overfitting*. Hal yang sama juga terjadi pada kernel RBF, di mana seleksi fitur membuat model lebih stabil karena parameter kernel tidak lagi dipengaruhi oleh dimensi fitur yang terlalu banyak.

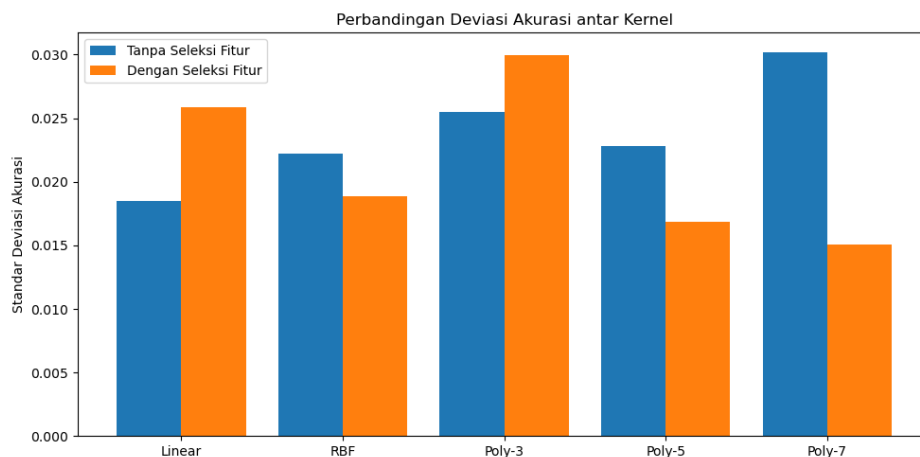
Sebaliknya, pada kernel sederhana seperti Linear dan Polynomial derajat 3, efek seleksi fitur tidak meningkatkan performa secara signifikan, bahkan sedikit menurunkan akurasi. Hal ini disebabkan karena data penyakit kardiovaskular memiliki pola yang cenderung *linearly separable*, sehingga setiap fitur masih memberikan kontribusi penting terhadap pembentukan garis keputusan. Menghapus sebagian fitur dapat mengurangi sedikit informasi yang berguna bagi model linear tersebut.

Analisis terhadap standar deviasi akurasi dari masing-masing kernel pada model SVM juga dilakukan untuk mengevaluasi kestabilan performa model baik sebelum maupun sesudah dilakukan seleksi fitur. Berdasarkan Tabel Tabel 4.19 dapat dilihat bahwa nilai rata-rata standar deviasi akurasi tanpa seleksi fitur adalah sebesar 0.0238, sedangkan setelah seleksi fitur menurun menjadi 0.0213. Penurunan ini menunjukkan bahwa model yang dibangun menggunakan fitur hasil seleksi memiliki variasi performa yang lebih kecil dibandingkan model yang menggunakan seluruh fitur.

Tabel 4.19 Standar Deviasi Akurasi antar Kernel

Kernel	Tanpa Seleksi Fitur	Dengan Seleksi Fitur
Linear	0.0185	0.0258
RBF	0.0222	0.0189
Polinomial d = 3	0.0255	0.0299
Polinomial d = 5	0.0228	0.0168
Polinomial d = 7	0.0302	0.0151
Rata-rata	0.0238	0.0213

Secara lebih rinci, kernel Polinomial derajat 7 menunjukkan penurunan standar deviasi paling signifikan, dari 0.0302 tanpa seleksi fitur menjadi 0.0151 setelah seleksi fitur. Demikian pula pada kernel Polinomial derajat 5, terjadi penurunan dari 0.0228 menjadi 0.0168. Namun, pada kernel Polinomial derajat 3 dan Linear, seleksi fitur justru menyebabkan sedikit peningkatan standar deviasi. Meskipun demikian, secara keseluruhan, seleksi fitur memberikan dampak positif terhadap kestabilan performa model, terutama pada kernel Polinomial berderajat lebih tinggi. Gambaran visual perbandingan standar deviasi akurasi antar kernel disajikan pada Gambar 4.15, yang menunjukkan secara jelas kecenderungan penurunan deviasi setelah seleksi fitur pada sebagian besar kernel yang diuji.



Gambar 4.15 Perbandingan Standart Deviasi Akurasi antar Kernel

Berdasarkan keseluruhan hasil evaluasi, dapat disimpulkan bahwa penerapan seleksi fitur menggunakan kombinasi *Chi-Square* dan RFE memberikan pengaruh positif terhadap stabilitas dan peningkatan performa model SVM, terutama pada kernel dengan tingkat kompleksitas menengah hingga tinggi seperti polinomial derajat 5 dan RBF. Meskipun pada kernel linear dan polinomial derajat 3 seleksi fitur menyebabkan sedikit penurunan akurasi, performa model secara umum tetap

kompetitif. Penurunan tersebut menunjukkan bahwa sebagian fitur yang dihilangkan masih memiliki peran penting dalam proses klasifikasi pada model sederhana. Di sisi lain, seleksi fitur terbukti mampu mengurangi *overfitting* pada kernel dengan kompleksitas tinggi, sebagaimana terlihat dari peningkatan akurasi pada kernel polinomial derajat 7. Selain itu, penurunan nilai rata-rata standar deviasi akurasi menunjukkan bahwa model dengan fitur terseleksi memiliki performa yang lebih stabil.

Dari sisi waktu pemrosesan, hasil pengujian menunjukkan bahwa model dengan seleksi fitur memiliki waktu komputasi yang lebih singkat dibandingkan model tanpa seleksi fitur. Hal ini terjadi karena jumlah fitur yang digunakan berkurang, sehingga proses perhitungan pada tahap pelatihan dan pengujian menjadi lebih efisien. Meskipun terdapat variasi waktu eksekusi antar percobaan yang disebabkan oleh kondisi sistem dan beban pemrosesan, tren secara keseluruhan menunjukkan peningkatan efisiensi setelah penerapan seleksi fitur. Dengan demikian, reduksi fitur tidak hanya berkontribusi terhadap kestabilan model, tetapi juga membantu mempercepat proses komputasi tanpa mengurangi kualitas hasil klasifikasi.

Dalam konteks medis, seluruh fitur pada data pasien pada dasarnya memiliki nilai penting karena setiap atribut dapat merepresentasikan kondisi klinis tertentu yang relevan terhadap diagnosis. Menghilangkan fitur tanpa pertimbangan dapat berisiko mengurangi informasi penting yang berdampak pada interpretasi medis dan keputusan klinis. Oleh karena itu, seleksi fitur dalam penelitian ini tidak dimaksudkan untuk sepenuhnya menggantikan fitur-fitur asli, melainkan untuk

menguji sejauh mana model masih dapat mempertahankan performa dengan subset fitur yang lebih kecil. Hasil pengujian menunjukkan bahwa meskipun seleksi fitur tidak secara langsung meningkatkan akurasi, namun mampu menurunkan rata-rata standar deviasi akurasi dari 0,0238 menjadi 0,0213. Penurunan ini mengindikasikan bahwa model menjadi lebih stabil dan memiliki variasi performa yang lebih rendah antar lipatan validasi, yang penting dalam konteks medis untuk memastikan konsistensi hasil klasifikasi ketika diterapkan pada data baru. Dengan demikian, dapat disimpulkan bahwa seleksi fitur memberikan kontribusi terhadap stabilitas model tanpa mengorbankan kualitas prediksi secara signifikan, serta tetap menjaga keseimbangan antara kompleksitas model dan keandalan dalam klasifikasi penyakit kardiovaskular.

4.6 Integrasi Islam

Penelitian ini mengembangkan model klasifikasi penyakit kardiovaskular menggunakan algoritma SVM yang berbasis kombinasi fitur dari RFE dan *Chi-Square*. Tujuan dari penelitian ini adalah untuk mengetahui pengaruh kombinasi seleksi fitur terhadap performa akurasi dalam klasifikasi penyakit jantung. Dalam perspektif keislaman, konsep akurasi dan ketelitian dalam penelitian ini sejalan dengan firman Allah SWT. dalam QS. Al-Anbiya ayat 47:

وَنَضَعُ الْمَوَازِينَ الْقِسْطَ لِيَوْمِ الْقِيَمَةِ فَلَا تُظْلَمُ نَفْسٌ شَيْئًا وَإِنْ كَانَ مِثْقَالُ حَبَّةٍ مِنْ خَرْدَلٍ أَتَيْنَا بِهَا وَكَفَى بِنَا حَسِيبِينَ

“Kami akan meletakkan timbangan (amal) yang tepat pada hari Kiamat, sehingga tidak seorang pun dirugikan walaupun sedikit. Sekalipun (amal itu) hanya seberat biji sawi, pasti Kami mendatangkannya. Cukuplah Kami sebagai pembuat perhitungan.” (Q.S. Al-Anbiya [21]: 47)

Menurut Tafsir Muharar al-Wajiz, ayat ini menegaskan bahwa Dalam menilai perbuatan hamba-hamba-Nya di akhirat, Allah SWT menjamin akan menegakkan keadilan yang sebenarnya. Dan Kami akan memasang timbangan yang tepat pada hari Kiamat, dengan data yang objektif dan akurat; maka tidak seorang pun dirugikan walau sedikit, sehingga tidak ada seorang hamba yang amal kebbaikannya dikurangi atau kejahatannya dilebih-lebihkan, sekalipun hanya seberat biji sawi, pasti Kami mendatangkan pahala untuk perbuatan baik dan hukuman untuk perbuatan jahat. Dan cukuplah Kami yang membuat perhitungan terhadap perbuatan manusia dengan seadil-adilnya, objektif, tepat, dan akurat. Ayat ini menegaskan bahwa Allah SWT. memiliki sistem perhitungan yang sempurna, di mana setiap amal sekecil apa pun tetap diperhatikan dan ditimbang dengan keadilan tanpa adanya kekeliruan.

Relevansinya dengan penelitian ini adalah bahwa prinsip keadilan, ketelitian, dan akurasi yang digambarkan dalam ayat tersebut menjadi cerminan dalam pengolahan data dan evaluasi model klasifikasi penyakit kardiovaskular. Sebagaimana Allah SWT menimbang amal perbuatan manusia dengan timbangan yang tepat tanpa ada yang terlewat, penelitian ini juga menekankan pentingnya pemrosesan data yang teliti dan objektif melalui tahapan seleksi fitur dengan metode RFE dan *Chi-Square* agar hanya fitur yang relevan yang dipertahankan. Hal ini bertujuan untuk memperoleh model SVM yang lebih akurat, seimbang, dan adil dalam melakukan prediksi, sehingga tidak ada data yang terabaikan maupun diperlakukan secara tidak proporsional. Dengan demikian, penelitian ini tidak

hanya menghasilkan kontribusi ilmiah berupa peningkatan performa akurasi model klasifikasi, tetapi juga selaras dengan nilai-nilai Islam yang menekankan pentingnya keadilan, ketelitian, dan keakuratan dalam setiap bentuk perhitungan dan pengambilan keputusan.

Penelitian ini juga dapat mendukung upaya pencegahan dan penanganan lebih cepat, terutama di lingkungan medis yang memerlukan keputusan berbasis data.

Upaya ini sejalan dengan firman Allah SWT dalam Surah Yunus ayat 57:

يَا أَيُّهَا النَّاسُ قَدْ جَاءَكُمْ مَوْعِظَةٌ مِنْ رَبِّكُمْ وَشِفَاءٌ لِمَا فِي الصُّدُورِ وَهُدًى وَرَحْمَةٌ لِّلْمُؤْمِنِينَ

“Wahai manusia, sungguh telah datang kepadamu pelajaran (Al-Qur’an) dari Tuhanmu, penyembuh bagi sesuatu (penyakit) yang terdapat dalam dada, dan petunjuk serta rahmat bagi orang-orang mukmin.” (Q.S. Yunus [10]: 57)

Menurut tafsir Ibnu Katsir, istilah "maw'izhah" dalam Al-Qur'an merujuk pada nasihat dan peringatan dari Allah SWT, sementara ungkapan "penyembuh bagi penyakit-penyakit dalam dada" mencakup penyakit hati seperti keraguan dan kebodohan, serta dapat dimaknai secara lebih luas sebagai petunjuk dalam mengatasi berbagai persoalan kehidupan, termasuk kesehatan fisik. Dalam konteks penelitian ini, ayat tersebut menjadi dasar spiritual bahwa segala bentuk ilmu pengetahuan, termasuk kedokteran dan teknologi kesehatan, adalah bagian dari rahmat dan petunjuk Allah SWT. Oleh karena itu, pengembangan model klasifikasi penyakit kardiovaskular tidak hanya merupakan aktivitas teknis semata, melainkan juga mengandung misi kemanusiaan dan keimanan, karena berkontribusi dalam menghadirkan solusi ilmiah atas penyakit yang pada hakikatnya merupakan bagian dari ilmu Allah SWT. Nilai kemanusiaan tersebut juga diperkuat dalam Q.S. Al-

Ma'idah ayat 32, yang menegaskan bahwa menyelamatkan satu nyawa seolah-olah telah menyelamatkan seluruh umat manusia:

وَمَنْ أَحْيَاهَا فَكَأَنَّمَا أَحْيَا النَّاسَ جَمِيعًا وَلَقَدْ جَاءَهُمْ رَسُولُنَا بِالْبَيِّنَاتِ ثُمَّ إِنَّ كَثِيرًا مِّنْهُمْ بَعْدَ ذَلِكَ فِي الْأَرْضِ لَمُسْرِفُونَ

“barangsiapa yang memelihara kehidupan seorang manusia, dia seakan-akan telah memelihara kehidupan semua manusia. Sungguh, rasul-rasul Kami benar-benar telah datang kepada mereka dengan (membawa) keterangan-keterangan yang jelas. Kemudian, sesungguhnya banyak di antara mereka setelah itu melampaui batas di bumi.” (Q.S. Al-ma'idah [5]: 32)

Ayat ini menjadi dasar prinsip penting dalam Islam mengenai perlindungan terhadap nyawa manusia. Meskipun penelitian ini tidak secara langsung menyembuhkan penyakit, teknologi yang dikembangkan memiliki peran signifikan dalam deteksi dini penyakit, yang pada akhirnya dapat mencegah risiko fatal akibat keterlambatan penanganan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil menunjukkan bahwa penerapan kombinasi metode RFE dan Chi-Square pada algoritma SVM mampu meningkatkan kinerja klasifikasi penyakit kardiovaskular. Berdasarkan hasil pengujian, kernel Polynomial derajat 7 mengalami peningkatan akurasi dari 77,23% menjadi 84,31%, sedangkan kernel RBF meningkat dari 83,66% menjadi 84,86% setelah penerapan seleksi fitur. Peningkatan ini terjadi karena proses seleksi fitur berhasil menghilangkan atribut yang tidak relevan atau berlebihan, sehingga model menjadi lebih fokus mengenali pola data yang signifikan dan terhindar dari *overfitting*. Efek ini paling terasa pada kernel dengan kompleksitas menengah hingga tinggi karena jenis kernel tersebut sangat sensitif terhadap jumlah dan kualitas fitur yang digunakan dalam pembentukan fungsi keputusan.

Sementara itu, kernel Linear tetap memberikan performa terbaik secara keseluruhan dengan akurasi 85,84% tanpa seleksi fitur, yang hanya sedikit menurun menjadi 85,29% setelah seleksi fitur. Hal ini disebabkan karena data penyakit kardiovaskular memiliki kecenderungan *linearly separable*, sehingga setiap fitur memberikan kontribusi penting terhadap pembentukan garis keputusan. Penghapusan sebagian fitur pada model linear justru mengurangi sedikit informasi yang relevan, sehingga tidak terjadi peningkatan akurasi yang signifikan. Namun demikian, penerapan seleksi fitur tetap memberikan dampak positif terhadap kestabilan performa model, yang terlihat dari penurunan nilai standar deviasi

akurasi dari 0,0238 menjadi 0,0213, menandakan hasil prediksi model menjadi lebih stabil terhadap variasi data pelatihan.

Dari sisi efisiensi komputasi, model dengan seleksi fitur menunjukkan waktu pemrosesan yang lebih cepat dibandingkan model tanpa seleksi fitur, karena jumlah atribut yang diolah lebih sedikit. Dengan demikian, proses pelatihan dan pengujian menjadi lebih efisien tanpa mengurangi kualitas hasil klasifikasi.

Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi metode Chi-Square dan RFE efektif dalam meningkatkan kestabilan, dan efisiensi komputasi algoritma SVM, terutama pada kernel dengan kompleksitas tinggi. Pendekatan ini dapat dijadikan alternatif pengembangan sistem pendukung keputusan medis berbasis *machine learning* yang lebih akurat, stabil, dan efisien untuk mendukung proses klasifikasi penyakit kardiovaskular.

5.2 Saran

Berdasarkan proses dan hasil yang telah dicapai dalam penelitian ini, peneliti menyadari masih terdapat berbagai keterbatasan yang perlu diperbaiki. Oleh karena itu, saran-saran berikut ditujukan bagi penelitian selanjutnya agar dapat memperoleh hasil yang lebih optimal:

1. Menggunakan *dataset* yang lebih besar dan beragam, *dataset* dengan jumlah data yang lebih banyak akan membantu model mempelajari pola secara lebih mendalam dan akurat. Selain itu, *dataset* yang besar juga mampu mengurangi risiko *overfitting*, sehingga model dapat melakukan generalisasi yang lebih baik terhadap data baru.

2. Melakukan eksplorasi algoritma klasifikasi lainnya selain SVM, seperti algoritma seperti *Decision Tree*, *Random Forest*, *Gaussian Naive Bayes*, atau *Neural Network (NN)* dapat dijadikan alternatif pembandingan. Setiap algoritma memiliki karakteristik dan keunggulan masing-masing yang mungkin dapat menghasilkan performa klasifikasi yang lebih tinggi pada *dataset* yang sama.
3. Mencoba eksplorasi metode seleksi fitur lainnya, seperti LASSO (*Least Absolute Shrinkage and Selection Operator*) atau metode berbasis *Tree* seperti *Random Forest* untuk membandingkan hasil seleksi fitur dan efektivitas model.
4. Melakukan optimasi *hyperparameter* secara menyeluruh, penyesuaian parameter seperti *C*, *gamma*, dan jenis *kernel* dalam SVM dapat sangat memengaruhi performa model. Oleh karena itu, penggunaan teknik seperti *Grid Search*, *Random Search*, atau *Bayesian Optimization* perlu dipertimbangkan untuk menemukan konfigurasi terbaik yang mampu memaksimalkan akurasi.

DAFTAR PUSTAKA

- Abdurrohman, M. F. (2024). Klasifikasi kanker kulit melanoma menggunakan Gray Level Co-Occurrence Matrix dan Ensemble Support Vector Machine berbasis citra kulit (Skripsi sarjana, Universitas Islam Negeri Maulana Malik Ibrahim Malang). <http://etheses.uin-malang.ac.id/72147/>
- Alifah, Siswantining, T., Sarwinda, D., & Bustamam, A. (2020). RFE and Chi-Square Based Feature Selection Approach for Detection of Diabetic Retinopathy: *Proceedings of the International Joint Conference on Science and Engineering (IJCSE 2020)*. International Joint Conference on Science and Engineering (IJCSE 2020), Surabaya, East Java, Indonesia. <https://doi.org/10.2991/aer.k.201124.069>
- Alizadehsani, R., Roshanzamir, M., Abdar, M., Beykikhoshk, A., Khosravi, A., Panahiazar, M., Koohestani, A., Khozeimeh, F., Nahavandi, S., & Sarrafzadegan, N. (2019). A database for using machine learning and data mining techniques for coronary artery disease diagnosis. *Scientific Data*, 6(1), 227. <https://doi.org/10.1038/s41597-019-0206-3>
- Almutiri, T., & Saeed, F. (2019). Chi Square and Support Vector Machine with Recursive Feature Elimination for Gene Expression Data Classification. *2019 First International Conference of Intelligent Computing and Engineering (ICOICE)*, 1–6. <https://doi.org/10.1109/ICOICE48418.2019.9035165>
- Amalia, N. A., Utami, I. T., & Wilandari, Y. (2024). ANALISIS SENTIMEN KEBIJAKAN PENYELENGGARA SISTEM ELEKTRONIK LINGKUP PRIVAT MENGGUNAKAN PENALIZED LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE. *Jurnal Gaussian*, 12(4), 560–569. <https://doi.org/10.14710/j.gauss.12.4.560-569>
- Arif, S. N. N., Siregar, A. M., Faisal, S., & Juwita, A. R. (2024). Klasifikasi Penyakit Serangan Jantung Menggunakan Metode Machine Learning K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1617. <https://doi.org/10.30865/mib.v8i3.7844>
- Bidul, W. J., Surono, S., & Kurniawan, T. B. (2024). *Comparative Evaluation of Feature Selection Methods for Heart Disease Classification with Support Vector Machine*. 10(2).
- Carolina Wibowo, A., Ardi Lestari, S., & Nurchim, N. (2024). ANALISIS PENGGUNAAN MACHINE LEARNING DALAM KLASIFIKASI PENENTUAN PENYAKIT JANTUNG. *Simtek : jurnal sistem informasi*

dan teknik komputer, 9(2), 97–101.
<https://doi.org/10.51876/simtek.v9i2.395>

- Desiani, A., Zayanti, D. A., Ramayanti, I., Ramadhan, F. F., & Giovillando. (2025). PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN LOGISTIC REGRESSION DALAM KLASIFIKASI KANKER PAYUDARA. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 4(1), 33–42. <https://doi.org/10.69916/jkbt.v4i1.191>
- Fahrudin, G. F., Suroso, S., & Soim, S. (2024). Pengembangan Model Support Vector Machine untuk Meningkatkan Akurasi Klasifikasi Diagnosis Penyakit Jantung. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 7(3), 1418–1428. <https://doi.org/10.32493/jtsi.v7i3.42254>
- Farida, L. N., & Bahri, S. (2024). Klasifikasi Gagal Jantung menggunakan Metode SVM (Support Vector Machine). *Komputika : Jurnal Sistem Komputer*, 13(2), 149–156. <https://doi.org/10.34010/komputika.v13i2.11330>
- Franke, T. M., Ho, T., & Christie, C. A. (2012). The Chi-Square Test: Often Used and More Often Misinterpreted. *American Journal of Evaluation*, 33(3), 448–458. <https://doi.org/10.1177/1098214011426594>
- Kocaoğlu, A. (2024). Efficient Optimization of a Support Vector Regression Model with Natural Logarithm of the Hyperbolic Cosine Loss Function for Broader Noise Distribution. *Applied Sciences*, 14(9), 3641. <https://doi.org/10.3390/app14093641>
- Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., & Wei, H. (2017). Selecting Feature Subsets Based on SVM-RFE and the Overlapping Ratio with Applications in Bioinformatics. *Molecules*, 23(1), 52. <https://doi.org/10.3390/molecules23010052>
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 143–149. <https://doi.org/10.11613/BM.2013.018>
- Natsir, F. M. (2024). *Analisis Deteksi Dini Penyakit Jantung dengan Pendekatan Support Vector Machine pada Data Pasien*.
- Ortega-Martorell, S., Riley, P., Olier, I., Raidou, R. G., Casana-Eslava, R., Rea, M., Shen, L., Lisboa, P. J. G., & Palmieri, C. (2022). Breast cancer patient characterisation and visualisation using deep learning and fisher information networks. *Scientific Reports*, 12(1), 14004. <https://doi.org/10.1038/s41598-022-17894-6>
- Pane, J. P., & Simorangkir, L. (2022). *Volume 4 Nomor 4, November 2022 e-ISSN 2715-6885; p-ISSN 2714-9757*
<http://jurnal.globalhealthsciencegroup.com/index.php/JPPP>. 4(4).

- Platt, J. C. (n.d.). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*.
- Rachmawati, C., Martini, S., & Artanti, K. D. (n.d.). *Analisis Faktor Risiko Modifikasi Penyakit Jantung Koroner Di RSUD Haji Surabaya Tahun 2019*.
- Santoso, I. B., Adrianto, Y., Sensusiaty, A. D., Wulandari, D. P., & Purnama, I. K. E. (2022). Ensemble Convolutional Neural Networks With Support Vector Machine for Epilepsy Classification Based on Multi-Sequence of Magnetic Resonance Images. *IEEE Access*, 10, 32034–32048. <https://doi.org/10.1109/ACCESS.2022.3159923>
- Sarra, R. R., Gorial, I. I., Manea, R. R., Korial, A. E., Mohammed, M., & Ahmed, Y. (n.d.). *Enhanced Stacked Ensemble-Based Heart Disease Prediction with Chi-Square Feature Selection Method*.
- Siddhartha, M. (2020). *Heart Disease Dataset (Comprehensive)* [Dataset]. IEEE DataPort. <https://doi.org/10.21227/DZ4T-CM36>
- Sinha, N., Jangid, T., Joshi, A. M., & Mohanty, S. P. (2022). *iCardo: A Machine Learning Based Smart Healthcare Framework for Cardiovascular Disease Prediction* (arXiv:2212.08022). arXiv. <https://doi.org/10.48550/arXiv.2212.08022>
- Sirait, D. T. C., & Astuti, W. (n.d.). *Analisis Perbandingan Reduksi Dimensi Principal Component Analysis (PCA) dan Partial Least Square (PLS) untuk Deteksi Kanker menggunakan Data Microarray*.
- Sundari, H., Amrustian, M. A., & Wicaksono, A. D. P. (2024). Penerapan Recursive Feature Elimination pada Support Vector Machine untuk Klasifikasi Kanker Payudara. *OPEN ACCESS*, 3(2).
- Tino, M. D. F., Herliyani Hasanah, & Tri Djoko Santosa. (2023). PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINES (SVM) DAN NEURAL NETWORK UNTUK KLASIFIKASI PENYAKIT JANTUNG. *INFOTECH journal*, 9(1), 232–235. <https://doi.org/10.31949/infotech.v9i1.5432>

LAMPIRAN

Lampiran I Source Code *Support Vector Machine* dengan SMO

```
def linear_kernel(x1, x2):
    return np.dot(x1, x2.T)

def polynomial_kernel(x1, x2, degree=3):
    return (1 + np.dot(x1, x2.T)) ** degree

def rbf_kernel(x1, x2, gamma=0.1):
    if x1.ndim == 1:
        x1 = x1[np.newaxis, :]
    if x2.ndim == 1:
        x2 = x2[np.newaxis, :]
    return np.exp(-gamma * np.sum((x1[:, np.newaxis] - x2)**2,
axis=2))

def smo_train(X, y, C=1.0, tol=1e-3, max_passes=5,
kernel_func=linear_kernel):
    m = X.shape[0]
    alpha = np.zeros(m)
    b = 0
    passes = 0
    K = kernel_func(X, X)

    while passes < max_passes:
        num_changed_alphas = 0
        for i in range(m):
            Ei = np.dot((alpha * y), K[:, i]) + b - y[i]
            if (y[i]*Ei < -tol and alpha[i] < C) or (y[i]*Ei >
tol and alpha[i] > 0):
                j = np.random.choice([x for x in range(m) if x
!= i])
                Ej = np.dot((alpha * y), K[:, j]) + b - y[j]

                alpha_i_old, alpha_j_old = alpha[i], alpha[j]

                if y[i] != y[j]:
                    L = max(0, alpha[j] - alpha[i])
                    H = min(C, C + alpha[j] - alpha[i])
                else:
                    L = max(0, alpha[i] + alpha[j] - C)
                    H = min(C, alpha[i] + alpha[j])
                if L == H:
                    continue

                eta = 2 * K[i, j] - K[i, i] - K[j, j]
                if eta >= 0:
                    continue

                alpha[j] -= y[j] * (Ei - Ej) / eta
```

```

        alpha[j] = np.clip(alpha[j], L, H)

        if abs(alpha[j] - alpha_j_old) < 1e-5:
            continue

        alpha[i] += y[i]*y[j] * (alpha_j_old -
alpha[j])

        b1 = b - Ei - y[i]*(alpha[i]-alpha_i_old)*K[i,
i] - y[j]*(alpha[j]-alpha_j_old)*K[i, j]
        b2 = b - Ej - y[i]*(alpha[i]-alpha_i_old)*K[i,
j] - y[j]*(alpha[j]-alpha_j_old)*K[j, j]

        if 0 < alpha[i] < C:
            b = b1
        elif 0 < alpha[j] < C:
            b = b2
        else:
            b = (b1 + b2) / 2

        num_changed_alphas += 1

    passes = passes + 1 if num_changed_alphas == 0 else 0

    return alpha, b

def smo_predict(X_train, y_train, alpha, b, X_test,
kernel_func=linear_kernel):
    sv = alpha > 1e-5
    alpha_sv = alpha[sv]
    y_sv = y_train[sv]
    X_sv = X_train[sv]

    K = kernel_func(X_test, X_sv)
    return np.sign(np.dot(K, alpha_sv * y_sv) + b)

```

Lampiran II *Graphical user interface*

Navigasi

Pilih Halaman:

Klasifikasi - All Features

Informasi Model

Sistem klasifikasi menggunakan semua fitur dengan Support Vector Machine (SVM).

Petunjuk Penggunaan

1. Pilih halaman klasifikasi yang diinginkan

2. Isi semua informasi yang diperlukan

3. Klik tombol 'Prediksi Risiko'

4. Lihat hasil penilaian risiko kardiovaskular Anda

Disclaimer

Sistem ini hanya untuk tujuan penelitian. Selalu konsultasikan dengan tenaga medis profesional untuk diagnosis dan pengobatan.

Cardiovascular Disease Classifier (All Features)

Informasi Pasien (Semua Fitur)

Data Pribadi

Usia (tahun)

Jenis Kelamin

50

Laki-laki

Gejala & Kondisi Fisik

Chest Pain Type

Maximum Heart Rate

Typical Angina

150

Exercise Induced Angina

ST Depression (Oldpeak)

Ya

1.00

Hasil Pemeriksaan Medis

Resting Blood Pressure (mmHg)

Resting ECG Results

120

Normal

Penilaian Risiko

Prediksi Risiko Kardiovaskular

RISIKO RENDAH
Tidak Terdeteksi Penyakit Jantung

Model menunjukkan risiko rendah penyakit kardiovaskular. Terus jaga pola hidup sehat!

Statistik Kinerja Model

Accuracy

Precision

Recall

F1-Score

85%

86%

88%

87%

Navigasi

Pilih Halaman:

Klasifikasi - Selected Features

Informasi Model

Petunjuk Penggunaan

Sistem klasifikasi menggunakan fitur terpilih dengan Support Vector Machine (SVM).

- Pilih halaman klasifikasi yang diinginkan
- Isi semua informasi yang diperlukan
- Klik tombol 'Prediksi Risiko'
- Lihat hasil penilaian risiko kardiovaskular Anda

Disclaimer

Sistem ini hanya untuk tujuan penelitian. Selalu konsultasikan dengan tenaga medis profesional untuk diagnosis dan pengobatan.

Cardiovascular Disease Classifier (Selected Features)

Data Pribadi

Jenis Kelamin

Perempuan

Gejala & Kondisi Fisik

Chest Pain Type	Maximum Heart Rate
Atypical Angina	143
Exercise Induced Angina	ST Depression (Oldpeak)
Tidak	2.00

Hasil Pemeriksaan Medis

Fasting Blood Sugar > 120 mg/dl	ST Slope
---------------------------------	----------

Penilaian Risiko

Prediksi Risiko Kardiovaskular

RISIKO TINGGI
Terdeteksi Penyakit Jantung

Model menunjukkan risiko tinggi penyakit kardiovaskular. Silakan konsultasi dengan dokter spesialis Jantung segera.

Statistik Kinerja Model

Accuracy	Precision	Recall	F1-Score
85%	86%	87%	86%