

SENASTIKA Universitas Malikussaleh

DIAGNOSIS PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

David Fadlianda¹, Aditya Prananto², Cut Anggel Eriska³, Syifa Anjanira⁴, Nada Syadzwin⁵, Munirul Ula⁶

^{1,2,3,4,5,6} Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh
Email: ¹david.210170272@mhs.unimal.ac.id, ²aditya.210170282@mhs.unimal.ac.id,
³cut.210170277@mhs.unimal.ac.id, ⁴syifa.210170270@mhs.unimal.ac.id,
⁵nada.210170276@mhs.unimal.ac.id, ⁶munirulula@unimal.ac.id

Abstrak

Penyakit jantung merupakan salah satu penyebab utama kematian global, dengan angka kejadian yang terus meningkat akibat perubahan gaya hidup. Permasalahan utama dalam diagnosa penyakit jantung terletak pada kompleksitas dan keragaman data medis yang tersedia dan sering kali terdiri dari berbagai variabel seperti tekanan darah, kadar kolesterol, riwayat kesehatan keluarga, dan faktor-faktor lainnya. Oleh karena itu, metode pembelajaran mesin seperti *SVM* sangat diperlukan untuk membantu mengolah data dengan efisien dan meningkatkan akurasi prediksi. Metode yang digunakan dalam penelitian ini meliputi beberapa tahapan, yaitu pengumpulan data medis, pra-pemrosesan data untuk mengatasi ketidakseimbangan dan kebisingan data, penerapan algoritma *SVM* untuk membangun model prediksi, serta evaluasi hasil dengan menggunakan dataset yang telah disediakan. Dataset yang digunakan mencakup berbagai parameter medis yang relevan. Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi yang akurat menggunakan metode pembelajaran mesin, khususnya algoritma Support Vector Machine (SVM), dalam mendeteksi risiko penyakit jantung. Penelitian ini bertujuan untuk mengatasi permasalahan kompleksitas dan keragaman data medis. Evaluasi dilakukan dengan menggunakan metrik-metrik seperti akurasi, precision, recall, dan F1-Score untuk menilai kinerja model dalam memprediksi risiko penyakit jantung. Hasil penelitian menunjukkan bahwa model *SVM* mampu mencapai akurasi sebesar 86,89%, dengan nilai precision, recall, dan F1-Score masing-masing sebesar 0,88. Ini menunjukkan bahwa model *SVM* memiliki kemampuan yang baik dalam mengklasifikasikan pasien dengan risiko tinggi maupun rendah terhadap penyakit jantung. Selain itu, model ini menunjukkan kinerja yang stabil dan dapat diandalkan, baik untuk pasien yang sudah terdiagnosis maupun yang belum menunjukkan gejala yang signifikan. Dengan demikian, penggunaan algoritma *SVM* dalam diagnosis penyakit jantung terbukti efektif, menawarkan solusi yang signifikan dalam membantu dokter dan tenaga medis untuk melakukan deteksi dini serta intervensi yang lebih tepat waktu.

Kata kunci: *Algoritma, Medis, Machine Learning, Penyakit Jantung, Support Vector Machine.*

Abstract

Cardiovascular disease is one of the leading causes of death globally, with the incidence rate continuing to increase due to lifestyle changes. The main problem in diagnosing heart disease lies in the complexity and diversity of available medical data, which often consists of various variables such as blood pressure, cholesterol levels, family health history, and other factors. Therefore, machine learning methods such as *SVM* are needed to help process data efficiently and improve prediction accuracy. The method used in this study includes several stages, namely medical data collection, data pre-processing to overcome data imbalance and noise, application of the *SVM* algorithm to build a prediction model, and evaluation of the results using the provided dataset. The dataset used includes various relevant medical parameters. The purpose of this study is to develop an accurate prediction model using machine learning methods, especially the Support Vector Machine (*SVM*) algorithm, in detecting the risk of heart disease. This study aims to overcome the problems of complexity and diversity of medical data. The evaluation was carried out using metrics such as accuracy, precision, recall, and F1-Score to assess the performance of the model in predicting the risk of heart disease. The results showed that the *SVM* model was able to achieve an accuracy of 86.89%, with precision, recall, and F1-Score values of 0.88 each. This shows that the *SVM* model has good ability in classifying patients with high and low risk of heart disease. In addition, this model shows stable and reliable performance, both for patients who have been diagnosed and those who have not shown significant symptoms. Thus, the use of the *SVM* algorithm in the diagnosis of heart disease has proven effective, offering a significant solution in helping doctors and medical personnel to carry out early detection and more timely intervention.

Keywords: *Algorithm, Medical, Machine Learning, Cardiovascular disease, Support Vector Machine.*

1. PENDAHULUAN

Penyakit jantung adalah penyebab utama kematian di seluruh dunia, dan kejadiannya terus meningkat seiring dengan perubahan gaya hidup dan kebiasaan makan masyarakat. Jantung adalah salah satu organ terpenting manusia, dan perannya adalah memompa darah ke seluruh bagian tubuh. Hal tersebut membuat jantung memiliki peranan yang sangat penting untuk tubuh manusia, oleh sebab itu penderita penyakit jantung akan merasakan ketidaknyamanan pada tubuh mereka. Kemudian penyakit jantung juga dapat menyebabkan penghambatan dari fungsi jantungnya seperti kardiovaskular, jantung koroner, bahkan yang paling parahnya dapat menyebabkan gagal jantung [1]. Penyakit jantung adalah penyakit mematikan yang menyerang hampir 12 juta orang di seluruh dunia. Oleh karena itu pemeriksaan sedari dini merupakan upaya yang sangat penting untuk dilakukan, dan diagnosa penyakit jantung menjadi salah satu hal yang sangat beresiko dikarenakan memiliki ikatan yang saling berkegantungan dari berbagai faktor [2]. Meskipun kemajuan dalam teknologi medis memungkinkan pengumpulan data yang lebih banyak dan beragam, masih terdapat kesulitan dalam mengolah dan menganalisis data tersebut untuk mengidentifikasi individu yang berisiko tinggi. Masalah utama yang dihadapi adalah bagaimana memanfaatkan data medis yang tersedia seperti riwayat kesehatan, faktor risiko, dan hasil pemeriksaan untuk memberikan gambaran yang jelas tentang risiko penyakit jantung secara keseluruhan.

Dalam konteks ini, algoritma *machine learning*, dengan metode *Support Vector Machine (SVM)*, menawarkan solusi yang menjanjikan. *SVM* dapat menganalisis data berukuran besar dan kompleks, membantu dalam mengidentifikasi pola yang mungkin tidak terlihat melalui metode tradisional. Dengan kemampuannya untuk menemukan batas optimal antara kelas-kelas data, *SVM* dapat digunakan untuk memprediksi risiko penyakit jantung secara lebih akurat.

Metode ini menggunakan prinsip dasar pengklasifikasi linier, yaitu kasus pengklasifikasi yang dapat diklasifikasikan secara linier. *SVM* dikembangkan untuk menangani masalah nonlinier dengan menerapkan teori kernel pada ruang kerja yang besar. Hyperplanes muncul pada ruang dengan dimensi besar, yang dapat memaksimalkan selisih (margin) antar dataset [4].

Penelitian sebelumnya yang dikembangkan oleh Herliani Hasanah dan Nurmaalityari, “Perbandingan tingkat akurasi support vector machine (SVM) dan algoritma C45 dalam memprediksi penyakit jantung”, menunjukkan bahwa algoritma SVM merupakan algoritma yang akurat dan akurat digunakan untuk memprediksi penyakit jantung hasilnya menunjukkan hal itu skor akurasi 87%. Sedangkan untuk algoritma C45 memiliki nilai akurasi sebesar 82% dalam menyelesaikan kasus klasifikasi penyakit jantung. Secara keseluruhan Algoritma SVM mempunyai nilai keakuratan yang lebih tinggi jika dibandingkan dengan algoritma C45 [5]. Penelitian terdahulu lainnya dengan menggunakan subjek penelitian yang berbeda dilakukan pada jurnal penelitian berjudul “Machine Learning Breast Cancer Diagnosis Menggunakan Algoritma Support Vector Machine” oleh Chalifa Chazar dan Bagus Erawan Widhiaputra dan SVM yang digunakan dalam penelitian ini Pembentukan sistem aplikasi ML menggunakan algoritma adalah dibahas. Diagnosis kanker payudara. Dataset yang digunakan adalah dataset Wisconsin Breast Cancer (Diagnosis). Dan hasil penelitian ini menunjukkan bahwa aplikasi ML ini dapat mendiagnosis kanker payudara dan mengambil keputusan prediktif berdasarkan dua kemungkinan: sel hidup dalam keadaan ganas atau jinak. Aplikasi ini juga dapat digunakan oleh petugas laboratorium untuk membuat diagnosis berdasarkan hasil FNA dari studi biopsi [6].

Dengan itu, tujuan pada penelitian ini untuk mengimplementasikan algoritma SVM guna memprediksi risiko penyakit jantung berdasarkan dataset yang telah disediakan. Penelitian ini bertujuan tidak hanya untuk menghasilkan model prediktif yang akurat, tetapi juga untuk memberikan wawasan mengenai faktor-faktor yang signifikan dalam pengembangan penyakit jantung. Penelitian ini dilaksanakan menggunakan *Google Colabs*, yang memungkinkan peneliti untuk mengakses lingkungan pemrograman yang interaktif dan efisien dalam mengolah data serta menerapkan algoritma. Melalui langkah-langkah seperti pengumpulan dan pra-pemrosesan data, penerapan algoritma SVM, serta evaluasi hasil, kami berharap dapat menyajikan analisis yang mendalam tentang variabel-variabel yang berpengaruh.

Secara Signifikansi penelitian ini terletak pada potensi kontribusinya terhadap bidang kesehatan, khususnya dalam meningkatkan akurasi diagnosis penyakit jantung dan membantu profesional medis untuk pengambilan keputusan yang lebih efektif dan simple. Dengan menyediakan model prediktif berbasis data, penelitian ini diharapkan dapat mendukung upaya pencegahan penyakit jantung dan memajukan kualitas layanan kesehatan di masyarakat. Implementasi algoritma SVM pada penelitian ini juga akan menunjukkan bagaimana teknologi machine learning dapat diintegrasikan dalam praktik medis untuk mencapai hasil yang lebih baik.

2. TINJAUAN PUSTAKA

Tinjauan ini berperan penting dalam memposisikan penelitian secara tepat, memastikan penggunaan metodologi yang valid, dan memperkuat argumen penggunaan machine learning metode SVM untuk diagnosis penyakit jantung. Analisis literatur akan mengidentifikasi celah pengetahuan, mendorong inovasi, dan berkontribusi pada peningkatan akurasi dalam deteksi dan klasifikasi penyakit jantung menggunakan dataset pasien.

2.1 Machine Learning

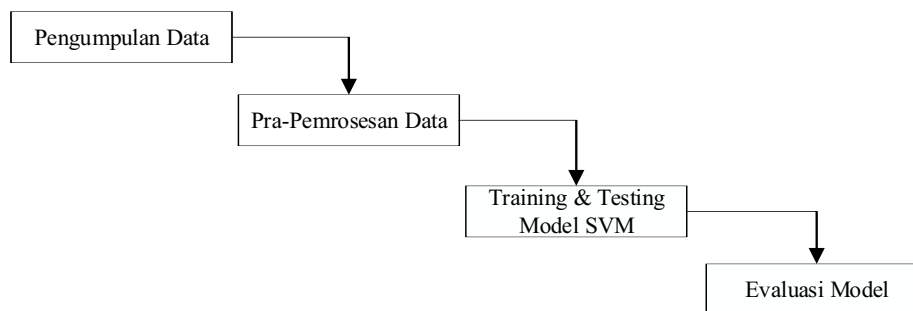
Machine learning dapat disebut dengan sebuah cabang kecerdasan buatan yang fokus terhadap pengembangan sistem yang dapat memahami dan berkembang secara mandiri. Inti dari teknologi ini adalah kemampuannya untuk mengidentifikasi pola kompleks dalam kumpulan data besar, tanpa perlu diprogram secara eksplisit untuk setiap skenario [7]. Dengan memanfaatkan algoritma canggih dan model statistik, sistem machine learning ini dapat membuat keputusan atau prediksi berdasarkan pengalaman sebelumnya, mirip dengan proses pembelajaran manusia. Seiring berjalannya waktu dan semakin banyaknya data yang diproses, kinerja sistem ini terus meningkat, memungkinkan aplikasinya dalam berbagai bidang termasuk dalam diagnosis dataset pasien penyakit jantung [8].

2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan sebuah algoritma supervised machine learning yang digunakan dalam sebuah tugas klasifikasi, yang mengklasifikasikan data menjadi dua kelas: +1 dan -1. SVM bekerja dengan mencari hyperplane optimal yang memaksimalkan jarak antara dua kelas dan memberikan pemisahan paling akurat. Algoritma ini dapat mengatasi nonlinier dengan menggunakan trik kernel sehingga sangat efektif terutama dalam menangani dataset yang kompleks dan besar dengan banyak fitur, menjadikannya pilihan populer dalam berbagai aplikasi klasifikasi [9].

3. METODE PENELITIAN

Melakukan penelitian yang berbasis data dan informasi objektif sangat penting sebagai landasan dalam penelitian tersebut. Dengan adanya data yang akurat, diharapkan hasil penelitian akan berkualitas tinggi [10]. Berikut adalah proses penelitian yang dapat digunakan dalam merangkum informasi tentang penelitian ini:



Gambar 1. Metode Penelitian

3.1 Pengumpulan Data

Dalam pengembangan sistem prediksi penyakit jantung, penelitian ini dilaksanakan melalui beberapa langkah. Pertama, data diperoleh dari sumber yang kredibel yang berupa file *Comma Separated Values (CSV)*. Dataset ini didapatkan melalui sebuah *website kaggle* [11] <https://www.kaggle.com/code/rafiromolo/prediksi-penyakit-jantung>. Dari hasil dataset ini didapatkan total hasil sebanyak 303 pasien. Setelah itu, data tersebut diproses untuk menjamin kualitas dan konsistensinya. Implementasi sistem ini dilakukan menggunakan bahasa pemrograman *Python* dengan sebuah platform *Google Collab* yang berbasis cloud untuk menulis tiap kode yang dibutuhkan pada algoritma *Support Vector Machine (SVM)* yang digunakan.

3.2 Pra-Pemrosesan Data

Pra-processing merupakan tahap yang sangat penting sebelum data bisa digunakan [12] untuk melatih algoritma *Support Vector Machine (SVM)*. Dengan metode ini dimana terdapat proses dalam memisahkan fitur dan target, membagi dataset menjadi training dan testing dan scaling data. Tujuannya adalah memastikan bahwa data dalam format yang benar, bersih, dan siap untuk digunakan oleh model.

3.3 Training & Testing Model

Proses Training Model, mempelajari data yang telah diproses menggunakan data latihan. Sementara itu, Testing Model adalah proses di mana model membuat prediksi berdasarkan data uji. Kedua proses ini sangat penting karena melalui pemrosesan dataset secara otomatis, *SVM* dapat menghasilkan model yang paling optimal [13].

3.4 Evaluasi Model

Pengujian model diperlukan untuk mengevaluasi kinerja metode support vector machine (SVM). Dalam studi ini, evaluasi didasarkan pada akurasi, yaitu metrik yang mengukur seberapa akurat model memprediksi hasil yang benar berdasarkan data.[14]. Akurasi ini akan menunjukkan tingkat keberhasilan model dalam memprediksi label

yang benar. Dalam pengujian metode *SVM* ini, yang juga akan memberikan persentase terhadap performa *confusion matrix* dari setiap rumus yang sudah ditentukan melalui metode *SVM* ini. Persamaan rumus tersebut dapat dilihat seperti berikut:

1. Akurasi merupakan suatu nilai rasio information tweet yang nyata saat dideteksi di dalam information pengujian. Dengan kata lainnya akurasi ini merupakan nilai yang secara langsung menunjukkan tingkat kedekatan antara suatu nilai prediksi sistem dengan suatu nilai prediksi manusia [15]. Nilai akurasi dapat dicari dengan persamaan sebagai berikut :

$$Accuracy = \left(\frac{TP + TN}{TP + FP + TN + FN} \right) \times 100 \% \quad (1)$$

2. Precision mengukur tingkat akurasi antara data yang diinginkan dengan hasil prediksi yang dihasilkan oleh model. Precision adalah rasio dari prediksi positif yang benar dibandingkan dengan seluruh prediksi yang dikategorikan sebagai positif. Konsep ini menjelaskan seberapa tepat hasil prediksi model terhadap data yang diinginkan, menunjukkan sejauh mana prediksi positif yang dibuat oleh model sesuai dengan kenyataan [16]. Nilai precision dapat dicari menggunakan persamaan dibawah ini :

$$Precision = \left(\frac{TP}{TP + FP} \right) \times 100 \% \quad (2)$$

3. Recall membantu kemampuan model untuk mengidentifikasi ulang informasi yang relevan. Recall atau sensitivitas dipergunakan untuk mendapat hasil perbandingan jumlah suatu prediksi positif yang akurat dibandingkan dengan jumlah semua kelas yang bernilai positif [17]. Adapun nilai recall dapat dicari menggunakan persamaan dibawah ini :

$$Recall = \left(\frac{TP}{TP + FN} \right) \times 100 \% \quad (3)$$

4. F1-Score merepresentasikan rata-rata harmonik dari precision dan recall yang telah diberi bobot. Dengan menggabungkan nilai precision dan recall, F1-Score menghasilkan satu metrik tunggal yang mencerminkan keseimbangan antara keduanya, memberikan gambaran yang lebih komprehensif mengenai kinerja model dalam prediksi positif. [18]. Nilai F1-Score dapat dicari menggunakan persamaan berikut :

$$F1 - Score = \left(\frac{2 \times precision \times recall}{precision + recall} \right) \times 100 \% \quad (4)$$

Untuk mengukur akurasi, presisi, recall dan F1-Score dapat menggunakan *confusion matrix*. *Confusion matrix* merupakan alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah dari nilai presisi klasifikasi terhadap kelas menggunakan algoritma yang digunakan [19]. Berikut telah dipresentasikan bentuk *confusion matrix* dari tabel dibawah ini:

Tabel 1. Bentuk Confusion matrix dari dua kelas

Confusion Matrix		Nilai Sebenarnya	
		TRUE	FALSE
Nilai prediksi	TRUE	TP (True Positive) Correct result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

4. ANALISIS DAN PEMBAHASAN

4.1 Analisis Atribut

Dari dataset ini yang akan digunakan untuk penelitian ini, merupakan sumber dari Kaggle, terdapat total 304 pasien dengan 14 atribut. Dari 14 atribut tersebut, 13 atribut digunakan untuk menggambarkan jenis penyakit yang diderita oleh setiap pasien, sementara 1 atribut berfungsi sebagai “target” untuk menunjukkan status kesehatan pasien terkait penyakit jantung. Atribut target ini memiliki 2 nilai: nilai 1 menandakan bahwa pasien memiliki penyakit jantung, sedangkan nilai 0 menandakan bahwa pasien tidak memiliki penyakit jantung. Data atribut yang lebih rinci dapat dijelaskan pada tabel berikut:

Tabel 2. Atribut Prediksi

No	ATRIBUT	Tipe Data	Keterangan
1	Age	int64	Usia
2	Sex	int64	Jenis Kelamin

3	Cp	int64	Sakit Dada
4	Trestbps	int64	Tekanan Darah
5	Chol	int64	Kolestrol
6	Fbs	int64	Gula Darah
7	Restecg	int64	Kondisi Jantung
8	Thalach	int64	Denyut Jantung
9	Exang	int64	Exercise Induced Angina
10	Oldpeak	float64	Uji Stress Jantung
11	Slope	int64	Uji Kemiringan Jantung
12	Ca	int64	Pembuluh darah koroner
13	Thal	int64	Deteksi Fungsi Jantung
14	Target	int64	Deteksi Penyakit Jantung

4.2 Preview Dataset

Untuk memberikan gambaran awal mengenai dataset yang digunakan dalam penelitian ini, penting untuk menampilkan setidaknya 5 data teratas agar proses *pre-processing* data dapat dilakukan dengan lebih mudah sebelum melakukan pemodelan dan visualisasi pemisahan. Tabel berikut menyajikan data teratas dari dataset, yang memberikan insight awal mengenai karakteristik data yang akan digunakan dalam analisis lebih lanjut. Dengan meninjau data ini, peneliti dapat memahami struktur dataset dan mengidentifikasi langkah-langkah yang diperlukan dalam *pre-processing* sebelum melanjutkan ke tahap pemodelan.

Tabel 3. Hasil Processing

Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal	Target
63.0	1.0	3.0	145.0	233.0	1.0	0.0	150.0	0.0	2.3	0.0	0.0	1.0	1.0
37.0	1.0	2.0	130.0	250.0	0.0	1.0	187.0	0.0	3.5	0.0	0.0	2.0	1.0
41.0	0.0	1.0	130.0	204.0	0.0	0.0	172.0	0.0	1.4	2.0	0.0	2.0	1.0
56.0	1.0	1.0	120.0	236.0	0.0	1.0	178.0	0.0	0.8	2.0	0.0	2.0	1.0
57.0	0.0	0.0	120.0	354.0	0.0	1.0	163.0	1.0	0.6	2.0	0.0	2.0	1.0

4.3 Preview Model SVM

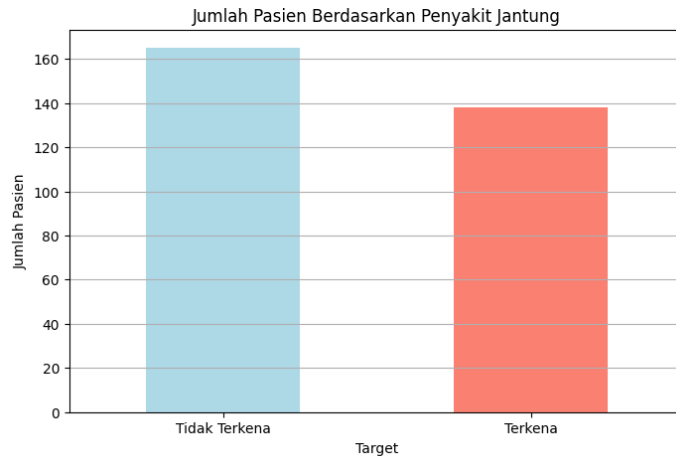
Langkah berikutnya adalah memisahkan dataset menjadi fitur (X) dan target (y), sehingga model dapat fokus pada variabel input untuk memprediksi hasil. Dengan visualisasi data menggunakan matplotlib, pola-pola penting yang mempengaruhi performa model dapat diidentifikasi. Setelah itu, model SVM dilatih menggunakan data tersebut, di mana pemahaman yang baik terhadap distribusi data sangat penting untuk memilih kernel yang tepat, dalam hal ini kernel linier. Hasil keseluruhan dari klasifikasi memperlihatkan perbedaan yang jelas antara pasien yang terkena dan tidak terkena penyakit jantung, yang akan divisualisasikan dalam bentuk tabel dan gambar berikut untuk memudahkan analisis.

Tabel 4. Fitur (X)

Age	Sex	Cp	Trestbps	Chol	Fbs	Restecg	Thalach	Exang	Oldpeak	Slope	Ca	Thal
63.0	1.0	3.0	145.0	233.0	1.0	0.0	150.0	0.0	2.3	0.0	0.0	1.0
37.0	1.0	2.0	130.0	250.0	0.0	1.0	187.0	0.0	3.5	0.0	0.0	2.0
41.0	0.0	1.0	130.0	204.0	0.0	0.0	172.0	0.0	1.4	2.0	0.0	2.0
56.0	1.0	1.0	120.0	236.0	0.0	1.0	178.0	0.0	0.8	2.0	0.0	2.0
57.0	0.0	0.0	120.0	354.0	0.0	1.0	163.0	1.0	0.6	2.0	0.0	2.0

Tabel 5. Target (Y)

Target (y)
1
1
1
1
1



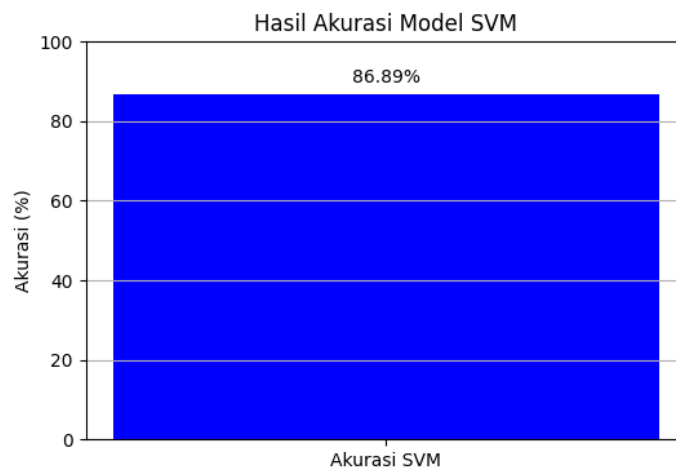
Gambar 2. Barplot Jumlah Data Pasien

Grafik di atas menampilkan jumlah pasien berdasarkan status penyakit jantung, yang dibagi menjadi dua kategori: "Tidak Terkena" dan "Terkena." Dari grafik ini, dapat diketahui bahwa jumlah pasien yang tidak terkena penyakit jantung (sekitar 160) dan lebih besar dibandingkan dengan pasien yang terkena penyakit jantung (sekitar 140).

Dalam konteks model SVM untuk diagnosis penyakit jantung, ketidakseimbangan ini dapat membuat model lebih akurat dalam memprediksi pasien yang tidak terkena, namun kurang sensitif terhadap pasien yang sebenarnya terkena penyakit. Untuk mengatasi ini, teknik pada SVM dapat diterapkan dalam evaluasi seperti precision, recall, dan F1-score yang penting digunakan untuk memastikan model tidak bias dan mampu mendeteksi risiko dengan baik, terutama pada pasien yang terkena penyakit jantung.

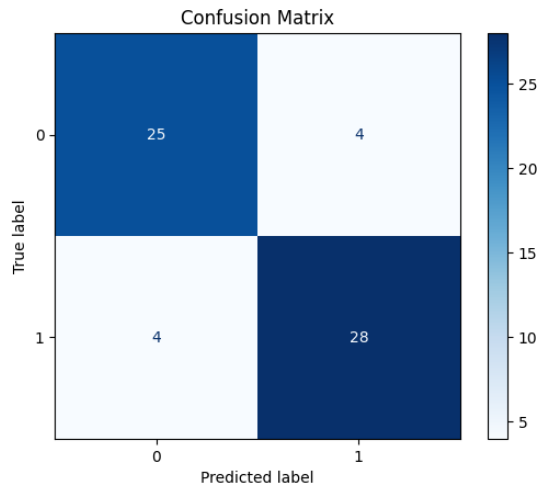
4.4 Hasil Pemodelan Data

Model SVM dilatih menggunakan kernel linier, dan akurasi model diukur serta divisualisasikan dalam bentuk diagram batang dengan nilai akurasi mencapai 86,89%. Visualisasi ini membantu memberikan gambaran yang lebih jelas mengenai performa model dalam mengklasifikasikan data secara akurat. Hasil akurasi model ditampilkan melalui gambar grafik seperti berikut, yang memperlihatkan perbandingan akurasi dengan metrik lainnya untuk evaluasi yang lebih komprehensif.



Gambar 3. Tingkat Akurasi Data

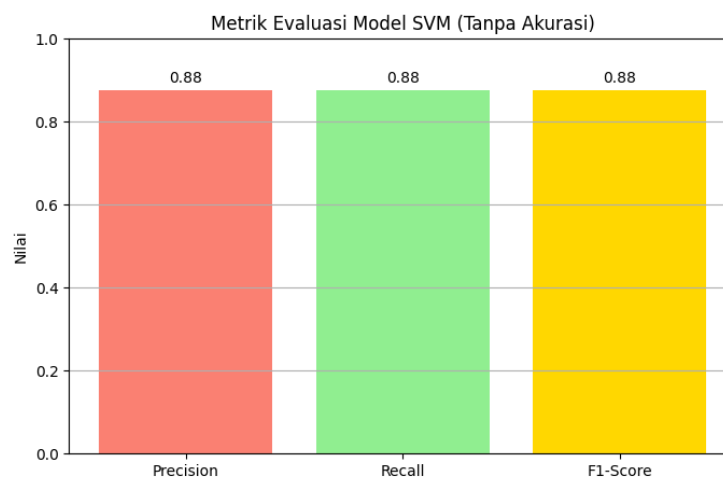
Sementara itu, dengan memanfaatkan confusion matrix, kita dapat menilai kinerja model klasifikasi, khususnya pada masalah klasifikasi biner atau multi-kelas. Confusion matrix menampilkan hasil prediksi model dengan membandingkan nilai aktual dengan nilai prediksi yang dihasilkan. Tabel ini mengkategorikan hasil prediksi ke dalam empat kategori: True Positives (TP), False Positives (FP), True Negatives (TN), dan False Negatives (FN). Dari informasi ini, kita dapat menghitung berbagai metrik evaluasi seperti precision, recall, dan F1-score, yang memberikan pemahaman yang jelas mengenai seberapa baik model dalam memprediksi kelas dengan akurat serta meminimalkan kesalahan.

Gambar 4. *Confusion Matrix*

Berikut adalah penjelasan dari setiap elemen pada matrix:

1. True Negatives (TN): Sebanyak 25 sampel kelas 0 diprediksi dengan benar sebagai kelas 0.
2. False Positives (FP): Sebanyak 4 sampel kelas 0 diprediksi secara salah sebagai kelas 1.
3. False Negatives (FN): Sebanyak 4 sampel kelas 1 diprediksi secara salah sebagai kelas 0.
4. True Positives (TP): Sebanyak 28 sampel kelas 1 diprediksi dengan benar sebagai kelas 1.

Dengan demikian, melalui confusion matrix, hasil prediksi model dapat dievaluasi untuk mendapatkan nilai metrik penting seperti precision, recall, dan F1-score. Berdasarkan evaluasi tersebut, model menghasilkan nilai precision sebesar 0,88, recall sebesar 0,88, dan F1-score sebesar 0,88. Nilai-nilai ini menunjukkan bahwa model mampu melakukan prediksi dengan konsistensi yang baik, seperti yang terlihat pada gambar matrix berikut:

Gambar 5. *Metrix Evaluasi Model*

Dalam situasi tertentu, model prediksi dapat menunjukkan keseimbangan yang baik antara hasil positif dan negatif, terutama ketika jumlah True Positives (TP), False Positives (FP), dan False Negatives (FN) hampir sama atau sangat kecil. Dalam kondisi ini, Precision, Recall, dan F1-Score cenderung setara, menunjukkan bahwa model dapat secara efektif mengidentifikasi kasus positif (precision) sekaligus mendeteksi sebagian besar kasus positif yang ada (recall). Hal ini sangat penting dalam aplikasi seperti diagnosa penyakit jantung, di mana keseimbangan antara prediksi yang benar dan kesalahan prediksi harus dijaga untuk meminimalkan risiko.

5. DISKUSI

5.1 Ketidakseimbangan

Distribusi data pada penelitian ini menunjukkan adanya ketidakseimbangan antara jumlah pasien yang tidak terkena penyakit jantung (160 pasien) dan yang terkena penyakit jantung (144 pasien) dari total 304 pasien. Ketidakseimbangan data ini memberikan tantangan tersendiri dalam membangun model prediksi yang akurat menggunakan algoritma Support Vector Machine (SVM). Secara umum, ketidakseimbangan seperti ini dapat

menyebabkan model lebih cenderung memprediksi kategori mayoritas, yaitu pasien yang tidak terkena penyakit jantung, dan mungkin kurang mampu mendeteksi pasien yang benar-benar berisiko. Hal ini dapat terlihat pada metrik seperti precision dan recall, di mana recall untuk pasien yang terkena penyakit jantung mungkin lebih rendah karena ketidakseimbangan ini.

5.2 Kinerja Model SVM

Model SVM yang diterapkan pada dataset ini menunjukkan hasil akurasi yang cukup baik, dengan nilai akurasi sebesar 86,89%. Namun, akurasi saja tidak cukup untuk mengevaluasi kinerja model, terutama dengan adanya ketidakseimbangan data. Precision untuk pasien yang terkena penyakit jantung mencapai 0,88, yang menunjukkan bahwa model cukup baik dalam menghindari prediksi false positive (memastikan bahwa pasien yang diprediksi terkena benar-benar berisiko). Namun, recall juga harus diperhatikan, karena model bisa saja gagal mendeteksi sejumlah pasien yang terkena penyakit jantung (false negative). Nilai recall yang lebih rendah akan menunjukkan bahwa model masih bisa melewatkan pasien yang benar-benar berisiko. Dalam hal ini, metrik F1-Score (kombinasi antara precision dan recall) digunakan untuk memberikan gambaran keseimbangan antara kedua aspek ini, yang dapat membantu menentukan apakah model SVM sudah optimal.

5.3 Perbandingan dengan Metode Lain

Penelitian ini menggunakan algoritma SVM dengan akurasi 86,89%, yang sejalan dengan penelitian oleh Herliani Hasanah dan Nurmaalitayari yang menunjukkan akurasi SVM sebesar 87%, lebih tinggi dibandingkan algoritma C45 yang hanya mencapai 82%. Hal ini menunjukkan bahwa SVM lebih unggul dalam memprediksi penyakit jantung dan pada penelitian oleh Chalifa Chazar dan Bagus Erawan Widhiaputra dalam diagnosis kanker payudara menggunakan SVM juga menunjukkan akurasi tinggi dalam membedakan sel ganas dan jinak. Kedua penelitian ini menegaskan bahwa SVM lebih efektif dibandingkan metode lain seperti C45 dalam menangani data medis yang kompleks, baik untuk penyakit jantung maupun kanker.

6. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan metode Support Vector Machine (SVM) dalam diagnosa penyakit jantung terbukti sangat efektif, dengan tingkat akurasi mencapai 86,89%. Pada penelitian ini, model SVM diuji dengan menggunakan pembagian dataset di mana 80% digunakan sebagai data training dan 20% sebagai data testing. Selain akurasi yang cukup tinggi, model ini juga menunjukkan performa yang konsisten dengan nilai precision, recall, dan F1-Score yang sama, yaitu 0,88. Hal ini menunjukkan bahwa model SVM mampu memprediksi secara akurat baik pasien yang menderita penyakit jantung maupun yang tidak, serta memberikan keseimbangan yang baik dalam mendeteksi kasus positif sekaligus meminimalkan kesalahan prediksi. Dengan demikian, hasil penelitian ini memperkuat efektivitas SVM sebagai alat bantu yang andal dalam mendiagnosa data pasien terkait penyakit jantung, memberikan kontribusi signifikan dalam bidang kesehatan untuk deteksi dini penyakit jantung.

DAFTAR PUSTAKA

- [1] E. Tasia, R. Zaid, I. Zarier, S. Kenia, and P. Loka, "Klasifikasi Penyakit Gagal Jantung Menggunakan Supervised Learning," *Sentimas*, pp. 1–7, 2023.
- [2] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [3] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics and Digital Expert (INDEX)*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.
- [4] W. Latuny, V. O. Lawalata, D. B. Paillin, and R. Ohoirenana, "Prediksi Fitur Kemasan Produk Minyak Kayu Putih Dengan Support Vector Machine (Svm)," *ALE Proceeding*, vol. 4, pp. 76–82, 2021, doi: 10.30598/ale.4.2021.76-82.
- [5] Herliyani, Hasanah and Nurmalitasari, "Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C45 dalam Prediksi Penyakit Jantung," *Prosiding Seminar Nasional Teknologi dan Sains*, vol. 2, pp. 13–18, 2023.
- [6] C. Chazar and B. E. Widhiaputra, "INFORMASI (Jurnal Informatika dan Sistem Informasi) Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," 2020.
- [7] S. Prasetyo and T. Dewayanto, "Penerapan Machine Learning, Deep Learning, Dan Data Mining Dalam Deteksi Kecurangan Laporan Keuangan-a Systematic Literature Review," *Diponegoro Journal of Accounting*, vol. 13, no. 3, pp. 1–12, 2024, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/accounting>
- [8] R. Kurniawan, P. B. Wintoro, Y. Mulyani, and M. Komarudin, "Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 11, no. 2, pp. 233–236, 2023, doi: 10.23960/jitet.v11i2.3034.

- [9] A. C. Nugraha and M. I. Irawan, "Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost)," *Jurnal Sains dan Seni ITS*, vol. 12, no. 1, 2023, doi: 10.12962/j23373520.v12i1.107032.
- [10] Jasmarizal, Junadhi, Rahmaddeni, and M. Khairul Anam, "Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap Produk Skincare," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 1438–1450, 2024, doi: 10.33022/ijcs.v13i1.3654.
- [11] J. Teknik Elektro dan Komputasi *et al.*, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 2, pp. 207–214, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/ELKOM/article/view/7691>
- [12] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Eduatic - Scientific Journal of Informatics Education*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [13] Lukman Priyambodo *et al.*, "Klasifikasi Kematangan Tanaman Hidroponik Pakcoy Menggunakan Metode SVM," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 1, pp. 153–160, 2022, doi: 10.29207/resti.v6i1.3828.
- [14] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
- [15] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021, doi: 10.30865/mib.v5i2.2937.
- [16] H. Jurnal, A. Rusdy Prasetyo, and B. S. Aditya, "JURNAL ILMIAH TEKNIK MESIN, ELEKTRO DAN KOMPUTER ANALISIS PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DAN CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK SISTEM DETEKSI KATARAK," vol. 3, pp. 1–10, 2023.
- [17] S. N. N. Arif, A. M. Siregar, S. Faisal, and A. R. Juwita, "Klasifikasi Penyakit Serangan Jantung Menggunakan Metode Machine Learning K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1617, Jul. 2024, doi: 10.30865/mib.v8i3.7844.
- [18] E. Hasibuan and E. A. Heriyanto, "ANALISIS SENTIMEN PADA ULASAN APLIKASI AMAZON SHOPPING DI GOOGLE PLAY STORE MENGGUNAKAN NAIVE BAYES CLASSIFIER," *JTS*, vol. 1, no. 3, 2022.
- [19] C. Emilia Sukmawati Universitas Buana Perjuangan Karawang Kab Karawang, A. Rizky Pratama Universitas Buana Perjuangan Karawang Kab Karawang, and N. Nurjanah Universitas Buana Perjuangan Karawang Kab Karawang, "Techno Xplore Jurnal Ilmu Komputer dan Teknologi Informasi ALGORITMA NAÏVE BAYES, RANDOM FOREST DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI PENYAKIT SAPI," 2024.