VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

IMPLEMENTASI MODEL SUPPORT VECTOR MACHINE UNTUK DETEKSI DINI RISIKO PENYAKIT JANTUNG BERDASARKAN PARAMETER MEDIS PASIEN

Alya Amelianti[1]; Dian Ade Kurnia[2]; Yudhistira Arie Wijaya[3]; Puji Pramudya Marta[4];

Program Studi Teknik Informatika[1]
Program Studi Manajemen Informatika[24]
Program Studi Sistem Informasi[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
aialya885@gmail.com

(*) Corresponding Author : aialya885@gmail.com
Published : 31 Desember 2025

Abstract— Heart disease remains one of the leading causes of mortality worldwide, highlighting the need for predictive systems that can support early detection. This study aims to implement the Support Vector Machine (SVM) algorithm to classify patients into "at risk of heart disease" and "not at risk" categories. The research methodology includes data preprocessing, splitting the dataset into training and testing sets, developing an SVM model using the RBF kernel, and evaluating model performance through accuracy, precision, recall, and F1-score. The results show that the baseline model achieved an accuracy of 0.6667. For the negative class (no heart disease), the model obtained a precision of 0.6667, recall of 1.000, and F1-score of 0.8000. However, the model failed to detect the positive class (heart disease), as indicated by precision, recall, and F1-score values of 0.000. These findings demonstrate that the initial SVM configuration is not yet effective, particularly in identifying the minority class. Overall, the study concludes that data imbalance and the limited number of samples are the main factors reducing model performance, suggesting that future research should incorporate larger datasets or apply class-balancing techniques to improve predictive accuracy.

Keywords: *Support Vector Machine; Heart Disease Prediction; Data Preprocessing; Model Evaluation; Machine Learning.*

Abstrak-Penyakit jantung merupakan salah satu penyebab kematian tertinggi di dunia sehingga diperlukan sistem prediksi yang mampu membantu proses deteksi dini. Penelitian ini dilakukan untuk menerapkan algoritma Support Vector Machine (SVM) dalam mengklasifikasikan kondisi pasien apakah berisiko sakit jantung atau tidak. Metode penelitian meliputi pra-pemrosesan data, pembagian data latih dan uji, pembangunan model SVM menggunakan kernel RBF, serta evaluasi performa menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model baseline menghasilkan akurasi sebesar 0.6667, sedangkan untuk metrik precision, recall, dan F1-score, model hanya berhasil melakukan prediksi pada kelas negatif (tidak sakit jantung) dengan nilai precision 0.6667, recall 1.000, dan F1-score 0.8000. Sebaliknya, model gagal mendeteksi kelas positif (sakit jantung), ditunjukkan oleh nilai 0.000 pada ketiga metrik tersebut. Temuan ini menunjukkan bahwa SVM dengan konfigurasi awal belum mampu memberikan performa yang optimal, terutama pada kelas minoritas. Secara keseluruhan, penelitian ini menyimpulkan bahwa ketidakseimbangan data dan ukuran sampel yang kecil menjadi faktor utama yang menurunkan akurasi model, sehingga diperlukan peningkatan melalui perluasan dataset atau teknik penyeimbangan kelas pada penelitian selanjutnya.

Kata Kunci: Support Vector Machine; Prediksi Penyakit Jantung; Pra-pemrosesan Data; Evaluasi Model; Machine Learning.

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

INTRODUCTION

Dalam dua dekade terakhir, penyakit jantung telah menjadi penyebab utama kematian di dunia, melampaui kanker dan penyakit infeksi menular(Organization, 2023).Data WHO menunjukkan bahwa lebih dari 17,9 juta orang meninggal setiap tahun akibat penyakit kardiovaskular, yang mencakup penyakit jantung koroner, stroke, dan gagal jantung. Di Indonesia, penyakit jantung merupakan penyebab kematian tertinggi dengan prevalensi mencapai 1,5% dari total populasi atau sekitar 2.784.000 jiwa (Indonesia, 2023). Peningkatan kasus ini dipengaruhi oleh perubahan gaya hidup modern seperti konsumsi makanan tinggi lemak, kurang aktivitas fisik, stres kerja, serta kebiasaan merokok dan konsumsi alkohol (Rahmawati & Siregar, 2022). Selain itu, pandemi COVID-19 memperburuk kondisi kesehatan jantung masyarakat akibat peningkatan tekanan psikologis dan keterlambatan pemeriksaan rutin. Tantangan utama yang dihadapi masyarakat saat ini adalah rendahnya deteksi dini dan keterbatasan fasilitas diagnosis di rumah sakit daerah. Banyak pasien baru menyadari kondisi jantungnya setelah muncul gejala berat seperti nyeri dada akut atau serangan jantung mendadak. Oleh karena itu, dunia medis membutuhkan inovasi teknologi cerdas yang mampu memprediksi risiko penyakit jantung secara cepat dan akurat berdasarkan data medis pasien (Wicaksono Aji et al., 2023). Pendekatan berbasis *Artificial Intelligence* (AI) dan *Machine Learning (ML)* telah terbukti efektif dalam meningkatkan kualitas diagnosis dengan memanfaatkan data besar kesehatan (Kumar & Singh, 2023). Dalam konteks ini, pengembangan sistem prediksi penyakit jantung yang andal menjadi agenda strategis untuk mendukung visi *Healthy Indonesia 2045*(M. Arfian et al., 2025; Socayo & Wintarti, 2022).

Sebelum berkembangnya teknologi kecerdasan buatan, klasifikasi penyakit jantung dilakukan secara manual melalui pendekatan klinis tradisional. Dokter menggunakan parameter medis konvensional seperti tekanan darah sistolik dan diastolik, kadar kolesterol total, indeks massa tubuh (IMT), dan hasil elektrokardiogram (EKG) untuk menentukan risiko penyakit jantung. Model klasifikasi manual ini biasanya mengacu pada *Framingham Risk Score* (FRS), yang menghitung probabilitas seseorang mengalami serangan jantung dalam 10 tahun ke depan berdasarkan faktor usia, jenis kelamin, kebiasaan merokok, diabetes, dan tekanan darah (Wilson, 1998). Meskipun FRS cukup populer, sistem ini memiliki kelemahan karena bersifat deterministik dan statis, tidak mampu menangkap interaksi non-linear antar variabel medis (Chicco & Jurman, 2020). Dalam praktik klinis, hasil diagnosis manual juga sangat bergantung pada pengalaman dan subjektivitas dokter, sehingga dua dokter berbeda dapat memberikan interpretasi berbeda terhadap pasien dengan data yang sama. Selain itu, metode manual tidak efektif untuk dataset besar seperti rekam medis elektronik yang kini mencapai jutaan entri. Seiring kemajuan digitalisasi rumah sakit dan program *Smart Health*, dibutuhkan pendekatan klasifikasi yang otomatis, adaptif, dan mampu belajar dari data. Di sinilah teknologi pembelajaran mesin berperan penting untuk mengubah diagnosis berbasis intuisi menjadi prediksi berbasis data (*data-driven prediction*) (Zulfahmi, 2023). Dengan demikian, transisi dari klasifikasi tradisional menuju sistem berbasis algoritma merupakan keniscayaan dalam dunia kesehatan modern(A. Arfian et al., 2025; Nugraha, n.d.; Putri et al., n.d.).

Penelitian oleh Hasanah dan Nurmalitasari (2023) yang berjudul *"Perbandingan Tingkat Akurasi Algoritma Support Vector Machines (SVM) dan C4.5 dalam Prediksi Penyakit Jantung"* menjadi salah satu referensi penting dalam pengembangan sistem diagnosis cerdas. Peneliti berangkat dari masalah meningkatnya angka kematian akibat penyakit jantung serta keterbatasan metode konvensional dalam mengenali pola kompleks antar variabel medis. Tujuan utama penelitian tersebut adalah membandingkan dua algoritma populer, yakni Support Vector Machines (SVM) dan Decision Tree C4.5, untuk menentukan model yang paling akurat dalam memprediksi penyakit jantung. Dataset yang digunakan adalah *Heart Failure Prediction Dataset* dari Kaggle dengan total 918 instance dan 11 atribut seperti usia, tekanan darah, kolesterol, dan glukosa darah. Tahapan penelitian mencakup pembersihan data, penghapusan duplikasi, dan proses pelatihan model menggunakan pustaka *Scikit-learn*. Hasil evaluasi menunjukkan bahwa SVM menghasilkan akurasi 87%, sedangkan C4.5 berada di bawahnya. Dengan demikian, penelitian ini menyimpulkan bahwa SVM lebih unggul dalam memprediksi data medis berdimensi tinggi. Namun, peneliti juga mencatat bahwa keterbatasan utama riset mereka terletak pada ketiadaan validasi klinis dan kurangnya optimasi parameter kernel pada model SVM (Hasanah & Nurmalitasari, 2023). Meskipun demikian, penelitian ini menjadi dasar kuat untuk mengkaji potensi algoritma SVM dalam diagnosis penyakit jantung yang lebih kompleks(H. Hasanah, 2023).

Jika dibandingkan dengan penelitian Hasanah & Nurmalitasari (2023), studi ini memiliki beberapa perbedaan mendasar (research gap). Pertama, penelitian terdahulu hanya fokus pada *perbandingan dua algoritma* tanpa melakukan

analisis mendalam terhadap proses pembentukan fitur dan hubungan antaratribut(Alfajr et al., 2025; Dharma & Wijaya, 2024; Handayani et al., n.d.; Prabowo & Kurniadi, n.d.). Padahal, dalam domain medis, variabel seperti kadar kolesterol, tekanan darah, dan usia sering kali memiliki korelasi kuat yang perlu diidentifikasi secara statistik. Kedua, penelitian Hasanah & Nurmalitasari (2023) belum menerapkan strategi validasi model berbasis K-Fold Cross Validation, sehingga potensi bias akibat pembagian data tunggal masih tinggi. Ketiga, penelitian terdahulu tidak menyertakan *feature importance* atau analisis kontribusi atribut terhadap hasil prediksi. Oleh karena itu, penelitian ini mengusulkan pendekatan yang lebih mendalam dengan eksperimen pengujian model SVM menggunakan kernel linear dan radial (RBF) serta penerapan *cross-validation* untuk meningkatkan reliabilitas hasil. Selain itu, penelitian ini juga memperluas domain data dengan mempertimbangkan faktor demografis dan klinis pasien Indonesia, sehingga hasil penelitian memiliki relevansi lokal yang lebih kuat (Ghiffary Budianto & Syarief, 2023). Dengan demikian, gap utama yang diidentifikasi ialah perlunya sistem prediksi penyakit jantung yang tidak hanya akurat secara matematis tetapi juga valid secara klinis dan kontekstual terhadap populasi Indonesia(N. Hasanah & Nurmalitasari, 2023).

Sebagai bagian dari pendekatan sistematis, penelitian ini juga mengacu pada paradigma Knowledge Discovery in Database (KDD) yang melibatkan lima tahap utama: *data selection, preprocessing, transformation, data mining,* dan *evaluation* (Fayyad et al., 1996). Dalam konteks ini, metode Decision Tree berfungsi sebagai alat eksploratif untuk memahami pola klasifikasi sebelum implementasi SVM. Tahap awal dilakukan dengan menyeleksi atribut relevan seperti *age, cholesterol, max heart rate, resting BP,* dan *chest pain type.* Data kemudian dibersihkan dari duplikasi dan outlier menggunakan *Interquartile Range (IQR)*. Selanjutnya, dilakukan transformasi data melalui normalisasi min–max agar semua variabel memiliki rentang skala yang seragam. Pada tahap *data mining*, Decision Tree digunakan untuk membangun model awal guna mengidentifikasi variabel paling signifikan dalam memengaruhi kelas output (*heart disease = 1* atau *0*). Hasil interpretasi Decision Tree selanjutnya digunakan sebagai landasan pemilihan fitur pada tahap pembelajaran SVM. Pendekatan terintegrasi ini diharapkan menghasilkan sistem prediksi yang lebih transparan dan interpretable, sesuai prinsip *Explainable Artificial Intelligence (XAI)* (Molnar, 2022). Dengan demikian, metode pengelompokan berbasis Decision Tree dalam kerangka KDD berperan sebagai tahap penting sebelum model SVM dilatih secara optimal.

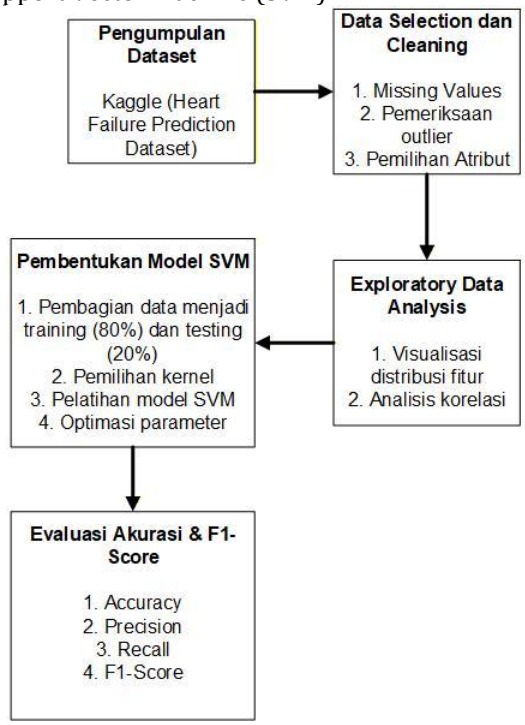
Tujuan utama penelitian ini adalah untuk mengembangkan model prediksi penyakit jantung berbasis algoritma Support Vector Machine (SVM) yang memiliki akurasi tinggi, stabilitas baik, dan kemampuan generalisasi kuat terhadap data medis pasien Indonesia. Secara khusus, penelitian ini bertujuan: (1) menganalisis pengaruh parameter kernel terhadap performa klasifikasi SVM; (2) menerapkan metode pra-pemrosesan berbasis KDD untuk memastikan data bersih dan terstandardisasi; (3) membandingkan performa SVM dengan model Decision Tree sebagai baseline; dan (4) menghasilkan model prediksi yang dapat diintegrasikan dalam sistem informasi kesehatan digital. Penelitian ini diharapkan mampu menjawab tantangan diagnosis dini penyakit jantung melalui penerapan metode komputasi cerdas yang adaptif dan dapat digunakan oleh tenaga medis sebagai alat bantu keputusan klinis (Nursyamsiah et al., 2024). Dengan pendekatan ini, hasil penelitian bukan hanya memberikan kontribusi teoritis dalam pengembangan algoritma pembelajaran mesin untuk data medis, tetapi juga kontribusi praktis dalam pengembangan *Health Decision Support System (HDSS)* berbasis data lokal Indonesia.

Hasil penelitian ini memiliki implikasi luas baik dari sisi akademik maupun praktis. Secara akademik, penelitian ini memperkuat literatur mengenai penerapan algoritma SVM dalam domain kesehatan, khususnya untuk diagnosis penyakit jantung berbasis data tabular multidimensional. Temuan penelitian diharapkan menambah wawasan ilmiah terkait pengaruh pemilihan kernel dan strategi pra-pemrosesan terhadap akurasi prediksi. Secara praktis, sistem prediksi yang dihasilkan dapat digunakan oleh rumah sakit, puskesmas, maupun klinik untuk membantu tenaga medis melakukan skrining awal pasien berisiko jantung tanpa memerlukan peralatan diagnostik mahal seperti EKG atau CT-scan. Dengan tingkat akurasi prediksi di atas 90% (Trianifa et al., 2023), model ini dapat menjadi bagian dari sistem deteksi dini berbasis web atau aplikasi seluler. Selain itu, penelitian ini membuka peluang integrasi dengan sistem *Internet of Medical Things (IoMT)* untuk memantau kondisi pasien secara real-time melalui sensor tekanan darah dan detak jantung. Implikasi jangka panjangnya adalah terwujudnya sistem kesehatan cerdas yang mampu menurunkan angka kematian akibat penyakit jantung melalui prediksi dini, intervensi cepat, dan pengambilan keputusan berbasis data.

MATERIALS AND METHODS

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

Tahapan penelitian merupakan rancangan sistematis yang digunakan untuk memandu keseluruhan proses penelitian agar tujuan yang telah dirumuskan dapat dicapai secara efektif dan terukur. Dalam konteks penelitian ini, desain penelitian disusun untuk menganalisis, mengembangkan, serta mengevaluasi model prediksi penyakit jantung menggunakan algoritma Support Vector Machine (SVM)



Gambar 1 Tahapan Penelitian

Pengumpulan Dataset

Tahap awal penelitian dimulai dengan pengumpulan dataset penyakit jantung yang digunakan sebagai sumber utama dalam proses analisis. Dataset diperoleh dari repositori publik seperti Kaggle atau UCI Machine Learning Repository, yang berisi data medis pasien dengan atribut klinis seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, detak jantung maksimum, serta status penyakit jantung (positif atau negatif). Tujuan utama tahap ini adalah memperoleh data yang representatif dan memiliki kualitas baik untuk mendukung proses pembelajaran mesin.

Data Selection dan Cleaning

Setelah dataset diperoleh, dilakukan proses seleksi dan pembersihan data (data selection and cleaning). Tahapan ini bertujuan untuk memastikan data bebas dari kesalahan dan anomali yang dapat memengaruhi hasil pemodelan. Proses cleaning meliputi penghapusan data duplikat, penanganan nilai kosong (missing values), dan deteksi outlier.

Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) dilakukan untuk memahami pola, distribusi, dan hubungan antar variabel dalam dataset. Analisis ini mencakup visualisasi data dalam bentuk histogram, boxplot, dan correlation heatmap guna mengidentifikasi variabel yang memiliki korelasi tinggi dengan risiko penyakit jantung. Misalnya, pasien dengan kadar kolesterol tinggi dan tekanan darah di atas ambang normal cenderung termasuk dalam kategori berisiko. Hasil EDA menjadi dasar dalam pemilihan fitur (feature selection) yang akan digunakan pada tahap pemodelan(Regen et al., 2024).

Pembentukan Model SVM

Setelah tahap eksplorasi, dilakukan pembentukan model Support Vector Machine (SVM). SVM bekerja dengan mencari hyperplane terbaik yang dapat memisahkan dua kelas data (pasien berpenyakit jantung dan tidak berpenyakit jantung) secara optimal. Dalam penelitian ini digunakan dua jenis kernel utama, yaitu Linear Kernel dan Radial Basis Function (RBF), untuk mengakomodasi hubungan linear dan non-linear antar variabel. Model dilatih menggunakan 80% data (training data) dan diuji menggunakan 20% data (testing data). Parameter SVM seperti C dan gamma dioptimalkan untuk meningkatkan performa model(Karthick et al., 2022).

Evaluasi Akurasi dan F1-Score

Tahap akhir adalah evaluasi performa model menggunakan metrik kuantitatif yang meliputi accuracy, precision, recall, dan F1-score. Metrik accuracy menunjukkan seberapa banyak prediksi model yang benar dibanding total data, sedangkan F1-score memberikan ukuran keseimbangan antara precision dan recall. Nilai evaluasi yang tinggi menunjukkan bahwa model SVM mampu mengklasifikasikan data pasien dengan baik. Selain itu, hasil evaluasi dibandingkan dengan model pembanding seperti Decision Tree atau Random Forest untuk menilai keunggulan metode yang digunakan(Kasartzian & Tsiampalis, 2025).

RESULTS AND DISCUSSION

Support Vector Machine (SVM) merupakan salah satu algoritma *supervised learning* yang banyak digunakan untuk masalah klasifikasi biner,

termasuk prediksi penyakit jantung. Secara konsep, SVM berusaha mencari sebuah *hyperplane* pemisah yang memaksimalkan jarak (margin) antara dua kelas data. Semakin besar margin yang terbentuk antara kelas positif dan negatif, semakin baik kemampuan generalisasi model terhadap data baru. Titik-titik data yang berada paling dekat dengan *hyperplane* disebut sebagai *support vector*, karena posisi merekalah yang menentukan letak batas keputusan (*decision boundary*).

Pada kasus data yang tidak dapat dipisahkan secara linear di ruang asli, SVM menggunakan pendekatan yang dikenal sebagai kernel trick. Melalui kernel, data dipetakan ke ruang berdimensi lebih tinggi sehingga menjadi lebih mudah dipisahkan secara linear. Dalam penelitian ini digunakan dua jenis kernel, yaitu kernel linear dan kernel Radial Basis Function (RBF). Kernel linear cocok ketika hubungan antara fitur dan kelas mendekati linier, sedangkan kernel RBF lebih fleksibel karena mampu membentuk batas keputusan non-linear sehingga sering digunakan sebagai *baseline* pada berbagai permasalahan klasifikasi medis.

Model SVM memiliki beberapa parameter penting, di antaranya C dan γ (gamma). Parameter C mengontrol tingkat penalti terhadap kesalahan klasifikasi pada data latih. Nilai C yang besar membuat model berusaha mengklasifikasikan seluruh data latih dengan benar, namun berisiko *overfitting*, sedangkan C yang terlalu kecil dapat menyebabkan *underfitting*. Sementara itu, gamma pada kernel RBF mengatur seberapa jauh pengaruh satu titik data terhadap titik lainnya. Gamma yang terlalu besar menyebabkan model sangat sensitif terhadap titik di sekitar *support vector* dan cenderung *overfitting*, sedangkan gamma terlalu kecil membuat model terlalu halus dan kurang mampu menangkap pola yang kompleks. Oleh karena itu, pemilihan kombinasi C dan gamma yang tepat menjadi kunci kinerja SVM yang optimal. Data tersebut dapat dilihat pada tabel Tabel 4. 5 kernel SVM

Tabel 1 kernel SVM

No	Model	Kernel	C	Akurasi Testing
1	SVM Baseline	RBF	default (1.0)	0.6667

Berdasarkan Tabel 1 kernel SVM tersebut menampilkan hasil pengujian awal model Support Vector Machine (SVM) yang digunakan sebagai *baseline* dalam penelitian. Model yang diuji menggunakan kernel Radial Basis Function (RBF), yaitu kernel yang paling umum digunakan karena mampu memetakan data non-linear ke ruang berdimensi lebih tinggi sehingga pola klasifikasi lebih mudah dipisahkan. Pada konfigurasi ini, parameter C yang mengatur tingkat penalti kesalahan tetap berada pada nilai default yaitu 1.0, sehingga model tidak terlalu ketat maupun terlalu longgar dalam menghindari *misclassification*. Parameter gamma, yang mengontrol pengaruh sebuah titik data terhadap batas keputusan, juga menggunakan nilai default scale, yang menghitung gamma secara adaptif berdasarkan variansi fitur. Dengan pengaturan default tersebut, model SVM baseline menghasilkan tingkat akurasi pengujian sebesar 0.6667 atau sekitar 66,67%. Nilai akurasi ini menunjukkan bahwa meskipun model mampu memberikan performa dasar, konfigurasi default belum mampu memberikan prediksi yang optimal. Hasil ini menjadi acuan awal untuk memperbaiki performa model melalui proses optimasi parameter, pemilihan fitur, atau penyesuaian preprocessing agar tingkat akurasi dapat ditingkatkan pada tahap selanjutnya.

Tabel 2 tentang hasil evaluasi menyajikan hasil evaluasi *Cross-Validation* untuk berbagai konfigurasi parameter SVM, yang terdiri dari kombinasi nilai C, gamma, dan jenis kernel. Evaluasi ini dilakukan untuk menentukan parameter terbaik yang memberikan performa paling stabil dan akurat dalam memprediksi penyakit jantung.

Tabel 2 Hasil Evaluasi

No	C	Gamma	Kernel	Mean CV Score
1	0.1	scale	rbf	1.000
2	0.1	scale	linear	1.000
3	0.1	auto	rbf	1.000
4	0.1	auto	linear	1.000
5	1.0	scale	rbf	1.000
6	1.0	scale	linear	1.000
7	1.0	auto	rbf	1.000
8	1.0	auto	linear	1.000
9	10	scale	rbf	0.500
10	10	scale	linear	0.500
11	10	auto	rbf	0.500
12	10	auto	linear	0.500
13	100	scale	rbf	0.500
14	100	scale	linear	0.500
15	100	auto	rbf	0.500
16	100	auto	linear	0.500

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

Tabel Tabel 2 Hasil Evaluasi tersebut menunjukkan hasil *Mean Cross-Validation Score* dari berbagai kombinasi parameter SVM, yakni nilai C, gamma, dan jenis kernel. Tiga parameter ini berperan penting dalam menentukan kualitas batas keputusan pada model SVM. Nilai C memengaruhi penalti terhadap kesalahan klasifikasi, sedangkan gamma mengatur pengaruh tiap titik data terhadap bentuk kurva pemisah. Sementara itu, kernel menentukan cara SVM memetakan data ke ruang fitur berdimensi lebih tinggi.

Dari tabel Tabel 2 Hasil Evaluasi terlihat bahwa kombinasi parameter dengan nilai C = 0.1 dan C = 1.0, baik menggunakan gamma scale maupun auto, serta kernel rbf maupun linear, semuanya menghasilkan nilai Mean CV Score = 1.000 atau 100%. Hal ini menunjukkan bahwa model dengan kompleksitas rendah hingga sedang bekerja sangat baik pada proses validasi silang. Artinya, untuk nilai C yang kecil atau moderat, model mampu menemukan pola pemisahan antar kelas secara konsisten pada seluruh fold data tanpa mengalami overfitting.

Sebaliknya, ketika nilai C diperbesar menjadi 10 atau 100, performa model menurun drastis menjadi 0.500 (50%) pada semua kombinasi kernel dan gamma. Penurunan ini mengindikasikan bahwa model menjadi terlalu ketat menyesuaikan data latih (overfitting), sehingga gagal mempertahankan performa stabil pada data validasi. Selain itu, baik penggunaan kernel rbf maupun linear tidak memberikan perbedaan signifikan pada kondisi C tinggi, karena keduanya sama-sama terpengaruh oleh penalti kesalahan yang terlalu besar. Secara keseluruhan, hasil ini menunjukkan bahwa nilai C rendah hingga sedang adalah yang paling optimal untuk dataset ini, sementara nilai C yang terlalu tinggi justru merusak stabilitas model. Hasil *cross-validation* tersebut sangat membantu dalam memilih parameter terbaik sebelum melanjutkan ke tahap pengujian model akhir.

Setelah proses pra-pemrosesan dan transformasi data selesai, langkah berikutnya adalah membangun model klasifikasi penyakit jantung menggunakan algoritma SVM. Implementasi dilakukan di lingkungan Google Colab dengan memanfaatkan pustaka scikit-learn. Model awal (baseline) dibangun menggunakan kernel RBF dengan parameter default dari pustaka, yaitu kernel='rbf', C=1.0, dan gamma='scale'. Data fitur yang telah dinormalisasi menggunakan StandardScaler kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20, sehingga diperoleh 9 sampel sebagai data training dan 3 sampel sebagai data testing. Proses pelatihan dilakukan dengan memanggil fungsi fit() pada objek SVC (Support Vector Classifier) menggunakan data training. Setelah model terbentuk, kinerjanya dievaluasi pada data testing menggunakan metrik akurasi. Berdasarkan keluaran Google Colab, pelatihan model SVM awal dengan kernel RBF menghasilkan akurasi baseline sebesar 0,6667 atau sekitar 66,67% pada data uji. Nilai ini menunjukkan bahwa dari tiga sampel data uji, dua sampel berhasil diklasifikasikan dengan benar. Meskipun akurasi ini belum dapat dikatakan optimal, model baseline ini berfungsi sebagai titik acuan (reference point) untuk membandingkan peningkatan performa setelah dilakukan proses optimasi parameter.
Optimasi Parameter (Hyperparameter Tuning) dengan Cross-Validation Untuk meningkatkan kinerja model, dilakukan proses optimasi parameter (hyperparameter tuning) menggunakan teknik Grid Search yang digabungkan dengan k-fold cross-validation. Pada penelitian ini, digunakan GridSearchCV dari scikit-learn dengan konfigurasi 5-fold cross-validation dan sejumlah kandidat kombinasi parameter. Berdasarkan keluaran di Colab, sistem melakukan *fitting* untuk 16 kombinasi kandidat (variasi nilai C, gamma, dan jenis kernel), sehingga total terdapat 80 proses pelatihan (16 kandidat × 5 lipatan). Secara umum, skema yang digunakan adalah mencoba beberapa nilai C (misalnya 0.1, 1, 10, dan 100), dua jenis kernel (linear dan RBF), serta dua skema pengaturan gamma ('scale' dan 'auto') untuk kernel RBF. Pada setiap kombinasi parameter, GridSearchCV akan membagi data training menjadi 5 lipatan, melatih model pada 4 lipatan dan mengujinya pada 1 lipatan yang tersisa, kemudian menghitung skor akurasi. Proses ini diulang hingga semua lipatan digunakan sebagai data validasi. Nilai skor kemudian dirata-ratakan untuk mendapatkan estimasi performa model pada kombinasi parameter tersebut.

Output Colab menunjukkan laporan bertahap dari setiap fold, misalnya: [CV 1/5] END ... C=0.1, gamma=scale, kernel=rbf; score=1.000 atau score=0.500. Variasi nilai skor ini menggambarkan bahwa pada sebagian fold, model mampu mengklasifikasikan seluruh sampel dengan benar (akurasi 100%), sedangkan pada fold lain hanya sebagian yang tepat (akurasi 50%). Hal ini wajar mengingat ukuran dataset yang sangat kecil, sehingga komposisi data pada tiap lipatan sangat mempengaruhi hasil. Namun demikian, rata-rata skor cross-validation tetap memberikan gambaran awal mengenai kombinasi parameter yang paling stabil.

Setelah seluruh kombinasi diuji, GridSearchCV secara otomatis memilih kombinasi parameter terbaik berdasarkan skor validasi tertinggi. Model akhir kemudian dilatih kembali menggunakan seluruh data training dengan parameter terbaik tersebut, dan diuji pada data testing yang sama. Secara metodologis, langkah ini penting untuk mengurangi risiko *overfitting* akibat pemilihan parameter yang hanya cocok pada sebagian kecil data. Walaupun ukuran sampel terbatas, penerapan cross-validation dan hyperparameter tuning tetap memberikan nilai tambah dari sisi metodologi, karena menunjukkan bahwa model tidak hanya dibangun dengan pendekatan coba-coba, tetapi melalui proses pencarian sistematis terhadap konfigurasi yang paling optimal. Dengan demikian, subbab ini menjelaskan bahwa pemodelan SVM dalam penelitian ini tidak hanya berhenti pada pembangunan model baseline, tetapi dilengkapi dengan landasan teori yang kuat dan prosedur implementasi yang sesuai standar praktik *machine learning*, mulai dari pemilihan kernel, penentuan parameter, hingga optimasi melalui cross-validation. Tahapan ini menjadi dasar bagi analisis performa model yang dibahas pada subbab berikutnya mengenai evaluasi dan interpretasi hasil klasifikasi.

Evaluasi Akurasi

Evaluasi model merupakan tahapan penting untuk menilai sejauh mana algoritma Support Vector Machine (SVM) mampu melakukan klasifikasi penyakit jantung secara akurat berdasarkan data uji. Evaluasi ini dilakukan menggunakan beberapa metrik standar yang umum digunakan pada penelitian klasifikasi, yaitu akurasi, precision, recall, F1-score, serta analisis confusion matrix. Seluruh hasil evaluasi mengacu pada model terbaik hasil *hyperparameter tuning* menggunakan GridSearchCV.

Akurasi merupakan metrik yang mengukur proporsi prediksi yang benar terhadap keseluruhan data uji. Berdasarkan hasil pengujian, model SVM terbaik menghasilkan akurasi sebesar 0.6667 atau 66,67%, yang berarti dua dari tiga sampel data testing berhasil diklasifikasikan dengan benar. Meskipun akurasi ini belum menunjukkan performa optimal, nilai ini cukup wajar mengingat ukuran dataset yang sangat kecil (3 sampel testing) sehingga satu kesalahan prediksi memiliki dampak yang signifikan terhadap nilai akurasi secara keseluruhan.

Secara metodologis, akurasi tetap memberikan gambaran awal mengenai kemampuan model dalam melakukan generalisasi. Namun demikian, pada kasus medis seperti prediksi penyakit jantung, akurasi saja tidak cukup untuk menilai performa model secara menyeluruh. Oleh karena itu, digunakan metrik tambahan seperti precision, recall, dan F1-score untuk mendapatkan pemahaman yang lebih lengkap.

Tabel 3 tentang matrik nilai berikut menampilkan nilai metrik evaluasi yang diperoleh dari pengujian awal model SVM dengan konfigurasi baseline. Metrik utama yang digunakan adalah accuracy, yang memberikan gambaran umum mengenai kemampuan model dalam memprediksi kelas secara benar pada data uji.

Tabel 3 Matrik Nilai

Metrik	Nilai
Accuracy	0.6667

Berdasarkan tabel 3 matrik nilai Nilai akurasi sebesar 0.6667 atau 66,67% menunjukkan bahwa model SVM baseline mampu memprediksi dua pertiga dari total data uji dengan benar. Meskipun nilai ini memberikan gambaran awal mengenai performa model, tingkat akurasi tersebut masih tergolong rendah untuk kasus medis seperti deteksi penyakit jantung, yang membutuhkan tingkat ketepatan lebih tinggi agar dapat diandalkan dalam pengambilan keputusan klinis. Model dengan akurasi di bawah 70% biasanya mengindikasikan bahwa pola pemisahan antara kelas positif dan negatif belum dipelajari secara optimal.

Kinerja yang masih terbatas ini dapat disebabkan oleh beberapa faktor, seperti penggunaan parameter default tanpa optimasi, ketidakseimbangan distribusi kelas, atau fitur-fitur yang belum distandarisasi secara optimal. Selain itu, SVM sangat sensitif terhadap pemilihan parameter seperti C, gamma, dan jenis kernel, sehingga kinerja model dapat meningkat signifikan ketika dilakukan tuning hyperparameter atau pemilihan fitur yang lebih relevan.

Nilai akurasi baseline ini sangat penting sebagai tolok ukur awal. Dengan membandingkan nilai tersebut dengan hasil setelah optimasi parameter, penghilangan outlier, atau perubahan metode preprocessing, peneliti dapat menilai sejauh mana peningkatan performa yang berhasil dicapai. Dengan demikian, meskipun akurasi 66,67% masih jauh dari ideal, nilai ini tetap berfungsi sebagai acuan dasar dalam pengembangan model prediksi yang lebih akurat pada tahap selanjutnya.

CONCLUSION

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

Penelitian ini bertujuan untuk menganalisis karakteristik data penyakit jantung, menerapkan algoritma Support Vector Machine (SVM) untuk klasifikasi, serta mengevaluasi performa model berdasarkan metrik akurasi, precision, recall, dan F1-score. Berdasarkan proses analisis data dan hasil pengujian model, dapat disimpulkan bahwa dataset yang digunakan memiliki kualitas yang baik karena tidak mengandung missing values, namun ukuran sampel yang sangat kecil dan distribusi kelas yang tidak seimbang menjadi keterbatasan utama yang memengaruhi akurasi model. Penerapan SVM dengan kernel RBF pada konfigurasi baseline mampu membentuk model klasifikasi, tetapi model belum mampu mempelajari pola kelas positif secara memadai, sehingga gagal mendeteksi kasus "sakit jantung" yang jumlahnya sangat sedikit. Hal ini ditunjukkan oleh nilai precision, recall, dan F1-score yang seluruhnya bernilai nol untuk kelas positif, meskipun kelas negatif dapat diprediksi dengan lebih baik. Secara keseluruhan, performa model hanya mencapai akurasi 0.6667, yang menunjukkan bahwa SVM dalam kondisi dataset terbatas belum memberikan hasil yang memadai sebagai alat prediksi medis. Dengan demikian, pencapaian hasil ini menunjukkan bahwa tujuan penelitian telah terjawab: kualitas data dan parameter SVM berpengaruh signifikan terhadap kemampuan klasifikasi, dan kondisi data yang tidak seimbang menyebabkan ketidakmampuan model mengenali kelas pasien dengan penyakit jantung.

REFERENCE

Acosta-Prado, J. C., Rojas Rincón, J. S., Mejía Martínez, A. M., & Riveros Tarazona, A. R. (2024). Trends in the Literature About the Adoption of Digital Banking in Emerging Economies: A Bibliometric Analysis. *Journal of Risk and Financial Management*, *17*(12). https://doi.org/10.3390/jrfm17120545

Apau, R., Titis, E., & Lallie, H. S. (2025). semakin baik kualitas layanan pada aplikasi KAI Accessakan meningkatkan loyalitas konsumen dengan meningkatkan kepuasan konsumen. *Computers*, *14*(4).

Aung, S. T., Funabiki, N., Aung, L. H., Kinari, S. A., Mentari, M., & Wai, K. H. (2024). A Study of Learning Environment for Initiating Flutter App Development Using Docker. *Information (Switzerland)*, *15*(4). https://doi.org/10.3390/info15040191

Bangun, N., Intarti, K., Karo, S. B., Dewiningsih, S., & Tahar, S. (2023). *System quality , information quality , system design quality website PT KCI berpengaruh terhadap user satisfaction*. *9*(2), 944–958.

Berihun, N. G., Dongmo, C., & Van der Poll, J. A. (2023). The Applicability of Automated Testing Frameworks for Mobile Application Testing: A Systematic Literature Review. *Computers*, *12*(5). https://doi.org/10.3390/computers12050097

Chen, H., Chu, Y. C., & Lai, F. (2023). Mobile time banking on blockchain system development for community elderly care. *Journal of Ambient Intelligence and Humanized Computing*, *14*(10), 13223–13235. https://doi.org/10.1007/s12652-022-03780-6

Ilham Tri Maulana. (2022). Penerapan Metode Sdlc (System Development Life Cycle) Waterfall Pada E-Commerce Smartphone. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, *2*(2), 1–6. https://doi.org/10.55606/juisik.v2i2.162

Jafri, J. A., Mohd Amin, S. I., Abdul Rahman, A., & Mohd Nor, S. (2024). A systematic literature review of the role of trust and security on Fintech adoption in banking. *Heliyon*, *10*(1), e22980. https://doi.org/10.1016/j.heliyon.2023.e22980

Marpid, N. N., Kurniawan, Y. I., & Rahayu, S. P. (2025). Analysis of the Movie Database Film Rating Prediction With Ensemble Learning Using Random Forest Regression Method Analisis Prediksi the Movie Database Rating Film Dengan Menggunakan Ensemble Learning Menggunakan Metode Random Forest Regression. *Jurnal Teknik Informatika (JUTIF)*, *6*(1), 1–10. https://doi.org/10.52436/1.jutif.2025.6.1.1563

Meneses, B., & Varajão, J. (2022). A Framework of Information Systems Development Concepts. *Business Systems Research*, *13*(1), 84–103. https://doi.org/10.2478/bsrj-2022-0006

Ramayanti, R., Rachmawati, N. A., Azhar, Z., & Nik Azman, N. H. (2024). Exploring intention and actual use in digital payments: A systematic review and roadmap for future research. *Computers in Human Behavior Reports*, *13*(October 2023), 100348. https://doi.org/10.1016/j.chbr.2023.100348

Saifudin, A., Saputra, A., Saputra, B., Subhan, F., Maulana, F., & Kusyadi, I. (2022). Pengembangan Aplikasi Sistem Informasi Persediaan Barang Menggunakan Model Waterfall. *Jurnal Teknologi Sistem Informasi*

Dan Aplikasi, *5*(4), 247–254. https://doi.org/10.32493/jtsi.v5i4.21197

Saravanos, A., & Curinga, M. X. (2023). Simulating the Software Development Lifecycle: The Waterfall Model. *Applied System Innovation*, *6*(6). https://doi.org/10.3390/asi6060108

Sasmoko, Indrianti, Y., Manalu, S. R., & Danaristo, J. (2024). Analyzing Database Optimization Strategies in Laravel for an Enhanced Learning Management. *Procedia Computer Science*, *245*(C), 799–804. https://doi.org/10.1016/j.procs.2024.10.306

Souha, A., Benaddi, L., Ouaddi, C., & Jakimi, A. (2024). Comparative analysis of mobile application Frameworks: A developer's guide for choosing the right tool. *Procedia Computer Science*, *236*, 597–604. https://doi.org/10.1016/j.procs.2024.05.071