Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Prediksi Risiko Penyakit Jantung Menggunakan Support Vector Machine dengan Seleksi Fitur dan Optimasi Hyperparamete

Nurhadi Surojudin[1*], Sufajar Butsianto, Rizal Ainun Yaqin

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi
Ruko Bekasi Mas, Jl. Ahmad Yani, Marga Jaya, Kec. Bekasi Selatan, Kota Bekas, Jawa Barat, Indonesia
Email: [1,*]nurhadi@pelitabangsa.ac.id, [2]sufajar@pelitabangsa.ac.id, [3]rizal.ay@gmail.com
Email Penulis Korespondensi: nurhadi@pelitabangsa.ac.id
Submitted: **29/06/2026;** Accepted: **31/07/2026;** Published: **31/07/2026**

Abstrak−Penyakit jantung merupakan salah satu penyebab utama kematian di dunia sehingga diperlukan metode prediksi yang mampu membantu proses diagnosis secara lebih cepat dan akurat. Penelitian ini bertujuan mengembangkan model prediksi risiko penyakit jantung menggunakan algoritma Support Vector Machine (SVM) yang dikombinasikan dengan Recursive Feature Elimination (RFE) sebagai metode seleksi fitur dan GridSearchCV untuk optimasi hyperparameter. Dataset yang digunakan adalah Cleveland Heart Disease Dataset yang terdiri atas 303 data pasien dengan 13 atribut prediktor dan 1 atribut target. Tahapan penelitian meliputi pengumpulan dataset, preprocessing data, seleksi fitur, optimasi hyperparameter, pembangunan model, serta evaluasi performa menggunakan Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil eksperimen menunjukkan bahwa kombinasi RFE dan GridSearchCV menghasilkan model SVM dengan nilai Accuracy sebesar 86,89%, Precision 85,71%, Recall 85,71%, F1-score 85,71%, dan ROC-AUC 93,18%. Seleksi fitur berhasil mengurangi atribut yang kurang relevan sehingga meningkatkan efisiensi model, sedangkan optimasi hyperparameter menghasilkan kombinasi parameter yang lebih optimal dibandingkan parameter bawaan. Hasil penelitian menunjukkan bahwa integrasi RFE dan GridSearchCV mampu meningkatkan kualitas model prediksi risiko penyakit jantung serta berpotensi diterapkan sebagai pendukung pengambilan keputusan pada sistem diagnosis berbasis machine learning.

Kata kunci: Support Vector Machine; Recursive Feature Elimination; GridSearchCV; Seleksi Fitur; Penyakit Jantung

Abstract−Heart disease remains one of the leading causes of mortality worldwide, highlighting the need for accurate prediction models to support early diagnosis and clinical decision-making. This study proposes a heart disease risk prediction model based on Support Vector Machine (SVM) integrated with Recursive Feature Elimination (RFE) for feature selection and GridSearchCV for hyperparameter optimization. The study utilized the Cleveland Heart Disease Dataset, consisting of 303 patient records, 13 predictive attributes, and one target variable. The research workflow included dataset collection, data preprocessing, feature selection, hyperparameter optimization, model development, and performance evaluation using Accuracy, Precision, Recall, F1-score, and ROC-AUC. Experimental results demonstrate that the proposed model achieved an Accuracy of 86.89%, Precision of 85.71%, Recall of 85.71%, F1-score of 85.71%, and ROC-AUC of 93.18%. Feature selection successfully reduced irrelevant attributes, improving model efficiency, while hyperparameter optimization produced a more effective parameter configuration than the default settings. These findings indicate that integrating RFE with GridSearchCV enhances the predictive performance of SVM and provides a promising approach for supporting heart disease diagnosis using machine learning techniques.

Keywords: Support Vector Machine; Recursive Feature Elimination; GridSearchCV; Feature Selection; Heart Disease

1. PENDAHULUAN

Penyakit jantung (heart disease) masih menjadi salah satu penyebab utama kematian di berbagai negara, termasuk Indonesia [1], [2]. Organisasi kesehatan dunia melaporkan bahwa penyakit kardiovaskular menyumbang proporsi terbesar dari angka kematian global setiap tahunnya [3], [4], [5]. Tingginya prevalensi penyakit jantung dipengaruhi oleh berbagai faktor risiko seperti usia, hipertensi, kadar kolesterol, diabetes, kebiasaan merokok, hingga gaya hidup yang kurang sehat [6], [7]. Kondisi tersebut mendorong pentingnya deteksi dini terhadap risiko penyakit jantung sehingga penanganan medis dapat dilakukan lebih cepat sebelum terjadi komplikasi yang lebih serius [8], [9]. Oleh karena itu, pemanfaatan teknologi kecerdasan buatan (Artificial Intelligence) dan Machine Learning menjadi salah satu alternatif yang banyak dikembangkan untuk membantu proses prediksi risiko penyakit jantung secara lebih cepat, objektif, dan efisien [10], [11].

Di bidang Machine Learning, berbagai algoritma klasifikasi telah diterapkan untuk memprediksi risiko penyakit jantung, di antaranya Decision Tree [12], Naïve Bayes [13], [14], Random Forest, K-Nearest Neighbor, Artificial Neural Network, Extreme Gradient Boosting, dan Support Vector Machine (SVM) [14], [15]. Di antara berbagai metode tersebut, SVM dikenal memiliki kemampuan yang baik dalam menangani data berdimensi tinggi serta menghasilkan model klasifikasi yang stabil meskipun jumlah data relatif terbatas [16]. Karakteristik tersebut menjadikan SVM banyak digunakan pada berbagai penelitian kesehatan, termasuk klasifikasi penyakit jantung menggunakan Cleveland Heart Disease Dataset.

Meskipun demikian, performa SVM sangat dipengaruhi oleh kualitas fitur yang digunakan dan pemilihan nilai hyperparameter [17], [18]. Penggunaan seluruh fitur pada dataset tidak selalu menghasilkan performa terbaik karena beberapa variabel dapat bersifat redundan, kurang informatif, atau bahkan menimbulkan noise. Selain itu, penggunaan nilai hyperparameter bawaan (default parameter) sering kali belum mampu menghasilkan model dengan tingkat akurasi yang optimal. Oleh sebab itu, diperlukan strategi untuk memilih fitur-fitur yang paling

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

relevan sekaligus melakukan optimasi hyperparameter sehingga model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik.

Beberapa penelitian terdahulu telah menunjukkan bahwa SVM merupakan salah satu algoritma yang memiliki performa tinggi dalam klasifikasi penyakit jantung [19]. Penelitian pertama menerapkan SVM pada dataset penyakit jantung tanpa proses seleksi fitur dan memperoleh tingkat akurasi yang cukup baik, namun masih ditemukan beberapa kesalahan klasifikasi pada data pengujian. Penelitian berikutnya menggunakan metode Recursive Feature Elimination (RFE) sebagai teknik seleksi fitur sebelum proses klasifikasi sehingga mampu mengurangi jumlah fitur tanpa menurunkan performa model [20]. Penelitian lain mengombinasikan SVM dengan Grid Search Cross Validation untuk memperoleh kombinasi hyperparameter terbaik dan berhasil meningkatkan nilai akurasi dibandingkan penggunaan parameter bawaan. Selanjutnya, penelitian yang memanfaatkan metode Randomized Search menunjukkan bahwa optimasi hyperparameter dapat mempercepat proses pencarian parameter optimal dengan performa yang tetap kompetitif [21]. Selain itu, penelitian terbaru juga mengombinasikan teknik seleksi fitur dengan algoritma klasifikasi berbasis SVM untuk meningkatkan akurasi sekaligus mengurangi kompleksitas model [22].

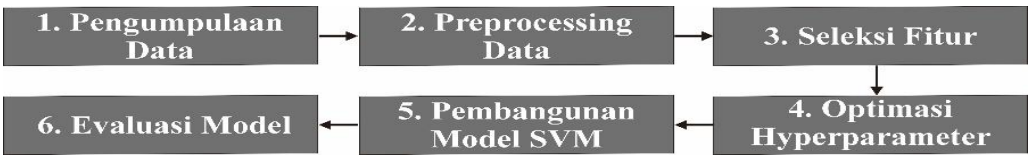
Berdasarkan penelitian-penelitian tersebut dapat disimpulkan bahwa seleksi fitur maupun optimasi hyperparameter secara terpisah telah memberikan kontribusi terhadap peningkatan performa model klasifikasi. Namun demikian, masih terdapat penelitian yang hanya menerapkan salah satu pendekatan tersebut sehingga belum mengevaluasi pengaruh kombinasi keduanya terhadap performa SVM secara komprehensif. Selain itu, sebagian penelitian hanya melaporkan nilai akurasi tanpa membandingkan metrik evaluasi lain seperti precision, recall, F1-score, dan Area Under Curve (AUC). Kondisi tersebut menunjukkan masih adanya peluang penelitian untuk mengembangkan model prediksi penyakit jantung yang lebih optimal melalui integrasi seleksi fitur dan optimasi hyperparameter.

Penelitian ini mengusulkan penerapan algoritma Support Vector Machine yang dipadukan dengan seleksi fitur dan optimasi hyperparameter pada Cleveland Heart Disease Dataset [23]. Tahap seleksi fitur dilakukan untuk memilih atribut yang paling berpengaruh terhadap prediksi penyakit jantung sehingga mampu mengurangi dimensi data dan menghilangkan fitur yang kurang relevan. Selanjutnya, optimasi hyperparameter dilakukan menggunakan Grid Search Cross Validation guna memperoleh kombinasi parameter terbaik, seperti nilai C, gamma, dan jenis kernel, sehingga model memiliki performa klasifikasi yang lebih optimal.

Kontribusi utama penelitian ini adalah mengevaluasi pengaruh kombinasi seleksi fitur dan optimasi hyperparameter terhadap peningkatan kinerja algoritma SVM dalam memprediksi risiko penyakit jantung. Evaluasi dilakukan dengan membandingkan empat skenario, yaitu SVM standar sebagai model dasar (baseline), SVM dengan seleksi fitur, SVM dengan optimasi hyperparameter, serta SVM yang mengombinasikan seleksi fitur dan optimasi hyperparameter. Performa setiap model dianalisis menggunakan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil penelitian diharapkan dapat memberikan alternatif model klasifikasi yang lebih akurat sekaligus menjadi referensi bagi pengembangan sistem pendukung keputusan pada bidang kesehatan, khususnya untuk deteksi dini risiko penyakit jantung.

2. METODOLOGI PENELITIAN

Penelitian ini bertujuan mengembangkan model prediksi risiko penyakit jantung menggunakan algoritma Support Vector Machine (SVM) dengan penerapan seleksi fitur dan optimasi hyperparameter. Dataset yang digunakan adalah Cleveland Heart Disease Dataset yang terdiri atas 303 data pasien dengan 13 atribut prediktor dan satu atribut target yang menunjukkan status penyakit jantung.



Gambar 1. Tahapan Penelitian

Alur penelitian yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Penelitian diawali dengan pengumpulan dataset, dilanjutkan dengan preprocessing data, seleksi fitur menggunakan Recursive Feature Elimination (RFE), optimasi hyperparameter menggunakan GridSearchCV, pembangunan model Support Vector Machine (SVM), dan diakhiri dengan evaluasi performa menggunakan berbagai metrik klasifikasi.

2.1 Pengumpulan Dataset

Tahap pertama penelitian adalah pengumpulan Cleveland Heart Disease Dataset yang berasal dari Cleveland Clinic Foundation dan dipublikasikan melalui UCI Machine Learning Repository. Dataset ini dipilih karena merupakan benchmark dataset yang banyak digunakan dalam penelitian prediksi penyakit jantung, memiliki atribut klinis yang lengkap, serta tersedia secara publik sehingga hasil penelitian dapat direproduksi dan dibandingkan dengan penelitian lain.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Dataset terdiri atas 303 data pasien dengan 13 atribut prediktor dan 1 atribut target. Selanjutnya, data diproses melalui tahapan preprocessing, seleksi fitur menggunakan Recursive Feature Elimination (RFE), optimasi hyperparameter menggunakan GridSearchCV, pembangunan model Support Vector Machine (SVM), dan evaluasi performa untuk memperoleh model prediksi yang optimal. Variabel yang digunakan dalam penelitian ini beserta deskripsinya disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Variabel	Deskripsi
1	age	Usia pasien (tahun)
2	sex	Jenis kelamin (1 = laki-laki, 0 = perempuan)
3	cp	Jenis nyeri dada (chest pain)
4	trestbps	Tekanan darah saat istirahat (mmHg)
5	chol	Kadar kolesterol serum (mg/dL)
6	fbs	Gula darah puasa (>120 mg/dL)
7	restecg	Hasil pemeriksaan elektrokardiogram saat istirahat
8	thalach	Denyut jantung maksimum yang dicapai
9	exang	Angina akibat aktivitas fisik (exercise-induced angina)
10	oldpeak	Depresi segmen ST akibat aktivitas fisik
11	slope	Kemiringan segmen ST saat puncak latihan
12	ca	Jumlah pembuluh darah utama yang terdeteksi melalui fluoroskopi
13	thal	Status thalassemia
14	target	Status penyakit jantung (0 = tidak berisiko, 1 = berisiko)

Berdasarkan hasil eksplorasi data, pada tabel 1, dataset terdiri atas 303 observasi, dengan 165 data pada kelas 0 (tidak berisiko penyakit jantung) dan 138 data pada kelas 1 (berisiko penyakit jantung). Distribusi kelas yang relatif seimbang mengurangi potensi class imbalance, sehingga model dapat dilatih tanpa memerlukan teknik data balancing seperti oversampling atau undersampling.

Cleveland Heart Disease Dataset dipilih karena memiliki ukuran yang relatif ringan, atribut klinis yang lengkap, dan telah banyak digunakan sebagai benchmark pada penelitian prediksi penyakit jantung. Hal ini memungkinkan hasil penelitian dibandingkan secara langsung dengan penelitian terdahulu yang menggunakan dataset yang sama.

2.2 Preprocessing Data

Tahap preprocessing bertujuan untuk menyiapkan data sebelum proses pelatihan model. Tahapan ini meliputi pemeriksaan struktur data, missing value, data duplikat, pembagian data, dan standardisasi. Hasil pemeriksaan menunjukkan bahwa dataset terdiri atas 303 data pasien dengan 13 atribut prediktor dan 1 atribut target. Seluruh atribut bertipe numerik, tidak memiliki missing value, dan tidak terdapat data duplikat sehingga seluruh data dapat digunakan dalam penelitian.

Selanjutnya, dataset dibagi menjadi 80% data latih dan 20% data uji menggunakan metode train-test split dengan parameter random_state = 42 dan stratify = y agar proporsi kelas tetap terjaga. Proses standardisasi dihitung menggunakan Persamaan (1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Pada Persamaan (1), x menyatakan nilai atribut sebelum proses standardisasi, sedangkan μ merupakan nilai rata-rata (mean) dari atribut tersebut. Simbol σ menunjukkan simpangan baku (standard deviation) atribut yang digunakan sebagai ukuran penyebaran data. Hasil proses standardisasi dinyatakan dengan z, yaitu nilai atribut yang telah ditransformasikan menggunakan metode Z-score sehingga memiliki rata-rata mendekati 0 dan simpangan baku mendekati 1. Proses ini bertujuan untuk menyamakan skala antaratribut agar setiap variabel memberikan kontribusi yang seimbang pada proses pemodelan.

Melalui proses preprocessing, seluruh atribut memiliki skala yang seragam sehingga meningkatkan stabilitas proses pelatihan dan kemampuan algoritma Support Vector Machine (SVM) dalam membentuk hyperplane yang optimal. Standardisasi juga mengurangi pengaruh perbedaan rentang nilai antar atribut, mendukung proses optimasi hyperparameter secara lebih efektif, serta meningkatkan kemampuan generalisasi model. Data yang telah diproses kemudian digunakan sebagai masukan pada tahap seleksi fitur menggunakan Recursive Feature Elimination (RFE) untuk memperoleh atribut yang paling relevan bagi proses klasifikasi penyakit jantung secara lebih akurat dan efisien.

2.3 Seleksi Fitur Menggunakan Recursive Feature Elimination (RFE)

Setelah tahap preprocessing, dilakukan seleksi fitur menggunakan Recursive Feature Elimination (RFE) untuk memilih atribut yang paling relevan terhadap klasifikasi penyakit jantung. RFE bekerja secara iteratif dengan menghilangkan fitur yang memiliki kontribusi paling rendah berdasarkan bobot yang dihasilkan oleh Support

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Vector Machine (SVM) dengan kernel linear sebagai estimator. Proses ini bertujuan mengurangi fitur yang kurang relevan sehingga meningkatkan efisiensi dan kemampuan generalisasi model. Hasil proses seleksi fitur beserta status fitur yang dipilih dan tidak dipilih ditunjukkan pada Tabel 2.

Tabel 2. Ranking Hasil Seleksi Fitur Menggunakan RFE

Feature	Ranking	Status
age	1	Dipilih
sex	1	Dipilih
cp	1	Dipilih
thalach	1	Dipilih
exang	1	Dipilih
oldpeak	1	Dipilih
ca	1	Dipilih
thal	1	Dipilih
trestbps	>1	Tidak dipilih
chol	>1	Tidak dipilih
fbs	>1	Tidak dipilih
restecg	>1	Tidak dipilih
slope	>1	Tidak dipilih

Berdasarkan Tabel 2, fitur yang dipilih didominasi oleh atribut yang berkaitan langsung dengan kondisi klinis penyakit jantung, seperti cp, thalach, exang, oldpeak, ca, dan thal. Seleksi fitur berhasil mengurangi atribut yang kurang relevan sehingga menurunkan kompleksitas model dan meningkatkan efisiensi proses optimasi hyperparameter. Fitur-fitur terpilih tersebut selanjutnya digunakan sebagai masukan pada tahap GridSearchCV untuk memperoleh kombinasi parameter terbaik pada algoritma Support Vector Machine (SVM).

2.4 Optimasi Hyperparameter Menggunakan GridSearchCV

Setelah diperoleh fitur terbaik menggunakan Recursive Feature Elimination (RFE), dilakukan optimasi hyperparameter pada algoritma Support Vector Machine (SVM) menggunakan GridSearchCV. Optimasi dilakukan dengan menguji berbagai kombinasi parameter kernel, C, dan gamma menggunakan 5-Fold Cross Validation. Kombinasi parameter yang menghasilkan nilai rata-rata akurasi tertinggi dipilih sebagai model terbaik untuk meningkatkan performa klasifikasi dan kemampuan generalisasi SVM.

Parameter yang dioptimasi meliputi kernel, C, dan gamma karena ketiganya berpengaruh terhadap performa klasifikasi SVM. Proses optimasi dilakukan menggunakan GridSearchCV dengan evaluasi 5-Fold Cross Validation untuk memperoleh kombinasi parameter terbaik berdasarkan nilai Accuracy. Berdasarkan hasil eksperimen, kernel RBF dengan nilai C dan gamma terbaik menghasilkan performa klasifikasi yang lebih optimal dibandingkan kombinasi parameter lainnya, sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Hasil Optimasi Hyperparameter

Parameter	Nilai Terbaik
Kernel	RBF
C	10
Gamma	0.001
Cross Validation Accuracy	83.87%

Berdasarkan tabel 3, GridSearchCV menghasilkan kombinasi parameter terbaik, yaitu kernel RBF, C = 10, dan gamma = 0.001, dengan nilai cross-validation accuracy sebesar 83,87%. Kombinasi parameter tersebut kemudian digunakan untuk membangun model Support Vector Machine (SVM) dan dievaluasi menggunakan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC untuk mengukur performa prediksi risiko penyakit jantung

2.5 Pembangunan Model Support Vector Machine

Model klasifikasi dibangun menggunakan algoritma Support Vector Machine (SVM) karena memiliki kemampuan yang baik dalam menangani klasifikasi biner dan menghasilkan kemampuan generalisasi yang tinggi. Pada penelitian ini, model menggunakan fitur hasil seleksi Recursive Feature Elimination (RFE) serta kombinasi hyperparameter terbaik hasil GridSearchCV, yaitu kernel RBF, C = 10, dan gamma = 0.001. Seluruh proses pelatihan model dilakukan menggunakan pustaka Scikit-learn pada lingkungan Google Colaboratory.

Secara matematis, tujuan utama algoritma Support Vector Machine adalah mencari hyperplane optimal yang mampu memisahkan dua kelas dengan margin terbesar. Persamaan hyperplane dinyatakan sebagai berikut.

$$f(x) = \boldsymbol{w}^T \boldsymbol{x} + b \quad (2)$$

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Pada Persamaan (2), $f(x)$ menyatakan fungsi keputusan (decision function) yang digunakan untuk menentukan kelas suatu data. Simbol **w** merupakan vektor bobot (weight vector) yang menentukan orientasi hyperplane, sedangkan $\boldsymbol{x}$ menyatakan vektor fitur masukan yang merepresentasikan karakteristik setiap data. Adapun b merupakan nilai bias (bias) yang berfungsi menggeser posisi hyperplane sehingga model dapat membentuk batas keputusan (decision boundary) yang optimal dalam memisahkan kedua kelas. Hyperplane optimal diperoleh melalui proses optimasi dengan meminimalkan fungsi objektif sebagai berikut.

$$min\frac{1}{2} \parallel \boldsymbol{w} \parallel^2 + C \sum_{i=1}^{n} \xi_i \quad (3)$$

Dengan kendala

$$y_i(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (4)$$

Pada Persamaan (3) dan Persamaan (4), **C** merupakan parameter regularisasi yang mengendalikan keseimbangan antara memaksimalkan margin dan meminimalkan kesalahan klasifikasi. Simbol ξ_i menyatakan variabel slack yang digunakan untuk mengakomodasi data yang tidak dapat dipisahkan secara sempurna. Selanjutnya, $\mathbf{y}_i$ menunjukkan label kelas dari setiap data pelatihan, sedangkan **n** menyatakan jumlah data pelatihan yang digunakan dalam proses pembelajaran model. Melalui formulasi tersebut, algoritma Support Vector Machine (SVM) berupaya memperoleh hyperplane optimal yang memiliki kemampuan generalisasi yang baik terhadap data baru. Persamaan tersebut menunjukkan bahwa algoritma Support Vector Machine (SVM) berupaya memperoleh margin maksimum dengan meminimalkan kesalahan klasifikasi melalui pengaturan parameter C. Pada penelitian ini digunakan kernel Radial Basis Function (RBF) karena mampu menangani data yang tidak dapat dipisahkan secara linear dengan memetakan data ke ruang berdimensi lebih tinggi, sebagaimana dinyatakan pada Persamaan (5).

$$K(x_i, x_j) = exp\left(\quad -\gamma \parallel x_i - x_j \parallel^2 \right) \quad (5)$$

Pada Persamaan (5), $K(x_i, x_j)$ menyatakan nilai fungsi kernel yang digunakan untuk mengukur tingkat kemiripan antara dua data, yaitu x_i dan x_j, yang masing-masing merepresentasikan pasangan data masukan. Sementara itu, γ (gamma) merupakan parameter kernel yang mengatur tingkat pengaruh setiap data terhadap pembentukan decision boundary. Nilai gamma yang sesuai akan menghasilkan batas keputusan yang mampu memisahkan kedua kelas secara lebih optimal.

Parameter gamma berperan dalam menentukan bentuk batas keputusan (decision boundary). Nilai gamma yang kecil menghasilkan batas keputusan yang lebih halus (smooth), sedangkan nilai gamma yang besar menghasilkan batas keputusan yang lebih kompleks sehingga mampu mengikuti pola data secara lebih rinci. Berdasarkan hasil optimasi menggunakan GridSearchCV, nilai gamma = 0.001 memberikan performa terbaik pada penelitian ini.

2.6 Evaluasi Model

Tahap terakhir adalah mengevaluasi performa model Support Vector Machine (SVM) menggunakan 61 data uji (20%) untuk mengukur kemampuan generalisasi model. Penelitian ini membandingkan tiga model, yaitu Baseline SVM, SVM + RFE, dan SVM + RFE + GridSearchCV. Evaluasi diawali dengan pembentukan Confusion Matrix yang terdiri atas True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), kemudian dilanjutkan dengan perhitungan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC untuk menentukan model dengan performa terbaik.

a. Accuracy

Accuracy menunjukkan persentase prediksi yang benar terhadap seluruh data pengujian. Nilai Accuracy dihitung menggunakan Persamaan (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Semakin tinggi nilai Accuracy menunjukkan bahwa model semakin baik dalam melakukan klasifikasi terhadap seluruh data.

b. Precision

Precision menunjukkan kemampuan model dalam menghasilkan prediksi positif yang benar dari seluruh data yang diprediksi sebagai positif. Persamaan Precision ditunjukkan pada Persamaan (7).

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Nilai Precision yang tinggi menunjukkan bahwa model memiliki tingkat kesalahan False Positive yang rendah.

c. Recall

Recall atau Sensitivity menunjukkan kemampuan model dalam mendeteksi seluruh data positif yang terdapat pada dataset. Persamaan Recall ditunjukkan pada Persamaan (8).

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Semakin tinggi nilai Recall menunjukkan bahwa semakin banyak pasien yang benar-benar berisiko penyakit jantung berhasil dideteksi oleh model.

d. Score

F1-score merupakan rata-rata harmonik antara Precision dan Recall sehingga mampu memberikan gambaran keseimbangan kedua metrik tersebut. Persamaan F1-score ditunjukkan pada Persamaan (9).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

F1-score banyak digunakan ketika distribusi kelas tidak sepenuhnya seimbang karena mempertimbangkan ketepatan maupun sensitivitas model secara bersamaan.

e. Receiver Operating Characteristic (ROC) dan Area Under the Curve (AUC)

Selain menggunakan metrik berbasis Confusion Matrix, penelitian ini juga mengevaluasi kemampuan model menggunakan Receiver Operating Characteristic (ROC) dan Area Under the Curve (AUC).

Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai threshold. Kurva ini menunjukkan kemampuan model dalam membedakan kelas positif dan negatif. Nilai TPR dan FPR dihitung menggunakan Persamaan (10) dan Persamaan (11).

$$TPR = \frac{TP}{TP+FN} \quad (10)$$

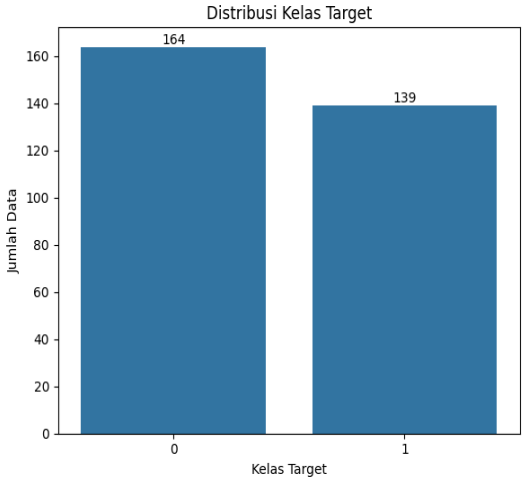
$$FPR = \frac{FP}{FP+TN} \quad (11)$$

Luas area di bawah kurva ROC dinyatakan sebagai Area Under the Curve (AUC). Nilai AUC berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan kemampuan model yang semakin baik dalam membedakan kedua kelas. Sebaliknya, nilai AUC mendekati 0,5 menunjukkan bahwa model memiliki kemampuan klasifikasi yang hampir sama dengan prediksi secara acak

3. HASIL DAN PEMBAHASAN

3.1 Analisis Dataset dan Hasil Preprocessing

Tahap awal penelitian dilakukan melalui Exploratory Data Analysis (EDA) terhadap Cleveland Heart Disease Dataset yang terdiri atas 303 data pasien, 13 atribut prediktor, dan 1 atribut target. Hasil analisis menunjukkan bahwa seluruh atribut bertipe numerik, tidak memiliki missing value maupun data duplikat, sehingga seluruh data dapat digunakan pada proses pelatihan dan pengujian model. Distribusi kelas target pada dataset ditunjukkan pada Gambar 2.



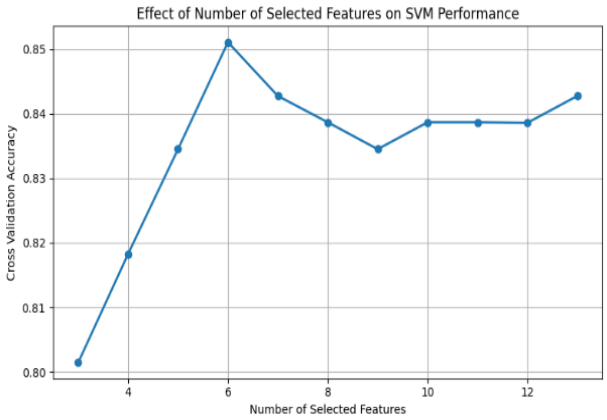
Gambar 2. Distribusi Kelas Target pada Cleveland Heart Disease Dataset

Berdasarkan gambar 2, terdapat 165 data pada kelas 0 dan 138 data pada kelas 1, sehingga dataset memiliki distribusi yang relatif seimbang tanpa memerlukan teknik data balancing. Selanjutnya, data dibagi menjadi 242 data latih (80%) dan 61 data uji (20%) menggunakan train-test split dengan parameter random_state = 42 dan stratify. Setelah itu, dilakukan standardisasi menggunakan StandardScaler untuk menyamakan skala atribut sebelum pelatihan model. Hasil analisis menunjukkan bahwa dataset memiliki kualitas yang baik sehingga layak digunakan pada tahap seleksi fitur menggunakan Recursive Feature Elimination (RFE) dan optimasi hyperparameter menggunakan GridSearchCV.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

3.2 Hasil Seleksi Fitur dan Optimasi Hyperparameter

Setelah tahap preprocessing, dilakukan seleksi fitur menggunakan Recursive Feature Elimination (RFE) untuk memilih atribut yang paling relevan terhadap klasifikasi penyakit jantung. RFE menggunakan Support Vector Machine (SVM) dengan kernel linear sebagai estimator, kemudian mengevaluasi jumlah fitur 3–13 menggunakan 5-Fold Cross Validation untuk memperoleh kombinasi fitur dengan nilai Cross Validation Accuracy terbaik. Hubungan antara jumlah fitur yang dipilih dan nilai Cross Validation Accuracy ditunjukkan pada Gambar 3.



Gambar 3. Pengaruh Jumlah Fitur terhadap Akurasi Cross Validation

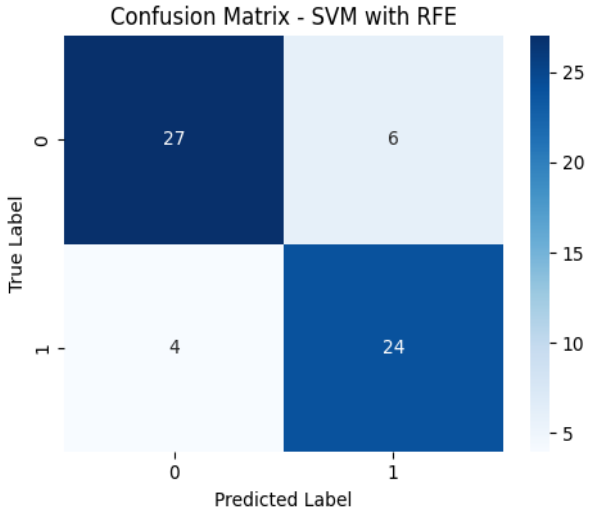
Berdasarkan Gambar 3, model dengan delapan fitur dipilih karena memberikan kemampuan generalisasi yang lebih baik, meskipun nilai Cross Validation Accuracy tertinggi (85,11%) diperoleh pada enam fitur. Selanjutnya, fitur terpilih digunakan dalam proses optimasi hyperparameter menggunakan GridSearchCV.

3.3 Hasil Pengujian

Tahap pengujian dilakukan untuk mengevaluasi kemampuan model dalam mengklasifikasikan pasien ke dalam kategori berisiko dan tidak berisiko penyakit jantung. Pengujian dilakukan menggunakan 20% data uji yang tidak pernah digunakan selama proses pelatihan model. Evaluasi dilakukan terhadap tiga model, yaitu Baseline Support Vector Machine (SVM), SVM dengan Recursive Feature Elimination (RFE), serta SVM dengan kombinasi RFE dan GridSearchCV sebagai model usulan.

3.3.1 Hasil Pengujian Baseline SVM

Pengujian pertama dilakukan menggunakan algoritma **Support Vector Machine** dengan parameter bawaan (default parameter) tanpa proses seleksi fitur maupun optimasi hyperparameter. Model ini digunakan sebagai **baseline** untuk mengetahui performa awal algoritma SVM sebelum dilakukan pengembangan metode. Hasil evaluasi performa model baseline ditunjukkan pada Gambar 4.

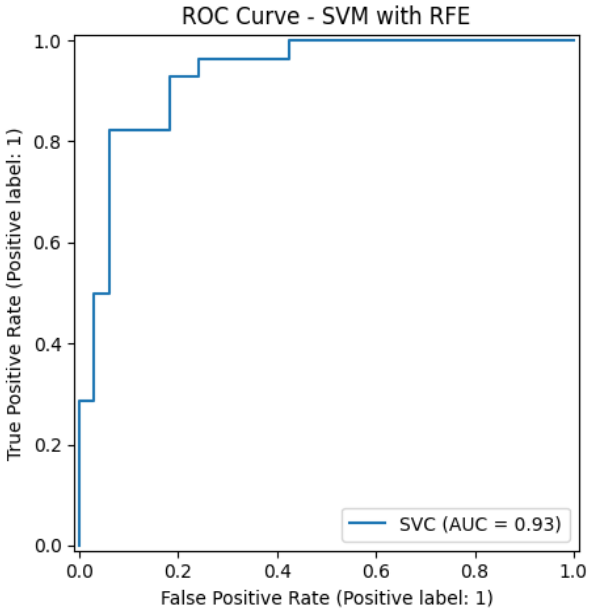


Gambar 4. Confusion Matrix Baseline Support Vector Machine

Berdasarkan Gambar 4, Confusion Matrix menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar, meskipun masih terdapat beberapa kesalahan klasifikasi (False Positive dan False Negative). Hal ini menunjukkan bahwa penggunaan parameter bawaan belum menghasilkan decision boundary

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

yang optimal. Selanjutnya, kemampuan diskriminasi model baseline dievaluasi menggunakan kurva ROC dan nilai ROC-AUC, sebagaimana ditunjukkan pada Gambar 5.

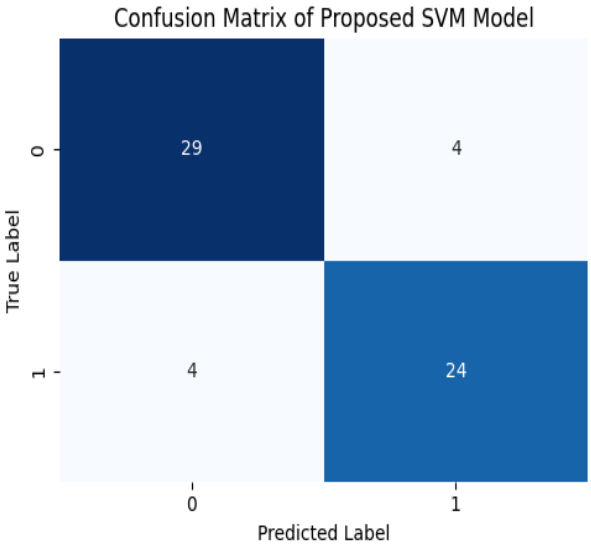


Gambar 5. ROC Curve Baseline Support Vector Machine

Berdasarkan Gambar 5, model baseline memperoleh nilai ROC-AUC sebesar 94,70%, yang menunjukkan kemampuan klasifikasi yang baik, meskipun masih berpotensi ditingkatkan melalui seleksi fitur dan optimasi hyperparameter.

3.3.2 Hasil Pengujian Model Usulan

Tahap berikutnya adalah pengujian model usulan yang mengombinasikan Recursive Feature Elimination (RFE) dan GridSearchCV. Seleksi fitur digunakan untuk mengurangi atribut yang kurang relevan, sedangkan GridSearchCV digunakan untuk memperoleh kombinasi parameter terbaik, yaitu kernel RBF, C = 10, dan gamma = 0.001. Hasil evaluasi performa model usulan menggunakan kurva ROC disajikan pada Gambar 6.



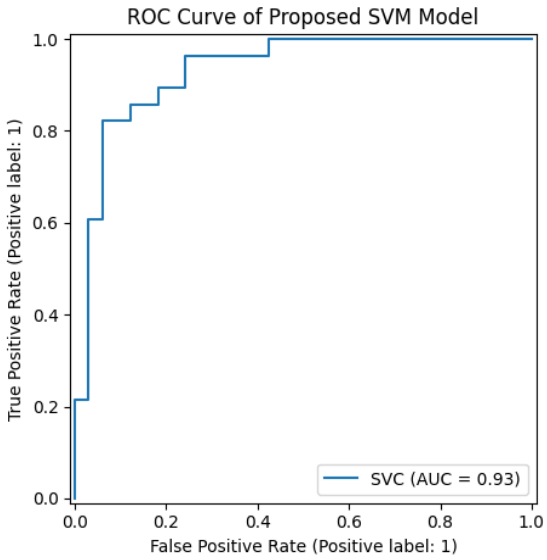
Gambar 6. Confusion Matrix Model Usulan

Berdasarkan gambar 6, jumlah prediksi yang benar meningkat dibandingkan model baseline. Hal tersebut menunjukkan bahwa kombinasi seleksi fitur dan optimasi hyperparameter mampu menghasilkan batas keputusan yang lebih baik sehingga mengurangi jumlah kesalahan klasifikasi.

Tahap berikutnya adalah evaluasi model usulan yang mengombinasikan Recursive Feature Elimination (RFE) dan GridSearchCV. Seleksi fitur dan optimasi hyperparameter diterapkan untuk meningkatkan kemampuan

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

klasifikasi model Support Vector Machine (SVM). Kurva ROC dan nilai ROC-AUC hasil evaluasi model usulan ditunjukkan pada Gambar 7.



Gambar 7. ROC Curve Model Usulan

Berdasarkan gambar 7, Kurva ROC menunjukkan bahwa model usulan memiliki kemampuan yang sangat baik dalam membedakan kedua kelas. Nilai ROC-AUC sebesar 93,18% menunjukkan bahwa model memiliki tingkat diskriminasi yang tinggi terhadap pasien yang berisiko maupun tidak berisiko penyakit jantung.

Setelah proses pelatihan dan pengujian selesai, kinerja model usulan dievaluasi menggunakan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil evaluasi model usulan disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model Usulan

Metrik	Nilai
Accuracy	86,89%
Precision	85,71%
Recall	85,71%
F1-score	85,71%
ROC-AUC	93,18%

Berdasarkan Tabel 4, model usulan memperoleh nilai Accuracy sebesar 86,89%, yang menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar. Nilai Precision sebesar 85,71% menunjukkan bahwa prediksi pasien yang dikategorikan berisiko penyakit jantung memiliki tingkat ketepatan yang tinggi. Sementara itu, nilai Recall sebesar 85,71% menunjukkan bahwa model mampu mendeteksi sebagian besar pasien yang benar-benar berisiko penyakit jantung. Nilai F1-score sebesar 85,71% mengindikasikan keseimbangan yang baik antara Precision dan Recall. Selain itu, nilai ROC-AUC sebesar 93,18% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan kedua kelas.

3.3.3 Perbandingan Performa Model

Untuk mengetahui pengaruh seleksi fitur dan optimasi hyperparameter, dilakukan perbandingan performa ketiga model yang digunakan dalam penelitian ini Untuk mengetahui pengaruh seleksi fitur dan optimasi hyperparameter, dilakukan perbandingan performa ketiga model yang digunakan dalam penelitian ini berdasarkan metrik Accuracy, Precision, Recall, F1-score, dan ROC-AUC. Hasil perbandingan tersebut disajikan pada Tabel 5.

Tabel 5. Perbandingan Performa Model

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Baseline SVM	85.25	80.65	89.29	84.75	94.70
SVM+RFE	83.61	80.00	85.71	82.76	93.07
SVM+RFE+GridSearchCV	86.89	85.71	85.71	85.71	93.18

Berdasarkan Tabel 5, model Baseline SVM memperoleh nilai Accuracy sebesar 85,25%, sedangkan model SVM dengan RFE menghasilkan Accuracy sebesar 83,61%. Setelah dilakukan optimasi hyperparameter menggunakan GridSearchCV, model usulan menghasilkan Accuracy sebesar 86,89%, yang merupakan nilai tertinggi dibandingkan model lainnya.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Selain meningkatkan Accuracy, model usulan juga menghasilkan nilai Precision dan F1-score yang lebih tinggi dibandingkan model baseline. Peningkatan tersebut menunjukkan bahwa kombinasi Recursive Feature Elimination dan GridSearchCV mampu meningkatkan kualitas klasifikasi dengan menghasilkan model yang lebih akurat dan lebih stabil.

Walaupun model baseline memiliki nilai ROC-AUC sedikit lebih tinggi, yaitu 94,70%, dibandingkan model usulan sebesar 93,18%, selisih tersebut relatif kecil. Sebaliknya, model usulan memberikan keseimbangan yang lebih baik antara Accuracy, Precision, Recall, dan F1-score, sehingga lebih sesuai digunakan sebagai model prediksi risiko penyakit jantung.

Secara keseluruhan, hasil penelitian menunjukkan bahwa penerapan seleksi fitur menggunakan Recursive Feature Elimination dan optimasi hyperparameter menggunakan GridSearchCV mampu meningkatkan performa algoritma Support Vector Machine dalam memprediksi risiko penyakit jantung. Kombinasi kedua metode tersebut menghasilkan model yang lebih optimal dibandingkan penggunaan algoritma SVM tanpa proses optimasi.

3.4 Perbandingan dengan Penelitian Terdahulu

Untuk mengevaluasi efektivitas model yang diusulkan, dilakukan analisis komparatif dengan beberapa penelitian terdahulu yang menggunakan algoritma machine learning pada prediksi penyakit jantung. Perbandingan dilakukan berdasarkan algoritma yang digunakan, metode optimasi, dataset, dan nilai accuracy yang diperoleh. Perbandingan hasil penelitian tersebut disajikan pada Tabel 6..

Tabel 6. Perbandingan Hasil Penelitian dengan Penelitian Terdahulu

Penelitian	Metode Terbaik	Dataset	Accuracy
Hidayat et al. [24]	SVM	Cleveland Heart Disease	85.00%
Abidin et al. [25]	SVM	Heart Disease	77.00%
Pratama & Putra [26]	Hybrid PSO–GWO + SVM	Heart Disease	91.38%
Bouqentar et al. [27]	Feature Engineering + Random Forest	Cleveland Heart Disease	90.16%
Nasution et al. [28]	SVM (Best among compared models)	UCI Heart Disease	87.00%
Penelitian ini	SVM + RFE + GridSearchCV	Cleveland Heart Disease	86.89%

Berdasarkan Tabel 6, model yang diusulkan memperoleh akurasi sebesar 86,89%, lebih tinggi dibandingkan penelitian Hidayat et al. [16] yang memperoleh akurasi 85,00% serta Abidin et al. [17] sebesar 77,00%. Namun demikian, penelitian Bouqentar et al. [20] dan Pratama & Putra [19] melaporkan akurasi yang lebih tinggi, yaitu masing-masing sebesar 90,16% dan 91,38%, sedangkan Nasution et al. [21] memperoleh akurasi 87,00% menggunakan algoritma SVM. Berdasarkan Tabel 6, model yang diusulkan memperoleh akurasi sebesar 86,89%, lebih tinggi dibandingkan penelitian Hidayat et al. [16] sebesar 85,00% dan Abidin et al. [17] sebesar 77,00%. Peningkatan tersebut menunjukkan bahwa kombinasi seleksi fitur menggunakan RFE dan optimasi hyperparameter melalui GridSearchCV mampu meningkatkan kemampuan klasifikasi dibandingkan penggunaan algoritma SVM tanpa optimasi tambahan.

Berbeda dengan penelitian terdahulu yang hanya menerapkan SVM atau hanya menggunakan satu tahap optimasi, penelitian ini mengintegrasikan Recursive Feature Elimination (RFE) sebagai seleksi fitur dan GridSearchCV sebagai optimasi hyperparameter dalam satu kerangka pemodelan. Pendekatan tersebut menghasilkan model yang lebih efisien karena hanya menggunakan fitur-fitur yang relevan sekaligus memperoleh kombinasi parameter optimal (kernel RBF, C = 10, gamma = 0.001).

Walaupun beberapa pendekatan berbasis metaheuristik dan ensemble menghasilkan akurasi yang lebih tinggi, metode yang diusulkan menawarkan keseimbangan yang baik antara performa klasifikasi, kompleksitas komputasi, dan kemudahan reproduksi. Integrasi RFE dan GridSearchCV mampu meningkatkan performa SVM tanpa memerlukan algoritma optimasi yang kompleks, sehingga berpotensi diterapkan pada sistem pendukung keputusan untuk prediksi risiko penyakit jantung.

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa penerapan Recursive Feature Elimination (RFE) dan GridSearchCV memberikan pengaruh positif terhadap performa algoritma Support Vector Machine (SVM) dalam memprediksi risiko penyakit jantung. Seleksi fitur menggunakan RFE berhasil mengurangi atribut yang kurang relevan sehingga model hanya memanfaatkan fitur yang memiliki kontribusi terbesar terhadap proses klasifikasi. Pengurangan jumlah fitur tersebut tidak hanya menurunkan kompleksitas model, tetapi juga meningkatkan efisiensi proses pelatihan dan optimasi.

Hasil penelitian menunjukkan bahwa penerapan kombinasi Recursive Feature Elimination (RFE) dan GridSearchCV pada algoritma Support Vector Machine (SVM) mampu meningkatkan kinerja klasifikasi dibandingkan penggunaan SVM tanpa optimasi. Model usulan memperoleh Accuracy sebesar 86,89%, lebih tinggi dibandingkan model Baseline SVM yang memperoleh 85,25%. Peningkatan tersebut menunjukkan bahwa proses seleksi fitur berhasil menghilangkan atribut yang kurang relevan, sedangkan optimasi hyperparameter mampu menghasilkan konfigurasi model yang lebih sesuai dengan karakteristik Cleveland Heart Disease Dataset.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

Jika dibandingkan dengan penelitian terdahulu, hasil penelitian ini sejalan dengan temuan Hidayat et al. [24] dan Abidin et al. [25], yang menunjukkan bahwa algoritma SVM memiliki kemampuan yang baik dalam klasifikasi penyakit jantung. Namun, penelitian ini memberikan peningkatan performa melalui integrasi RFE dan GridSearchCV, sehingga menghasilkan akurasi yang lebih tinggi dibandingkan kedua penelitian tersebut. Hal ini mengindikasikan bahwa penerapan seleksi fitur dan optimasi hyperparameter secara bersamaan lebih efektif dibandingkan penggunaan SVM tanpa proses optimasi.

Di sisi lain, akurasi model usulan masih berada di bawah penelitian Pratama dan Putra [26] yang menggunakan pendekatan Hybrid PSO–GWO + SVM serta Bouqentar et al. [27] yang menerapkan Feature Engineering dan Random Forest. Perbedaan tersebut dapat dipengaruhi oleh penggunaan metode optimasi yang lebih kompleks, teknik rekayasa fitur yang berbeda, maupun karakteristik data yang digunakan. Meskipun demikian, model yang diusulkan menawarkan pendekatan yang relatif sederhana dan mudah direproduksi karena hanya memanfaatkan RFE dan GridSearchCV yang tersedia pada pustaka scikit-learn, sehingga implementasinya lebih praktis tanpa memerlukan algoritma optimasi berbasis metaheuristik.

Temuan penelitian ini juga memperkuat hasil penelitian sebelumnya yang menyatakan bahwa kualitas fitur dan pemilihan hyperparameter merupakan faktor penting yang memengaruhi performa algoritma SVM. Dengan demikian, kombinasi RFE dan GridSearchCV dapat menjadi alternatif yang efektif untuk meningkatkan kemampuan klasifikasi penyakit jantung, terutama pada dataset berukuran kecil hingga menengah seperti Cleveland Heart Disease Dataset. Meskipun demikian, penelitian ini masih memiliki keterbatasan karena hanya menggunakan satu dataset dan satu algoritma klasifikasi. Penelitian selanjutnya dapat mengevaluasi pendekatan yang diusulkan pada dataset lain atau mengombinasikannya dengan metode optimasi yang lebih adaptif untuk memperoleh performa yang lebih tinggi.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi risiko penyakit jantung menggunakan algoritma Support Vector Machine (SVM) yang dipadukan dengan Recursive Feature Elimination (RFE) dan GridSearchCV. Penerapan RFE mampu mengidentifikasi atribut-atribut yang paling relevan sehingga mengurangi kompleksitas model dengan menghilangkan fitur yang kurang berkontribusi terhadap proses klasifikasi. Selanjutnya, optimasi hyperparameter menggunakan GridSearchCV berhasil memperoleh kombinasi parameter terbaik, yaitu kernel Radial Basis Function (RBF), C = 10, dan gamma = 0.001, yang menghasilkan performa lebih baik dibandingkan penggunaan parameter bawaan. Berdasarkan hasil pengujian menggunakan Cleveland Heart Disease Dataset, model usulan memperoleh Accuracy sebesar 86,89%, Precision 85,71%, Recall 85,71%, F1-score 85,71%, dan ROC-AUC 93,18%. Hasil tersebut menunjukkan bahwa integrasi seleksi fitur dan optimasi hyperparameter mampu meningkatkan kualitas klasifikasi serta menghasilkan keseimbangan yang baik antara ketepatan prediksi dan kemampuan mendeteksi pasien yang berisiko penyakit jantung. Dibandingkan model SVM tanpa optimasi, pendekatan yang diusulkan menghasilkan model yang lebih efisien, memiliki kemampuan generalisasi yang baik, serta mempertahankan performa klasifikasi dengan jumlah fitur yang lebih sedikit. Kontribusi utama penelitian ini adalah penerapan kerangka terintegrasi antara Recursive Feature Elimination (RFE) dan GridSearchCV pada algoritma SVM untuk meningkatkan performa prediksi risiko penyakit jantung. Pendekatan ini relatif sederhana, mudah direproduksi menggunakan pustaka Scikit-learn, dan berpotensi diterapkan sebagai komponen pendukung sistem diagnosis penyakit jantung berbasis machine learning. Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah sampel yang lebih besar dan lebih beragam, membandingkan metode optimasi lain seperti Bayesian Optimization atau Particle Swarm Optimization (PSO), serta mengevaluasi algoritma klasifikasi lain, seperti Extreme Gradient Boosting (XGBoost) dan LightGBM, guna memperoleh model prediksi yang lebih akurat, robust, dan memiliki kemampuan generalisasi yang lebih baik.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa atas dukungan yang diberikan selama pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada UCI Machine Learning Repository yang menyediakan Cleveland Heart Disease Dataset sebagai sumber data penelitian sehingga penelitian ini dapat dilaksanakan dengan baik.

REFERENCES

[1] R. Waluyo and A. S. Munir, “Optimasi Prediksi Kematian pada Gagal Jantung Analisis Perbandingan Algoritma Pembelajaran Ensemble dan Teknik Penyeimbangan Data pada Dataset,” Jurnal Sistem dan Teknologi Informasi (JustIN), vol. 12, no. 2, p. 365, Apr. 2024, doi: 10.26418/justin.v12i2.75158.

[2] A. S. Verdiana, M. Cerinda, R. Mansur, and F. Y. Cahyani, “Provincial Analysis of Years of Life Lost and Economic Impact of Heart Disease Among the Working-Age Population in Indonesia (2018-2023),” Indonesian Health Journal, vol. 5, no. 1, pp. 141–147, Mar. 2026, doi: 10.58344/ihj.v5i1.829.

[3] P. Joseph et al., “Cardiovascular disease in the Americas: the epidemiology of cardiovascular disease and its risk factors,” The Lancet Regional Health - Americas, vol. 42, p. 100960, Feb. 2025, doi: 10.1016/j.lana.2024.100960.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 1374−1385
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10518

[4] Y. Luo, J. Liu, J. Zeng, and H. Pan, "Global burden of cardiovascular diseases attributed to low physical activity: An analysis of 204 countries and territories between 1990 and 2019," Am. J. Prev. Cardiol., vol. 17, p. 100633, Mar. 2024, doi: 10.1016/j.ajpc.2024.100633.
[5] G. A. Mensah et al., "Global Burden of Cardiovascular Diseases and Risks, 1990-2022," J. Am. Coll. Cardiol., vol. 82, no. 25, pp. 2350–2473, Dec. 2023, doi: 10.1016/j.jacc.2023.11.007.
[6] T. S. Santi, J. E. Nelwan, and F. L. F. G. Langi, "Gambaran Faktor Risiko Kejadian Penyakit Jantung Koroner di Poliklinik Jantung Cardio Vascular and Brain Center Rumah Sakit Umum Pusat Prof. Dr. R. D. Kandou Manado," Jurnal Lentera: Penelitian dan Pengabdian Masyarakat, vol. 3, no. 2, pp. 79–84, Dec. 2022, doi: 10.57207/a6maa684.
[7] Erni Kurniasih, W. Jumaiyah, D. Purnamawati, Y. Sofiani, and E. Erwin, "Determinan Self Care Pasien Penyakit Jantung Koroner Setelah Intervensi Koroner Perkutan," Jurnal Ners, vol. 9, no. 3, pp. 4822–4826, Jul. 2025, doi: 10.31004/jn.v9i3.47105.
[8] A. Chakraborty, U. K. Das, S. Sazzad, P. Das, and Md. M. H. Khan, "Explainable machine learning for early heart disease risk prediction: Insights from a clinical dataset in Bangladesh," Intell. Based. Med., vol. 13, p. 100352, Mar. 2026, doi: 10.1016/j.ibmed.2026.100352.
[9] D. Sylvester Aondonenge et al., "Early Heart Disease Prediction Using Data Mining Techniques," Vokasi Unesa Bulletin of Engineering, Technology and Applied Science, vol. 2, no. 2, pp. 211–226, Jun. 2025, doi: 10.26740/vubeta.v2i2.36735.
[10] S. Samant, A. N. Panagopoulos, W. Wu, S. Zhao, and Y. S. Chatzizisis, "Artificial Intelligence in Coronary Artery Interventions: Preprocedural Planning and Procedural Assistance," Journal of the Society for Cardiovascular Angiography & Interventions, vol. 4, no. 3, p. 102519, Mar. 2025, doi: 10.1016/j.jscai.2024.102519.
[11] J. Zhang et al., "Artificial intelligence applied in cardiovascular disease: a bibliometric and visual analysis," Front. Cardiovasc. Med., vol. 11, Feb. 2024, doi: 10.3389/fcvm.2024.1323918.
[12] A. R. Raharja, Jayadi, A. Pramudianto, and Y. Muchsam, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang," Technologia Journal, vol. 1, no. 1, Feb. 2024, doi: 10.62872/cwgzp962.
[13] I. Arfyanti, T. Bustomi, and I. Haristyawan, "Perbandingan Kinerja Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Darah Tinggi," Building of Informatics, Technology and Science (BITS), vol. 6, no. 3, pp. 1987–1994, Dec. 2024, doi: 10.47065/bits.v6i3.6477.
[14] L. Nafi'ah and Z. Fatah, "Implementasi Algoritma Decision Tree Untuk Pendeteksian Penyakit Jantung," JUSIFOR : Jurnal Sistem Informasi dan Informatika, vol. 3, no. 2, pp. 160–165, Dec. 2024, doi: 10.70609/jusifor.v3i2.5729.
[15] R. Reátegui, C. Tandazo-Malla, R. Suárez, and L. Ramírez-Cerna, "Cardiovascular risk prediction via ensemble machine learning and oversampling methods," Sci. Rep., vol. 15, no. 1, p. 43576, Dec. 2025, doi: 10.1038/s41598-025-30895-5.
[16] T. H. Tanjung and M. Furqan, "Classification of Heart Disease Using Support Vector Machine," sinkron, vol. 8, no. 3, pp. 1803–1812, Jul. 2024, doi: 10.33395/sinkron.v8i3.13904.
[17] M. Fajri and A. Primajaya, "Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search," Journal of Applied Informatics and Computing, vol. 7, no. 1, pp. 14–19, Jan. 2023, doi: 10.30871/jaic.v7i1.5004.
[18] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," Information, vol. 15, no. 4, p. 235, Apr. 2024, doi: 10.3390/info15040235.
[19] V. Baviskar, M. Verma, P. Chatterjee, and G. Singal, "Efficient Heart Disease Prediction Using Hybrid Deep Learning Classification Models," IRBM, vol. 44, no. 5, p. 100786, Oct. 2023, doi: 10.1016/j.irbm.2023.100786.
[20] A. M. Priyatno and T. Widiyaningtyas, "A Systematic Literature Review: Recursive Feature Elimination Algorithms," JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), vol. 9, no. 2, pp. 196–207, Feb. 2024, doi: 10.33480/jitk.v9i2.5015.
[21] M. A. K. Raiaan et al., "A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks," Decision Analytics Journal, vol. 11, p. 100470, Jun. 2024, doi: 10.1016/j.dajour.2024.100470.
[22] Y.-W. Chen and C.-J. Lin, "Combining SVMs with Various Feature Selection Strategies," in Feature Extraction, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 315–324. doi: 10.1007/978-3-540-35488-8_13.
[23] A. Lakshmanarao, A. Srisaila, and T. S. R. Kiran, "Heart Disease Prediction using Feature Selection and Ensemble Learning Techniques," in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Feb. 2021, pp. 994–998. doi: 10.1109/ICICV50876.2021.9388482.
[24] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, H. T. Saputra, and F. Okmayura, "Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine," BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer, vol. 5, no. 2, pp. 161–168, Sep. 2024, doi: 10.37148/bios.v5i2.152.
[25] M. Abidin et al., "Classification of Heart (Cardiovascular) Disease using the SVM Method," Indonesian Journal of Modern Science and Technology, vol. 1, no. 1, pp. 9–15, Jan. 2025, doi: 10.64021/ijmst.1.1.9-15.2025.
[26] L. I. A. Pratama and A. T. Putra, "Optimizing Heart Disease Classification Using the Support Vector Machine Algorithm with Hybrid Particle Swarm and Grey Wolf Optimization," Recursive Journal of Informatics, vol. 3, no. 1, pp. 26–33, Mar. 2025, doi: 10.15294/rji.v3i1.737.
[27] M. A. Bouqentar et al., "Early heart disease prediction using feature engineering and machine learning algorithms," Heliyon, vol. 10, no. 19, p. e38731, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38731.
[28] N. Nasution, M. A. Hasan, and F. Bakri Nasution, "Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset," IT Journal Research and Development, vol. 9, no. 2, pp. 140–150, Apr. 2025, doi: 10.25299/itjrd.2025.17941.