Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 5 Iss. 4 October 2025, pp: 1531-1537
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Comparative Analysis of Linear, Polynomial, RBF, and Sigmoid Kernels in Support Vector Machine for Heart Disease Classification

Analisis Komparatif *Kernel Linear, Polynomial, RBF*, dan *Sigmoid* pada *Support Vector Machine* untuk Klasifikasi Penyakit Jantung

Adeline Faradisia[1*], Magdalena A. Ineke Pakereng[2]

[1,2]Universitas Kristen Satya Wacana

E-Mail: [1]672022027@student.uksw.edu, [2]ineke.pakereng@uksw.edu

Received Sep 25th 2025; Revised Oct 04th 2025; Accepted Oct 30th 2025; Available Online Nov 05th 2025
Corresponding Author: Adeline Faradisia

Abstract

Heart disease continues to be a major cause of mortality worldwide, highlighting the importance of early detection to minimize health risks. This research employs the Support Vector Machine (SVM) algorithm to classify heart disease risk by evaluating the performance of four kernel functions linear, polynomial, radial basis function (RBF), and sigmoid. The dataset used in this study is the open "Heart Disease" dataset from Kaggle, consisting of 303 patient records with 13 clinical attributes, age, sex, chest pain type (cp), resting blood pressure (trestbps), serum cholesterol (chol), fasting blood sugar (fbs), resting ECG (restecg), maximum heart rate (thalach), exercise-induced angina (exang), ST depression (oldpeak), slope, number of major vessels (ca), and thal, and one binary target indicating the presence of heart disease. The research workflow comprises dataset loading and exploratory analysis, followed by data preprocessing, SVM initialization, iteration over kernels, model training, test-set prediction, and performance evaluation. All kernels were consistently tuned using GridSearchCV with k-fold cross-validation to identify optimal hyperparameter configurations. The evaluation results reveal that the polynomial kernel delivers the best outcomes, with an accuracy of 88.52% and an F1-score of 89%. These results demonstrate that the polynomial kernel is the most effective option for heart disease classification with the SVM approach.

Keyword: Classification, GridSearchCV, Heart Disease, Support Vector Machine, SVM Kernel

Abstrak

Penyakit jantung masih menjadi salah satu penyebab utama kematian di berbagai belahan dunia, sehingga deteksi dini sangat diperlukan untuk menekan risiko yang mungkin timbul. Penelitian ini menerapkan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan risiko penyakit jantung dengan melakukan perbandingan kinerja empat jenis kernel, yaitu *linear*, *polynomial*, *radial basis function* (RBF), dan *sigmoid*. *Dataset* yang digunakan berasal dari *open dataset "Heart Disease"* di Kaggle yang berisi 303 data pasien dengan 13 atribut klinis, *age*, *sex*, jenis nyeri dada (*cp*), tekanan darah saat istirahat (*trestbps*), kolesterol serum *(chol*), gula darah puasa (*fbs*), hasil EKG istirahat (*restecg*), detak jantung maksimum (*thalach*), *angina* akibat olahra*ga (exang*), depresi ST (*oldpeak*), kemiringan segmen ST (*slope*), jumlah pembuluh darah utama (ca), dan *thal*, serta satu target biner yang menunjukkan ada/tidaknya penyakit jantung. Proses penelitian meliputi pemuatan *dataset* dan eksplorasi awal, dilanjutkan pra-pemrosesan data, inisialisasi SVM, iterasi kernel, pelatihan model, prediksi pada data uji, serta evaluasi performa. Seluruh kernel dituning secara konsisten menggunakan *GridSearchCV* guna memperoleh konfigurasi optimal. Hasil evaluasi menunjukkan bahwa kernel *polynomial* memberikan performa terbaik dengan akurasi 88,52% dan F1-*score* 89%. Dengan demikian, kernel *polynomial* dinilai sebagai pilihan paling optimal untuk klasifikasi penyakit jantung menggunakan metode SVM.

Kata Kunci: *GridSearchCV*, Kernel SVM, Klasifikasi, Penyakit Jantung, *Support Vector Machine*

1. PENDAHULUAN

Hipertensi adalah salah satu faktor risiko utama yang berperan besar terhadap timbulnya berbagai penyakit kardiovaskular serius, seperti stroke, gagal jantung, infark miokard, fibrilasi atrium, penyakit arteri perifer, hingga diseksi aorta [1]. Kondisi ini didefinisikan sebagai gangguan kronis dengan tekanan darah sistolik ≥140 mmHg atau diastolik ≥90 mmHg [2][3][4]. Hipertensi ditandai oleh disfungsi pembuluh darah

yang menghambat suplai oksigen serta nutrisi ke jaringan tubuh. Bahayanya, hipertensi kerap muncul secara tiba-tiba tanpa gejala awal [3]. Berdasarkan penyebabnya, hipertensi terbagi menjadi dua kategori, yaitu hipertensi primer yang mencakup 80–95% kasus tanpa penyebab pasti, dan hipertensi sekunder yang timbul akibat penyakit lain seperti stenosis arteri renalis, gangguan parenkim ginjal, feokromositoma, atau hiperaldosteronisme [4]. Faktor-faktor yang memengaruhi hipertensi antara lain keturunan, usia, jenis kelamin, pola makan, gaya hidup, kebiasaan merokok, konsumsi alkohol, stres, serta faktor lain yang mendukung [5].

Risiko penyakit jantung umumnya dipicu oleh hipertensi, kadar kolesterol tinggi, *diabetes mellitus*, obesitas, dan pola hidup tidak sehat [3]. Dalam upaya pencegahan maupun deteksi dini, perkembangan teknologi informasi, khususnya kecerdasan buatan (*artificial intelligence*), memiliki peranan penting dalam peningkatan layanan kesehatan. *Machine learning* menjadi salah satu pendekatan yang semakin banyak dimanfaatkan untuk klasifikasi penyakit jantung karena kemampuannya dalam mengolah data medis guna memprediksi dan mengidentifikasi risiko secara lebih tepat [4]. Dengan teknik ini, dapat ditemukan pola dan informasi signifikan yang mendukung pengambilan keputusan medis. Melalui penerapannya, sistem kesehatan diharapkan mampu menyediakan layanan yang lebih efektif, efisien, dan berbasis data akurat [5].

Salah satu metode *machine learning* yang banyak digunakan dalam klasifikasi data medis adalah *Support Vector Machine* (SVM). Algoritma ini membangun sebuah *hyperplane* yang paling optimal untuk memisahkan data menjadi dua kelas, dengan tujuan memaksimalkan margin antar kelas sehingga menghasilkan prediksi yang lebih akurat [6]. Prinsip kerja SVM berlandaskan struktural *risk minimization* serta memerlukan data latih (*training data*) untuk membangun model yang kemudian digunakan dalam meprediksi data uji (*testing data*) [7]. Evaluasi kinerja dilakukan dengan menganalisis sejumlah parameter, seperti jenis kernel, nilai C, dan *gamma*, guna memperoleh konfigurasi yang paling efektif. Kernel dalam SVM berfungsi memetakan data ke ruang berdimensi lebih tinggi agar pemisahan data dapat dilakukan secara optimal, terutama pada data yang tidak terpisah secara *linear*. Beberapa kernel yang umum dipakai meliputi *linear*, *polynomial*, *radial basis function* (RBF), dan *sigmoid*. Masing-masing kernel memiliki pendekatan serta karakteristik berbeda yang memengaruhi hasil klasifikasi [8].

Algoritma SVM telah banyak dimanfaatkan dalam klasifikasi data medis, terutama untuk deteksi penyakit jantung maupun gangguan kesehatan lainnya. Misalnya, penelitian oleh Dharmawan (2021) yang mengombinasikan SVM dengan *Particle Swarm Optimization* (PSO) berhasil mencapai akurasi sebesar 84,81% pada klasifikasi penyakit jantung [9]. Studi lain yang dilakukan oleh Natsir dkk. (2024) menunjukkan tingkat akurasi 98,44%, dengan keberhasilan mengklasifikasikan 126 dari 128 sampel data menggunakan metode SVM [10]. Sementara itu, Alexsander dkk. (2024) menggunakan algoritma SVM dalam prediksi penyakit jantung dengan skema pembagian data latih dan uji 80:20, menghasilkan akurasi sebesar 95% [11].

Selain penyakit jantung, SVM juga diterapkan pada deteksi penyakit lain. Pada klasifikasi *diabetes mellitus*, akurasi yang diperoleh mencapai 91,2% [12]. Dalam penelitian terkait klasifikasi penyakit kulit berbasis ekstraksi fitur ABCD, yaitu metode pengambilan ciri citra yang melihat keasimetrisan (*asymmetry*), ketegasan batas (*border*), keragaman warna (*color*), dan ukuran/diameter (*diameter*) lesi, ekstraksi tersebut digunakan untuk menghasilkan fitur citra yang lebih representatif sebelum dimasukkan ke SVM. Algoritma ini mencatat akurasi 86,42%, sensitivitas 86,42%, dan spesifisitas 96,60% [13]. Pada penelitian terkait klasifikasi stroke otak, Handayani dan Taufiq (2024) terlebih dahulu menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan data menjadi 4.733 *record* dengan skema pembagian data 80:20, lalu membandingkan *Logistic Regression* (LR), *Random Forest* (RF), SVM, dan *K-Nearest Neighbors* (KNN). Hasil pengujian menunjukkan bahwa LR, RF, dan SVM memperoleh akurasi yang sama, yaitu 95% dengan *recall* 100%, sedangkan KNN berada sedikit lebih rendah dengan akurasi 90% [14]. Dalam klasifikasi kanker payudara, kernel *linear* pada SVM memberikan hasil lebih baik dibanding LR, yaitu akurasi 96% dengan percentage split dan 98% menggunakan *K-fold Cross Validation* [15]. Pada klasifikasi kasus stunting balita, Jalil dkk. (2024) menggunakan kernel linear dengan hasil akurasi 82%, presisi 80%, dan *recall* 86% [16]. Sedangkan untuk klasifikasi kelenjar tiroid, kernel RBF pada SVM menghasilkan akurasi 97%, lebih tinggi dibandingkan KNN (92%) dan *Decision Tree* (91%) [17].

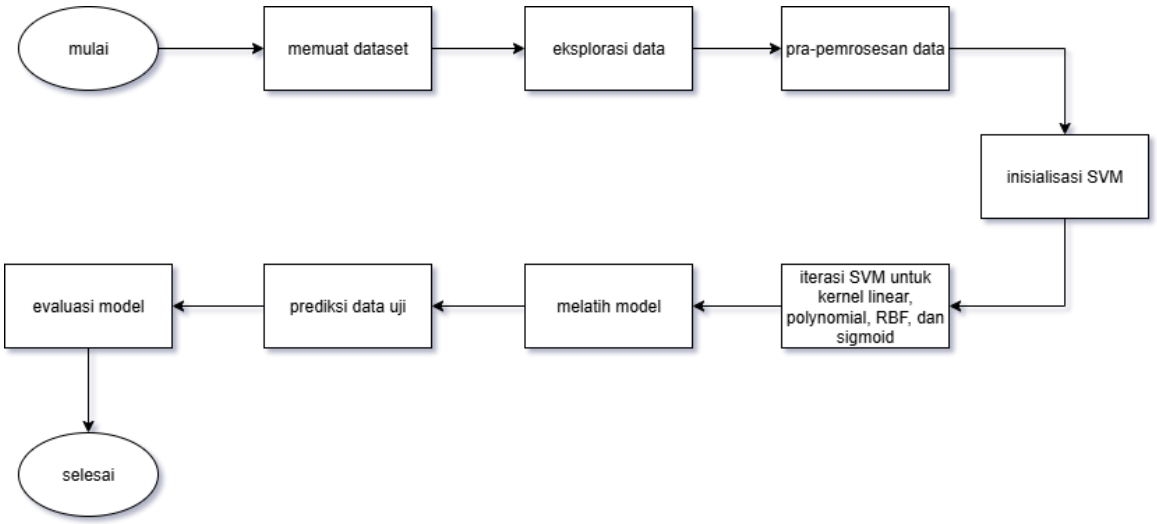
Beberapa penelitian secara khusus membandingkan kinerja kernel pada SVM. Studi oleh Rahayu dan Yamasari (2024) mengenai klasifikasi penyakit *stroke* menunjukkan bahwa kernel *polynomial* memberikan hasil terbaik dengan akurasi 78,86%, *precision* 73,98%, dan *recall* 56,75% menggunakan pembagian data 80:20 [18]. Penelitian lain terkait klasifikasi gagal jantung menemukan bahwa kernel *linear* memberikan akurasi tertinggi 86,92%, sedangkan kernel *polynomial* paling rendah dengan 85,83% [19]. Pada klasifikasi status gizi balita dengan optimasi *Grid Search Cross-Validation*, kernel RBF menghasilkan performa terbaik dengan akurasi 85,71% [20]. Prediksi anemia berbasis hemoglobin darah juga menunjukkan bahwa kernel *linear* memberikan performa terbaik dengan akurasi hingga 99,3% [21]. Dari beberapa penelitian tersebut terlihat bahwa pemilihan kernel SVM sangat dipengaruhi oleh karakteristik data dan jenis penyakit yang diklasifikasikan. Namun, sebagian besar studi tersebut hanya menguji satu atau dua kernel pada domain

medis tertentu, sehingga masih diperlukan pengujian komparatif yang terfokus pada kasus penyakit jantung dengan keempat kernel pada SVM.

Di luar bidang medis, pemanfaatan SVM juga meluas ke sektor lain, seperti pertanian dan pengolahan citra digital. Dalam penelitian klasifikasi penyakit daun pada tanaman kentang menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM), SVM mencapai akurasi 83,12%, terutama pada identifikasi penyakit *early blight*, *late blight*, dan *non-disease* [22]. Penelitian Nahak dkk. (2024) mengenai klasifikasi penyakit daun apel dengan metode *Multiclass* SVM yang membandingkan kernel polynomial dan *linear*. Hasilnya menunjukkan bahwa kernel *polynomial* menghasilkan akurasi tertinggi 76,67%, sedangkan kernel *linear* mencapai 75,23% [23]. Variasi hasil ini memperlihatkan bahwa pemilihan kernel SVM sangat dipengaruhi oleh karakteristik *dataset* dan skema evaluasi yang digunakan. Oleh karena itu, penelitian ini difokuskan pada analisis perbandingan kinerja keempat kernel tersebut dalam mengklasifikasi risiko penyakit jantung guna menentukan kernel yang paling optimal.

2. METODE PENELITIAN

Penelitian ini memanfaatkan *dataset* penyakit jantung dari *platform* Kaggle berjudul "*Heart Disease Prediction Dataset*" yang diunggah oleh Krish Ujeniya. *Dataset* tersebut memuat berbagai atribut kesehatan pasien, antara lain usia, jenis kelamin, tekanan darah, kadar kolesterol, serta fitur medis lain. Label target menunjukkan ada atau tidaknya indikasi penyakit jantung. Untuk mengolah data tersebut, digunakan algoritma SVM yang dikenal efektif dalam menangani data bersifat *linear* maupun *non-linear*. Fokus penelitian ini adalah membandingkan kinerja empat jenis kernel pada SVM, yaitu *linear*, *polynomial*, RBF, dan *sigmoid*. Proses penelitian diawali dengan pemuatan *dataset* dan eksplorasi awal. Setelah itu dilakukan tahap pra-pemrosesan yang meliputi pemisahan fitur (X) dan target (y). *Dataset* kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan *train_test_split*. Seluruh fitur dinormalisasi menggunakan *StandardScaler* agar berada pada skala yang seragam. Keseluruhan rangkaian tahapan penelitian tersebut digambarkan pada Gambar 1, yang menampilkan skema alur penelitian secara menyeluruh.



Gambar 1. Skema Alur Penelitian

Gambar 1 memperlihatkan skema alur klasifikasi yang digunakan dalam penelitian ini. Setelah tahap normalisasi data selesai dilakukan, model SVM diinisialisasi. Pemodelan kemudian dilaksanakan secara iteratif untuk setiap jenis kernel. Untuk memperoleh performa yang optimal, setiap kernel dituning menggunakan *GridSearchCV* dengan validasi silang 5-*fold* pada data latih. Pada tahap ini diuji beberapa kombinasi *hyperparameter*, yaitu C = {0.1, 1, 10, 100} untuk seluruh kernel, *gamma* = {1, 0.1, 0.01, *scale*} untuk kernel RBF, *polynomial*, dan *sigmoid*, serta *degree* = {2, 3, 4} untuk kernel *polynomial*. Model dengan parameter terbaik selanjutnya dievaluasi menggunakan data uji. Hasil prediksi kemudian dibandingkan dengan data aktual guna menghitung metrik evaluasi berupa akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi ini digunakan sebagai dasar untuk menilai efektivitas masing-masing kernel dalam klasifikasi risiko penyakit jantung.

Dalam penelitian ini, algoritma SVM diimplementasikan dengan empat jenis kernel berbeda. Masing-masing kernel memiliki fungsi pemetaan (*kernel function*) ke dalam ruang berdimensi tinggi. Kernel *linear* digunakan untuk data yang bersifat *linear* dan dapat dipisahkan dengan garis lurus. Kernel *polynomial* dan RBF berfungsi untuk menangani pola data *non-linear* yang lebih kompleks. Adapun kernel *sigmoid*

umumnya digunakan untuk memodelkan hubungan yang menyerupai fungsi aktivasi pada jaringan saraf. Untuk menggambarkan karakteristik teknis masing-masing kernel, fungsi matematis dari keempat kernel utama SVM disajikan pada Tabel 1.

Tabel 1. Fungsi Kernel pada Algoritma SVM

Fungsi Kernel	Nama Kernel
$K(x_i, x_j) = x_i T . x_j$	Kernel *linear*
$K(x_i, x_j) = \gamma x_i T x_j + r)^d, \gamma > 0$	Kernel *polynomial*
$K(x_i, x_j) = exp(-\gamma \|\| x_i - x_j \|\|^2), \gamma > 0$	Kernel RBF
$K(x_i, x_j) = tanh(\gamma x_i T x_j + r)$	Kernel *sigmoid*

3. HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini memuat berbagai fitur yang merepresentasikan karakteristik kesehatan pasien. Fitur-fitur tersebut menjadi dasar dalam penentuan apakah pasien berpotensi memiliki risiko penyakit jantung. Tabel 2 menyajikan rangkuman fitur yang tersedia dalam *dataset* yang diperoleh dari Kaggle.

Tabel 2. Fitur-fitur dalam *dataset*

Fitur	Deskripsi
age	Usia pasien
sex	Jenis kelamin pasien (0 = perempuan, 1 = laki-laki)
cp	Jenis nyeri dada (0-3, menunjukkan tingkat dan tipe nyeri dada)
trestbps	Tekanan darah saat istirahat dalam mm Hg
chol	Kadar kolesterol serum dalam mg/dl
fbs	Gula darah puasa (1 jika ≥ 120 mg/dl, 0 jika < 120 mg/dl)
restecg	Hasil elektrokardiografi istirahat (kategorikal)
thalach	Detak jantung maksimum yang dicapai
exang	Angina (nyeri dada) yang dipicu oleh olahraga (0 = tidak, 1 = ya)
oldpeak	Depresi ST yang disebabkan oleh olahraga relatif terhadap istirahat
slope	Kemiringan segmen ST saat latihan puncak
ca	Jumlah pembuluh darah utama (0-3) yang diwarnai oleh fluoroskopi
thal	Kondisi talasemia (dengan beberapa kategori)
target	Diagnosis penyakit jantung (1 = ada indikasi penyakit jantung, 0 = tidak ada indikasi penyakit)

Sebelum dilakukan pelatihan model, data dibersihkan dari nilai hilang dan dikonfirmasi kembali konsistensinya. Tabel 3 berikut menampilkan 10 sampel data awal untuk memperlihatkan variasi nilai pada setiap atribut.

Tabel 3. Sample data (10 baris pertama)

No	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
2	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
3	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
4	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
5	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
6	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
7	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
8	44	1	1	120	263	0	1	173	0	0.0	2	0	3	1
9	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
10	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1

Dataset yang digunakan berjumlah 303 *record* dan kemudian dibagi menjadi data latih 80% (242 data) dan data uji 20% (61 data). Pemisahan dilakukan sebelum penskalaan untuk mencegah data *leakage*. Standarisasi fitur numerik diterapkan menggunakan *StandardScaler* yang dipasang pada data latih dan diaplikasikan konsisten ke himpunan latih maupun uji. Prosedur ini memastikan seluruh fitur berada pada skala yang sebanding sehingga margin pemisah SVM tidak terdistorsi oleh perbedaan skala. Tahap pra-pemrosesan menjadi prasyarat sebelum model diinisialisasi dan dioptimasi.

Selanjutnya dilakukan inisialisasi model SVM dan iterasi fungsi kernel yang mencakup *linear*, *polynomial*, RBF, dan *sigmoid*. Optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* berbasis 5-*fold* pada data latih sehingga perbandingan antarkernel berlangsung adil. Rentang *C*, *gamma*, dan *degree* diseragamkan antarkernel dan rinciannya telah diuraikan pada bagian Metode. Pemilihan rentang tersebut

merepresentasikan regularisasi rendah–sedang–tinggi serta variasi lebar fungsi kernel *non-linear*, sehingga sensitivitas model terhadap perubahan parameter tetap teramati. Untuk setiap kernel, model dengan skor validasi rata-rata tertinggi pada skema 5-*fold* ditetapkan sebagai konfigurasi akhir. Konfigurasi terbaik ini kemudian dilatih ulang pada data latih dan dievaluasi pada data uji, sehingga interpretasi kinerja tidak bias oleh proses pemilihan model. Pendekatan ini diterapkan untuk meminimalkan bias pemilihan model dan meningkatkan reliabilitas perbandingan.

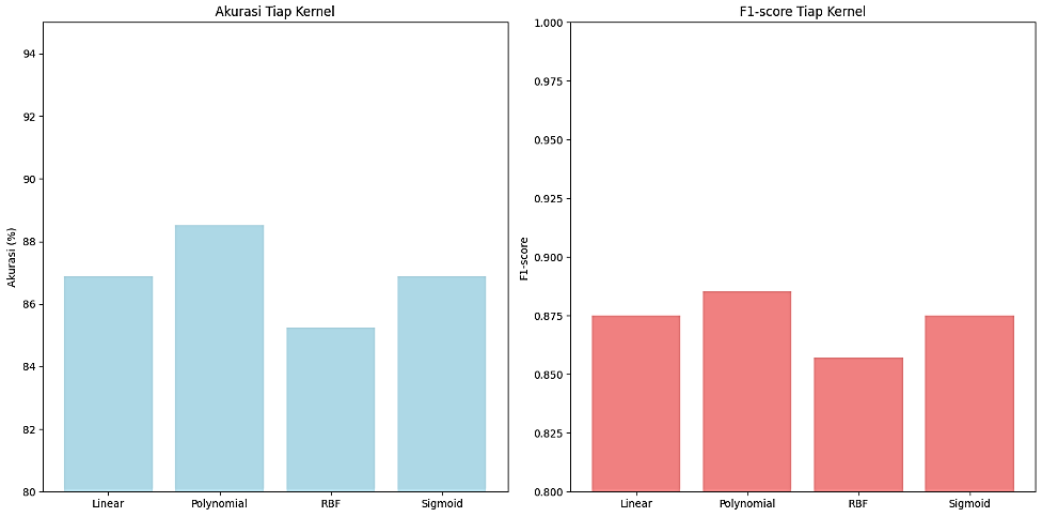
Pada fase pelatihan dan pengujian, model terpilih dilatih pada himpunan latih kemudian digunakan untuk menghasilkan prediksi pada data uji. Performa dievaluasi dengan membandingkan label prediksi terhadap label aktual menggunakan metrik akurasi, *precision*, *recall*, dan F1-*score*. Pertimbangan juga diberikan pada keseimbangan kesalahan Tipe I (*false positive*) dan Tipe II (*false negative*), yang berimplikasi pada keseimbangan risiko salah klasifikasi. Ringkasan hasil per kernel disajikan pada Tabel 4.

Tabel 4. Hasil Evaluasi Performa Model SVM pada Klasifikasi Risiko Penyakit Jantung

Kernel	*Best Parameters*	*Accuracy (%)*	*Recall*	*F1-score*
Linear	*C*: 1	86,89%	88%	87%
Polynomial	*C*: 1, *degree*: 3, *gamma* 0,1	88,52%	84%	89%
RBF	*C*: 10, *gamma*: 0.1	85,25%	84%	85%
Sigmoid	*C*: 1, *gamma*: *scale*	86,89%	88%	87%

Berdasarkan Tabel 4, kernel *polynomial* menunjukkan performa terbaik dengan akurasi sebesar 88,52% dan F1-score sebesar 89%. Hasil ini menegaskan bahwa kernel *polynomial* cukup efektif dalam memisahkan data dengan pola *non-linear*. Dengan memetakan data ke ruang berdimensi lebih tinggi, kernel ini membantu model membedakan kelas dengan lebih akurat, terutama ketika distribusi data tidak dapat dipisahkan secara langsung menggunakan garis lurus. Kernel *linear* dan *sigmoid* menghasilkan performa yang hampir sama, yakni akurasi 86,89% dengan F1-*score* 87%. Kernel *linear* tetap kompetitif karena sifatnya yang sederhana dan efisien secara komputasi, menunjukkan bahwa sebagian besar data masih memiliki pola pemisahan yang relatif linier. Sementara itu, meskipun kernel *sigmoid* memberikan hasil serupa dengan kernel *linear*, kinerjanya tidak menunjukkan keunggulan signifikan. Hal ini kemungkinan dipengaruhi oleh sensitivitas fungsi aktivasi *sigmoid* terhadap skala data dan parameter, sehingga performanya cenderung fluktuatif jika distribusi data tidak sesuai dengan karakteristik fungsi *sigmoid*.

Di sisi lain, kernel RBF memperoleh akurasi terendah sebesar 85,25%, meskipun secara umum kernel ini dikenal fleksibel dalam menangani data *non-linear*. Rendahnya performa RBF dalam penelitian ini kemungkinan disebabkan oleh parameter *gamma* hasil tuning yang belum optimal atau distribusi data yang kurang sesuai dengan fungsi *Gaussian*. Selain itu, kernel RBF relatif sensitif terhadap keberadaan *outlier*, yang dapat menurunkan kinerja model jika tidak ditangani dengan baik. Untuk memperjelas perbandingan, Gambar 2 menyajikan grafik batang yang menampilkan nilai akurasi dan F1-*score* dari setiap kernel yang diuji. Visualisasi ini memberikan gambaran yang lebih intuitif mengenai perbedaan kinerja antar kernel. Berdasarkan grafik tersebut, kernel *polynomial* tetap konsisten menunjukkan performa terbaik, diikuti oleh kernel *linear* dan *sigmoid* yang memberikan hasil hampir setara, sementara kernel RBF menampilkan performa paling rendah sesuai dengan hasil pada Tabel 4. Temuan ini memperkuat kesimpulan bahwa kernel *polynomial* merupakan pilihan paling efektif untuk *dataset* yang digunakan dalam penelitian ini.



Gambar 2. Visualisasi Grafik Batang Akurasi dan F1-*Score*

Secara keseluruhan, kernel *polynomial* tampil paling kuat pada metrik utama penelitian ini. Kinerja *linear* dan *sigmoid* berada pada tingkat yang relatif seimbang, sedangkan RBF sedikit tertinggal dibanding keduanya. Skema penalaan diterapkan secara seragam menggunakan *GridSearchCV* 5-*fold* pada data latih sehingga perbandingan antarkernel bersifat adil. Dengan pengaturan tersebut, perbedaan hasil lebih merefleksikan karakter pemetaan masing-masing kernel daripada variasi konfigurasi. Temuan ini konsisten dengan tujuan studi yang menilai empat kernel dalam satu skema eksperimen yang seragam.

Hasil tersebut sejalan dengan literatur yang menunjukkan kernel terbaik bergantung pada domain dan karakteristik data. Pada klasifikasi *stroke*, *polynomial* dilaporkan unggul [18]. Untuk gagal jantung dan anemia, *linear* mencapai performa tertinggi [19][21]. Sementara pada status gizi balita, RBF menjadi yang terbaik [20]. Rentang akurasi studi ini (sekitar 85–89%) masih dapat diperbandingkan dengan laporan-laporan tersebut, sehingga posisi hasil ini berada dalam kisaran yang wajar untuk *dataset* publik sejenis.

Dari sisi implikasi, F1-*score polynomial* yang lebih tinggi menunjukkan keseimbangan *precision*–*recall* yang lebih baik. Kondisi ini relevan untuk mengurangi risiko false negative pada konteks klinis, tanpa meningkatkan *false positive* secara berlebihan. *Linear* tetap menarik karena stabil dan efisien secara komputasi, sehingga berguna ketika sumber daya terbatas atau interpretabilitas *margin linear* diutamakan. *Sigmoid* menghasilkan kinerja mendekati *linear* namun cenderung kurang konsisten di seluruh metrik. RBF berpotensi meningkat apabila rentang *gamma* diperluas atau profil fitur lebih sesuai dengan pemetaan *Gaussian*.

Penelitian ini memiliki beberapa keterbatasan yang perlu dicatat. Ukuran data relatif kecil (303 *record*) dan penyeimbangan kelas tidak diterapkan secara khusus. Rentang *hyperparameter* yang dieksplorasi bersifat moderat, sehingga perluasan pencarian (misalnya *log-scale* untuk *C* dan *gamma*) berpotensi meningkatkan performa. Arah pengembangan lanjutan meliputi uji *SMOTE* bila ditemukan ketimpangan label, serta pembandingan dengan model *non*-SVM untuk memperkuat konteks komparatif. Selain itu, evaluasi *stratified cross-validation* penuh dapat dipertimbangkan guna menurunkan variansi estimasi.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma SVM dalam mengklasifikasikan risiko penyakit jantung dengan membandingkan empat jenis kernel, yakni *linear*, *polynomial*, RBF, dan *sigmoid*. Hasil analisis menunjukkan bahwa kernel *polynomial* memberikan performa terbaik dengan akurasi 88,52% dan F1-*score* 89%, menegaskan efektivitasnya dalam memodelkan pola data *non-linear*. Kernel *linear* dan *sigmoid* menghasilkan kinerja yang relatif seimbang, masing-masing dengan akurasi 86,89% dan F1-*score* 87%. Kernel RBF mencatat akurasi sedikit lebih rendah, yaitu 85,25%, yang menunjukkan bahwa pada konfigurasi parameter terbaik yang ditemukan, pemetaan RBF belum setepat kernel *polynomial* untuk *dataset* ini. Secara keseluruhan, kernel *polynomial* dapat dianggap sebagai pilihan paling optimal dalam klasifikasi risiko penyakit jantung pada *dataset* ini, dan capaian tersebut masih berada pada rentang akurasi 85–90% yang juga dilaporkan pada penelitian lain yang menggunakan *dataset* penyakit jantung publik. Meskipun kernel linear juga menampilkan hasil yang cukup baik, kernel *polynomial* tetap unggul terutama dari sisi F1-*score*, yang menunjukkan keseimbangan *precision* dan *recall* yang lebih sesuai untuk konteks medis, sementara kontribusi penelitian ini adalah menyajikan perbandingan empat kernel SVM dalam satu skema eksperimen yang seragam pada *dataset* publik.

REFERENSI

[1] R. M. Ubaidilah and T. A. Puspito, "Optimasi Prediksi Penyakit Jantung Dengan Naive Bayes dan Particle Swarm Optimization (PSO)," *Jurnal Informasi dan Komputer*, vol. XIII, no. 1, pp. 148-154, 2025.

[2] A. Nurmasani and Y. Pristyanto, "Algortime Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class," *Jurnal Pseudocode*, vol. VIII, no. 1, pp. 21-26, 2021.

[3] R. Sumara, N. A. Wibowo and Indarti, "Identifikasi Faktor Kejadian Penyakit Jantung Koroner Terhadap Wanita Usia ≤ 50 Tahun di RSU Haji Surabaya," *Jurnal Manajemen Asuhan Keperawatan*, vol. VI, no. 2, pp. 53-59, 2022.

[4] P. A. Sihotang and D. Sitanggang, "Perbandingan Algoritma C4.5 Dengan Naive Bayes Untuk Memprediksi Penyakit Jantung," *Jurnal TEKINKOM*, vol. VII, no. 2, pp. 899-908, 2024.

[5] D. H. Depari, Y. Widiastiwi and M. M. Santoni, "Perbandingan Model Desicion Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Jurnal Informatik*, vol. XVIII, no. 3, pp. 239-248, 2022.

[6] M. T. T. B. Sirait, N. S. Fathonah and M. N. Fauzan, "Pemanfaatan Algoritma Adasyn dan Support Vector Machine Dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru," JATI (Jurnal Mahasiswa Teknik Informatika), vol. VIII, no. 5, pp. 8773-8778, 2024.

[7] M. D. F. Tino, H. Hasanah and T. D. Santosa, "Perbandingan Algoritma Support Vector Machines (SVM) dan Neural Network Untuk Klasifikasi Penyakit Jantung," Infotech Journal, vol. IX, no. 1, pp. 232-235, 2023.

[8] K. Saputra, Zuriati and Sriyanto, "Perbandingan Kinerja Fungsi Kernel Algoritma Support Vector Machine Pada Klasifikasi Penyakit Padi," *Jurnal Teknika,* vol. XVII, no. 1, pp. 119-131, 2023.
[9] W. S. Dharmawan, "Komparasi Algoritma Klasifikasi SVM-PSO dan C4.5-PSO Dalam Prediksi Penyakit Jantung," *Jurnal Informatika, Manajemen dan Komputer,* vol. XIII, no. 2, pp. 31-41, 2021.
[10] F. M. Natsir, R. Y. Bakti and T. Wahyuni, "Analisis Deteksi Penyakit Jantung Dengan Pendekatan Support Vector Machine Pada Data pasien," Arus Jurnal Sains dan Teknologi (AJST), vol. II, no. 2, pp. 437-446, 2024.
[11] Alexsander, A. Nazri, R. A. Panbudi and Junadhi, "Implementasi Algoritma SVM Dalam Memprediksi Penyakit Stroke," *Journal Zetroem,* vol. VI, no. 2, pp. 1-5, 2024.
[12] H. S. Wafa, A. I. Hadiana and F. R. Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics and Digital Expert (INDEX),* vol. IV, no. 1, pp. 40-45, 2022.
[13] A. d. R. Wibisono, E. P. Mandyartha and M. M. A. Haromainy, "Klasifikasi Penyakit Kulit Berbasis Support Vector Machine Dengan Ekstraksi Fitur ABCD Rule," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika),* vol. X, no. 1, pp. 686-698, 2025.
[14] F. Handayani and R. M. Taufiq, "Komparasi Algoritma Menggunakan Teknik SMOTE Dalam Melakukan Klasifikasi Penyakit Stroke Otak," *Jurnal Computer Science and Information Technology (CoSciTech),* vol. V, no. 2, pp. 367-372, 2024.
[15] A. Desiani, D. A. Zayanti, I. Ramayanti, F. F. Ramadhan and Giovillando, "Perbandingan Algoritma Support Vector Machine (SVM) Dan Logistic Regression Dalam Klasifikasi Kanker Payudara," *Jurnal Kecerdasan Buatan dan Teknologi,* vol. IV, no. 1, pp. 33-42, 2025.
[16] A. Jalil, A. Homaidi and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech: Jurnal Teknologi Terapan,* vol. VIII, no. 8, pp. 2070-2079, 2024.
[17] Angel and D. E. Herwindiati, "Perbandingan Algoritma K-NN, SVM, Dan Decision Tree Dalam Klasifikasi Kelenjar Tiroid," *JTeksis (Jurnal Teknologi Dan Sistem Informasi Bisnis),* vol. VI, no. 4, pp. 866-871, 2024.
[18] S. Rahayu and Y. Yamasari, "Klasifikasi Penyakit Stroke Dengan Metode Support Vector Machine (SVM)," *JINACS (Journal of Informatics and Computer Science),* vol. V, no. 3, pp. 440-446, 2024.
[19] L. N. Farida and S. Bahri, "Klasifikasi Gagal Jantung Menggunakan SVM (Support Vector Machine)," *Komputika: Jurnal Sistem Komputer,* vol. XIII, no. 2, pp. 149-156, 2024.
[20] A. Nadroh, D. N. Triwibowo and R. B. B. Sumantri, "Klasifikasi Status Gizi Balita Menggunakan Algoritma Support Vector Machine Dengan Optimasi Grid Search Cross-Validation," *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi,* vol. VIII, no. 2, pp. 250-257, 2024.
[21] A. Wulandari, S. Wahyuni, D. Z. Haq and D. C. R. Novitasari, "Identifikasi Penyakit Anemia Menggunakan Metode Support Vector Machine Berdasarkan Hemoglobin Darah," *Jurnal Algoritme,* vol. V, no. 2, pp. 101-110, 2025.
[22] R. Marlita and D. Mustofa, "Implementasi Support Vector Machine Pada Klasifikasi Penyakit Daun Kentang Berbasis Fitur GLCM," *JIKA: Jurnal Ilmu Komputer dan Aplikasinya,* vol. I, no. 1, pp. 6-11, 2025.
[23] E. Nahak, R. P. Putra and F. Marisa, "Klasifikasi Penyakit Pada Tanaman Apel Melalui Citra Daun Menggunakan Metode Multiclass Support Vector Machine," *JATASI (Jurnal Teknik Informatika dan Sistem Informasi),* vol. XI, no. 3, pp. 401-408, 2024.