

MENGOPTIMALKAN PREDIKSI GAGAL JANTUNG DENGAN KOMBINASI SVM DAN FORWARD SELECTION

Verry Riyanto¹, Henny Destiana², Titin Prihatin³, Sugiono⁴, Ganda Wijaya⁵

¹Program Studi Sistem Informasi, Universitas Bina Sarana Informatika, , ²Program Studi Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, ^{3,4}Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika, ⁵Program Studi Sistem Informasi, Universitas Nusa Mandiri.

^{1,3,4,5}Jln. Kramat Raya No.98, Senen, Jakarta Pusat 10450, ²Jln. Jatiwaringin Raya No. 2, Jakarta Timur 13620

¹verry.vry@bsi.ac.id, ²henny.hnd@bsi.ac.id, ³titin.tpn@bsi.ac.id, ⁴sugiono.sg@bsi.ac.id,
⁵ganda.gws@nusamandiri.ac.id

Abstract

Heart failure is a critical health condition with a mortality rate that continues to increase, with the problem that traditional diagnosis is often less accurate and efficient, so more precise and efficient early diagnosis methods are needed. Previous research has improved prediction accuracy with various methods but is still limited in optimal feature selection and modeling efficiency. Therefore, this research aims to analyze kernel performance in Support Vector Machine (SVM) algorithms such as Dot, Radial, Polynomial and analyze the effectiveness of Forward Selection (FS) in selecting the most significant features to optimize heart failure risk predictions. The research results show that the Radial kernel with FS has the best performance with AUC 0.881, Accuracy 84.64%, and Recall 92.55%. The time and serum_creatinine features proved to be the most significant in improving model performance. This research proves that the combination of SVM and FS is able to produce a more precise and efficient solution in early diagnosis of heart failure compared to previous approaches. These results are expected to support the development of artificial intelligence-based prediction systems for more reliable clinical applications.

Keywords : Heart Failure, Support Vector Machine, Forward Selection, Kernel, Predictive Diagnosis.

Abstrak

Gagal jantung merupakan salah satu kondisi kesehatan kritis dengan angka kematian yang terus meningkat, dengan permasalahan yang ada diagnosis tradisional seringkali kurang akurat dan efisien sehingga diperlukan metode diagnosis dini yang lebih presisi dan efisien. Penelitian sebelumnya telah meningkatkan akurasi prediksi dengan berbagai metode namun masih terbatas dalam pemilihan fitur optimal dan efisiensi pemodelan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kinerja kernel pada algoritma *Support Vector Machine* (SVM) seperti *Dot*, *Radial*, *Polynomial* dan menganalisis efektivitas *Forward Selection* (FS) dalam memilih fitur paling signifikan guna mengoptimalkan prediksi risiko gagal jantung. Hasil penelitian menunjukkan bahwa kernel *Radial* dengan FS memiliki performa terbaik dengan AUC 0.881, Akurasi 84,64%, dan Recall 92,55%. Fitur *time* dan *serum_creatinine* terbukti paling signifikan dalam meningkatkan performa model. Penelitian ini membuktikan bahwa kombinasi antara SVM dan FS mampu menghasilkan solusi yang lebih presisi dan efisien dalam diagnosis dini gagal jantung dibandingkan pendekatan sebelumnya. Hasil ini diharapkan dapat mendukung pengembangan sistem prediksi berbasis kecerdasan buatan untuk aplikasi klinis yang lebih andal.

Kata kunci : Gagal Jantung, Support Vector Machine, Forward Selection, Kernel, Diagnosis Prediktif.

1. PENDAHULUAN

Penyakit gagal jantung menjadi salah satu penyebab utama angka kematian diseluruh negara [1]. Menurut data organisasi kesehatan dunia (WHO) dengan estimasi 17,9 juta kematian setiap tahun[2]. Berdasarkan Riset Kesehatan Dasar (Riskesdas) 2013 dan 2018, terdapat tren peningkatan prevalensi penyakit jantung di Indonesia, dari 0,5% pada 2013 menjadi 1,5% pada 2018 [3]. Kasus penyakit ini semakin meningkat seiring dengan pertambahan usia, populasi dan perubahan gaya hidup masyarakat [4]. Kondisi ini membutuhkan pendekatan yang lebih efektif untuk diagnosis dini guna mencegah komplikasi lebih lanjut[5]. Salah satu langkah penting adalah pemanfaatan teknologi kecerdasan buatan dengan metode *Support Vector Machine* (SVM) [6], untuk menaikkan keakuratan prediksi risiko gagal jantung [7].

Namun, kinerja algoritma SVM sangat dipengaruhi oleh pemilihan kernel dan fitur yang digunakan [8]. Kernel seperti dot, polynomial, dan radial memiliki karakteristik unik yang dapat memberikan hasil berbeda dalam klasifikasi [8]. Selain itu, metode feature selection seperti *Forward Selection* (FS) dapat membantu memilih fitur yang paling relevan untuk meningkatkan kinerja prediksi [9]. Kombinasi kedua pendekatan ini berpotensi menghasilkan keluaran akurasi yang lebih presisi dan efisien dibandingkan dengan metode tradisional [10].

Beberapa penelitian sebelumnya telah mengeksplorasi teknik optimasi lain untuk meningkatkan akurasi prediksi gagal jantung. Sebagai contoh, penelitian [11] membahas penerapan *Particle Swarm Optimization* (PSO) pada *Neural Network*. Hasilnya menunjukkan peningkatan akurasi sebesar 8,35% menjadi 85,61% dan *Area Under Curve* (AUC) meningkat menjadi 0,869. Penelitian ini menekankan pentingnya optimasi parameter untuk meningkatkan kinerja model diagnostik. Penelitian [12] mengintegrasikan *Random Forest* dengan PSO untuk meningkatkan akurasi prediksi risiko gagal jantung menjadi 87,33%. Teknik ini menunjukkan bahwa optimasi parameter melalui PSO dapat meningkatkan performa model secara signifikan dibandingkan *Random Forest* yang digunakan secara terpisah. Selanjutnya, penelitian [13] mengevaluasi penggunaan *Decision Tree* dengan pendekatan PSO. Hasilnya menunjukkan akurasi sebesar 85,29%, serta memberikan wawasan mendalam mengenai faktor-faktor yang memengaruhi risiko gagal jantung. Penelitian ini menggarisbawahi pentingnya kombinasi metode diagnostik untuk hasil yang lebih akurat dan

efisien. Terakhir, penelitian [14] membandingkan kinerja metode *Naive Bayes* dengan optimasi PSO. Hasil penelitian ini mendukung efektivitas pendekatan PSO dalam meningkatkan keandalan metode diagnostik tradisional.

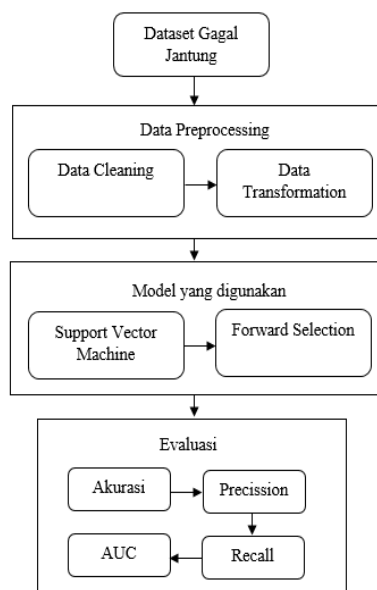
Meskipun penelitian-penelitian ini berhasil meningkatkan akurasi prediksi gagal jantung, masih terdapat permasalahan terkait evaluasi kinerja kernel SVM (dot, polynomial, radial) dan seberapa efektif metode FS dalam memilih fitur yang signifikan untuk meningkatkan akurasi model SVM. Kontribusi utama penelitian ini adalah memberikan analisis komprehensif mengenai optimasi model prediksi gagal jantung, yang belum banyak dibahas dalam penelitian sebelumnya, dengan pendekatan evaluatif terhadap performa berbagai kernel SVM serta efektivitas metode forward selection dalam pemilihan fitur.

Dengan demikian, penelitian ini berfokus pada upaya optimalisasi prediksi gagal jantung dengan mengevaluasi performa kernel SVM (dot, polynomial, radial) serta menganalisis efektivitas metode forward selection dalam memilih fitur yang signifikan. Temuan dalam penelitian ini diharapkan dapat memberikan dampak signifikan dalam pengembangan alat bantu diagnosis yang lebih akurat, sehingga relevan dengan kebutuhan klinis saat ini dengan memberikan solusi berbasis data yang tidak hanya memperbaiki akurasi [15], tetapi juga meningkatkan keandalan diagnosis dini gagal jantung.

2. METODOLOGI PENELITIAN

2.1. Skema Alur Penelitian

Skema alur penelitian ini menggambarkan langkah-langkah yang dilalui dalam rangka mengoptimalkan prediksi gagal jantung, menggunakan pendekatan SVM dan teknik FS untuk memilih fitur yang paling relevan. Proses penelitian ini terdiri dari berbagai tahapan yang saling terintegrasi terlihat pada Gambar 1.



Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Penelitian ini menggunakan data sekunder sebagai bahan analisis, yang berasal dari dataset medis terkait penyakit gagal jantung yang didapat dari UCI Repository. Data ini mencakup informasi pasien seperti usia, jenis kelamin. Fitur riwayat kesehatan mencakup tekanan darah tinggi, diabetes, dan hasil tes medis lainnya yang relevan untuk diagnosis gagal jantung. Pada data tersebut tercatat apakah pasien terdeteksi gagal jantung atau tidak.

2.3. Data Preprocessing

Pada tahap *preprocessing* data, langkah pertama yang dilakukan adalah data *cleaning*, yang mencakup penggantian nilai yang hilang menggunakan teknik *Replace Missing Value*. Meskipun tidak ditemukan nilai kosong dalam dataset, langkah ini tetap diterapkan sebagai langkah preventif untuk memastikan konsistensi dan kualitas data. Penggantian nilai dilakukan dengan cara yang tepat, seperti menggantikan nilai hilang dengan nilai rata-rata untuk data numerik atau nilai modus untuk data kategorikal. Selanjutnya, dilakukan juga data transformation untuk memastikan bahwa data berada dalam bentuk yang sesuai, seperti normalisasi atau standarisasi, sehingga siap untuk digunakan dalam analisis dan pelatihan model.

2.4. Model SVM dan FS

Pada tahap ini, dilakukan penerapan SVM untuk membangun model prediksi gagal jantung. SVM dipilih karena efektif dalam menangani data yang rumit dan memberikan hasil yang akurat, terutama dalam masalah klasifikasi [16]. Untuk meningkatkan performa model dilakukan pemilihan kernel SVM yang sesuai yaitu *dot*, *polynomial*, dan *radial* [17]. Setiap kernel memiliki karakteristik yang berbeda dalam mentransformasikan data ke dalam ruang dimensi yang lebih tinggi [18], dan pemilihan kernel yang tepat diharapkan dapat meningkatkan kemampuan model dalam memisahkan kelas-kelas pada data [19].

Selain itu, diterapkan pula seleksi fitur untuk memilih fitur-fitur yang paling relevan dalam membangun model prediksi [20]. Teknik FS digunakan sebagai metode seleksi fitur yang bersifat *bottom-up*, yaitu memulai dengan model kosong (tanpa fitur) lalu menambahkan fitur satu per satu, yang dimulai dengan memilih fitur yang memiliki kontribusi paling besar terhadap akurasi model dan kemudian menambahkannya satu per satu berdasarkan kriteria tertentu [21]. Mekanisme pemilihan fitur dalam FS dilakukan dengan cara mengevaluasi setiap fitur secara individual, lalu memilih fitur yang memberikan peningkatan terbesar terhadap akurasi dan nilai F1-score. Dengan menggunakan FS, hanya fitur-fitur yang paling signifikan yang dipilih sehingga mengurangi kemungkinan *overfitting* dan meningkatkan efisiensi model [22].

Proses ini bertujuan untuk menemukan kombinasi kernel SVM dan subset fitur yang optimal, yang dapat memberikan akurasi terbaik dalam memprediksi risiko gagal jantung. Setelah model selesai dilatih, tahap evaluasi dilakukan untuk menilai performa model menggunakan metrik-metrik yang ada pada datamining.

2.5. Evaluasi

Setelah model SVM dilatih dan fitur yang relevan telah dipilih melalui teknik FS, tahap selanjutnya adalah evaluasi model untuk mengukur kinerjanya dalam memprediksi risiko gagal jantung dengan melibatkan *confusion matrix* [23]. Evaluasi ini penting untuk menilai sejauh mana model dapat menggeneralisasi data [24].

3. HASIL DAN PEMBAHASAN

Bagian ini membahas penerapan algoritma SVM dengan berbagai kernel dalam mengklasifikasikan tingkat risiko penyakit gagal

jantung disertai teknik FS. Model tersebut dibangun, diterapkan, dan diuji dengan Kinerja model dievaluasi menggunakan matriks kebingungan dan ROC AUC.

3.1. Pengumpulan Data

Proses ini menggunakan dataset yang diambil dari sebuah web UCI repository yang merupakan kumpulan data sekunder. Data terdiri dari 299 entri dengan 12 variabel dan 1 label. Data ini diambil dari UCI Repository dan mencakup berbagai atribut yang relevan untuk diagnosis gagal jantung seperti usia, tekanan darah tinggi, diabetes, serta hasil tes medis lainnya seperti yang dapat dilihat pada Gambar 2.

Nama Variabel	Type	Units
age	Integer	years
anaemia	Binary	
creatinine_phosphokinase	Integer	mcg/L
diabetes	Binary	
ejection_fraction	Integer	%
high_blood_pressure	Binary	
platelets	Continuous	kiloplatelets/mL
serum_creatinine	Continuous	mg/dL
serum_sodium	Integer	mEq/L
sex	Binary	
smoking	Binary	
time	Integer	days
death_event	Binary	

Gambar 2. Dataset Gagal Jantung

Dataset ini telah diolah dan dipersiapkan lebih lanjut melalui proses data preprocessing untuk memastikan kualitas data yang optimal. Pada variabel terdapat *type* integer, binary dan continuous. Data integer bilangan bulat, binary nilai 0 atau 1, continuous bilangan desimal. Untuk units years mengukur usia pasien dalam satuan waktu tahun. Mikrogram per liter (mcg/L) digunakan untuk variabel creatinine_phosphokinase menunjukkan kadar enzim CPK dalam darah, yang diukur dalam mikrogram per liter. Kiloplatelets/mL (ribuan trombosit per mililiter) digunakan untuk variabel platelets yang berfungsi untuk mengukur jumlah trombosit dalam darah, dinyatakan dalam ribuan trombosit per mililiter darah. Miligram per desiliter (mg/dL) digunakan untuk variabel serum_creatinine yang berfungsi menunjukkan kadar kreatinin dalam darah, diukur dalam miligram per desiliter. Miliekuivalen per liter (mEq/L) digunakan untuk variabel serum_sodium yang berfungsi mengukur konsentrasi natrium dalam darah dalam satuan miliekuivalen per liter.

3.2. Data Preprocessing

Pada tahap ini setelah eksplorasi awal terhadap dataset, tidak ditemukan nilai kosong pada atribut apa pun. Namun, langkah *Replace Missing Value* tetap diterapkan sebagai bagian dari proses data *cleaning* untuk menangani potensi nilai kosong yang dapat memengaruhi hasil analisis. Setelah itu, dilakukan data *preprocessing* yang mencakup normalisasi menggunakan teknik *Min-Max Scaling* yang mentransformasi setiap fitur ke dalam rentang 0 hingga 1 untuk menyelaraskan skala fitur sehingga meningkatkan stabilitas dan kinerja model prediksi. Meskipun dataset ini tidak memiliki nilai kosong, langkah ini tetap diterapkan sebagai prosedur standar *pipeline preprocessing* untuk memastikan konsistensi data yang fleksibel jika terdapat perubahan dataset di masa mendatang. Proses data *cleaning* yang diterapkan pada dataset dapat ditunjukkan Gambar 3.

Sebelum Data Cleaning		
Nama Variabel	Non-Missing Values	Missing Values
age	299	0
anaemia	299	0
creatinine_phosphokinase	299	0
diabetes	299	0
ejection_fraction	299	0
high_blood_pressure	299	0
platelets	299	0
serum_creatinine	299	0
serum_sodium	299	0
sex	299	0
smoking	299	0
time	299	0
death_event	299	0
Setelah Data Cleaning		
Nama Variabel	Non-Missing Values	Missing Values
age	299	0
anaemia	299	0
creatinine_phosphokinase	299	0
diabetes	299	0
ejection_fraction	299	0
high_blood_pressure	299	0
platelets	299	0
serum_creatinine	299	0
serum_sodium	299	0
sex	299	0
smoking	299	0
time	299	0
death_event	299	0

Gambar 3. Cleaning Dataset

Setelah melalui proses data *cleaning* yang pada Gambar 3, Maka akan dilakukan pada tahap data *transformation*, langkah ini dilakukan untuk mengganti nilai kosong dengan rata-rata pada atribut numerik dan dengan mode pada atribut kategorikal. Transformasi ini bertujuan untuk memastikan data konsisten dan dapat diproses oleh algoritma *machine learning* yang tidak mendukung nilai kosong.

Setelah dilakukan data transformation, dataset berada dalam kondisi siap diproses oleh algoritma dapat ditunjukkan Gambar 4.

age	anaemia	creatinine	diabetes	ejection	high	platelets	serum	serum	sex	smoking	time	DEATH
		phosphokinase		fraction	blood_pressure		creatinine	sodium				EVENT
1	75	0	582	0	20	1	265000 1.9	130	1	0	4	1
2	55	0	7861	0	38	0	263358.03 1.1	136	1	0	6	1
3	65	0	146	0	20	0	162000 1.3	129	1	1	7	1
4	50	1	111	0	20	0	210000 1.9	137	1	0	7	1
5	65	1	160	1	20	0	327000 2.7	116	0	0	8	1
6	90	1	47	0	40	1	204000 2.1	132	1	1	8	1
7	75	1	246	0	15	0	127000 1.2	137	1	0	10	1
8	60	1	315	1	60	0	454000 1.1	131	1	1	10	1
--	--	--	--	--	--	--	--	--	--	--	--	--
297	80	1	123	0	35	1	388000 9.4	133	1	1	10	1
298	75	1	81	0	38	1	368000 4	131	1	1	10	1
299	62	0	231	0	25	1	253000 0.9	140	1	1	10	1

Gambar 4. Dataset Sebelum dan Sesudah Data Transformation

Pada Gambar 4, langkah *Replace Missing Value* dengan perlakuan data *transformation* diterapkan untuk menggantikan nilai kosong jika ada. Setelah dilakukan pemeriksaan pada dataset, tidak ditemukan nilai kosong, sehingga proses ini tidak mengubah data. Meskipun demikian, langkah ini tetap disertakan dalam *pipeline preprocessing* sebagai bagian dari prosedur standar untuk memastikan kesiapan data dalam berbagai kondisi. Langkah-langkah *preprocessing* ini membantu memastikan bahwa data bebas dari inkonsistensi dan siap untuk proses pelatihan dan pengujian model.

3.3. Model SVM dan Forward Selection

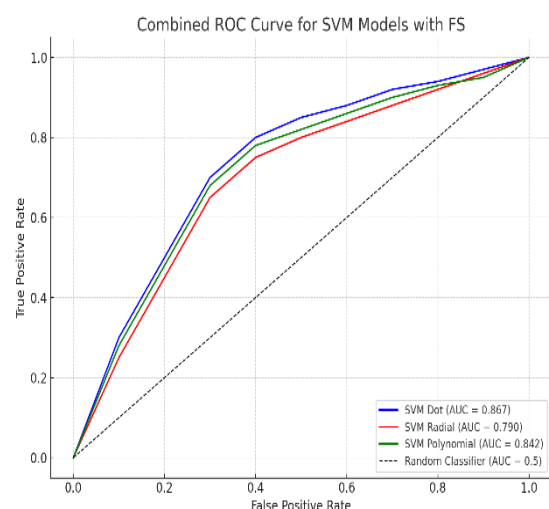
Pada tahap ini, model SVM diterapkan dengan tiga kernel berbeda dot, radial, dan polynomial. Setiap kernel memiliki karakteristik unik dalam mentransformasikan data ke dalam ruang dimensi yang lebih tinggi. Matriks kebingungan untuk model SVM dapat ditunjukkan pada Tabel 1.

TABEL I. CONFUSION MATRIX SVM

Kernel		True 1	True 0
Dot	Pred. 1	57	21
	Pred. 0	39	182
Radial	Pred. 1	8	3
	Pred. 0	88	200
Polynomial	Pred. 1	36	11
	Pred. 0	60	192

Berdasarkan Tabel I, performa model SVM dievaluasi dengan tiga jenis kernel dot (*linear kernel*), *radial* (*RBF kernel*), dan *polynomial kernel*. Setiap kernel menunjukkan akurasi dan kesalahan prediksi yang bervariasi dalam mendeteksi pasien gagal jantung. Kernel dot memiliki performa yang cukup baik dengan *True Negative* tinggi, namun jumlah *False Negative* (39) masih signifikan, menunjukkan beberapa kasus gagal jantung tidak terdeteksi dengan baik. Kernel radial memiliki *True Negative* tertinggi (200), tetapi performa untuk mendeteksi *True Positive* rendah (hanya 8). Hal ini mengindikasikan bahwa kernel radial cenderung lebih konservatif dalam memprediksi kasus gagal jantung. Kernel polynomial menunjukkan keseimbangan yang lebih baik antara *True Positive* (36) dan *True Negative* (192). Namun, jumlah *False Negative* masih cukup tinggi (60), yang menunjukkan perlunya optimasi lebih lanjut.

Sebagai langkah awal, model prediksi gagal jantung menggunakan SVM dievaluasi. Tujuannya adalah untuk memahami performa dasar dari setiap kernel SVM, termasuk kernel *Dot*, *Radial*, dan *Polynomial*. Grafik Receiver Operating Characteristic (ROC) memperlihatkan perbandingan performa model SVM tanpa seleksi fitur, dengan area di bawah kurva AUC yang digunakan sebagai metrik evaluasi.



Gambar 5. Kurva AUC SVM

Berdasarkan grafik Gambar 5 kernel SVM Dot menunjukkan performa dasar terbaik dengan AUC sebesar 0.867, diikuti oleh kernel Polynomial dengan AUC 0.842, dan kernel Radial dengan AUC 0.790. Meskipun hasil ini sudah cukup baik, masih ada peluang untuk meningkatkan performa model melalui teknik seleksi fitur. Oleh karena itu, tahap berikutnya adalah menerapkan teknik FS untuk memilih subset fitur terbaik yang lebih relevan, sehingga diharapkan dapat meningkatkan akurasi prediksi model dan meminimalkan kesalahan prediksi, khususnya pada kasus gagal jantung yang tidak terdeteksi (*False Negative*). Setelah diterapkan FS, bobot atribut yang diperoleh melalui penerapan teknik FS pada ketiga kernel SVM dapat dilihat Tabel II.

TABEL II. BOBOT ATRIBUT

Atribut	Bobot FS+SV M Dot	Bobot FS+SV M Radial	Bobot FS+SV M Polynomial
age	0	0	0
anaemia	0	0	1
creatinine_phosphokinase	0	0	1
diabetes	0	0	1
ejection_fraction	0	0	0
high_blood_pressure	0	0	1
platelets	0	0	0
serum_creatinine	0	1	1
serum_sodium	0	0	0
sex	0	0	1
smoking	0	0	0
time	1	1	1

Berdasarkan Tabel II, bobot atribut yang diperoleh setelah penerapan teknik FS pada ketiga kernel SVM (*dot*, *radial*, dan *polynomial*) menunjukkan perbedaan signifikan dalam pemilihan atribut yang berkontribusi terhadap prediksi gagal jantung. Pada kernel *dot* waktu pemantauan (*time*) merupakan faktor paling penting dalam menentukan risiko gagal jantung ketika menggunakan kernel *dot*. Atribut lain dianggap tidak berpengaruh signifikan. Kernel radial menekankan pentingnya kadar kreatinin dalam darah sebagai indikator utama risiko gagal jantung, diikuti oleh waktu pemantauan. Hal ini mengindikasikan bahwa kombinasi kedua atribut ini memberikan kontribusi dominan dalam model prediksi. Kernel polynomial cenderung mengakomodasi lebih banyak atribut sebagai faktor prediktif. Atribut *time* tetap menjadi yang paling konsisten dan signifikan di semua kernel, sementara atribut lain seperti *anaemia* dan *serum_creatinine* juga memiliki pengaruh penting.

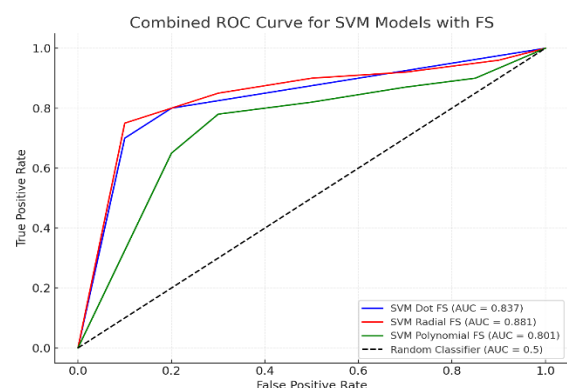
Setelah penerapan teknik FS, hasil matriks kebingungan model SVM dengan FS ditampilkan pada Tabel III.

TABEL III. CONFUSION MATRIX SVM+FS

Kernel		True 1	True 0
Dot	Pred. 1	59	10
	Pred. 0	37	193
Radial	Pred. 1	65	15
	Pred. 0	31	188
Polynomial	Pred. 1	45	9
	Pred. 0	51	194

Berdasarkan Tabel III, penerapan teknik FS pada model SVM dengan ketiga kernel (*dot*, *radial*, dan *polynomial*) menunjukkan peningkatan performa dibandingkan sebelum FS diterapkan. Dengan teknik FS, kernel *dot* berhasil mengurangi *False Positive* secara signifikan dari 21 menjadi 10. Namun, jumlah *False Negative* masih cukup tinggi (37), menunjukkan masih ada beberapa kasus gagal jantung yang tidak terdeteksi. Kernel radial menunjukkan peningkatan signifikan pada *True Positive* (dari 8 menjadi 65), menunjukkan kemampuan model dalam mendeteksi kasus gagal jantung meningkat drastis. Namun, *False Positive* sedikit bertambah (15), yang merupakan trade-off dari peningkatan sensitivitas model. Kernel polynomial mempertahankan *False Positive* yang rendah (9), namun memiliki *False Negative* yang lebih tinggi (51). Hal ini menunjukkan bahwa meskipun kernel ini memiliki akurasi tinggi dalam memprediksi kelas negatif, performa dalam mendeteksi kasus positif masih perlu dioptimalkan.

Berdasarkan analisis *Confusion Matrix* SVM dengan FS pada Tabel III, performa model SVM dengan tiga kernel menunjukkan variasi dalam mendeteksi kasus gagal jantung. Untuk memahami performa secara keseluruhan, akan diperlihatkan kurva ROC untuk ketiga kernel SVM (*dot*, *radial*, dan *polynomial*) pada Gambar 6.



Gambar 6. Kurva AUC SVM+FS

Berdasarkan hasil ROC Curve dan nilai AUC pada Gambar 6. Kernel radial menunjukkan performa terbaik dengan AUC 0.881, menjadikannya pilihan yang optimal untuk prediksi gagal jantung. Kernel dot memberikan keseimbangan performa dengan AUC 0.837, tetapi perlu perbaikan dalam mendeteksi kasus positif. Kernel polynomial memiliki performa paling rendah (AUC 0.801), menunjukkan bahwa model ini masih perlu optimasi lebih lanjut dalam mendeteksi *True Positive*.

3.4. Evaluasi

Untuk mengevaluasi performa model, dilakukan perbandingan antara SVM sebelum dan sesudah penerapan teknik FS. Evaluasi dilakukan dengan menggunakan empat metrik utama yang dapat dilihat pada Tabel IV dengan nilai berdasarkan persen (%).

TABEL IV. PERBANDINGAN SVM DAN SVM+FS

Algoritma	Accuracy	Precision	Recall	AUC
SVM (Dot)	79,92	82,89	89,71	0,867
SVM (Radial)	69,57	69,52	98,55	0,790
SVM (Polynomial)	76,25	76,54	94,60	0,842
SVM (Dot)+ FS	84,30	84,19	95,12	0,837
SVM (Radial)+ FS	84,64	86,51	92,55	0,881
SVM (Polynomial)+FS	79,93	79,50	95,57	0,801

Berdasarkan Tabel IV, dapat dilihat bahwa penerapan teknik FS berhasil meningkatkan performa model SVM untuk ketiga jenis kernel *Dot*, *Radial*, dan *Polynomial*. Peningkatan ini terlihat jelas pada seluruh metrik, yang menjadi indikator utama dalam mengevaluasi performa model. SVM kernel dot sebelum FS Accuracy sebesar 79,92%, Precision 82,89%, Recall 89,71%, dan AUC 0,867. Setelah FS meningkat dengan Accuracy menjadi 84,30%, dengan Precision 84,19%, Recall 95,12%, dan AUC 0,837. SVM kernel *radial* sebelum FS Accuracy hanya 69,57%, Precision 69,52%, dan Recall 98,55%, AUC 0,790. Setelah FS meningkat signifikan menjadi Accuracy 84,64%, Precision 86,51%, Recall 92,55%, dan AUC 0,881. SVM kernel polynomial sebelum FS Accuracy 76,25%, Precision 76,54%, Recall 94,60%, dan AUC 0,842. Setelah FS meningkat

menjadi Accuracy 79,93%, Precision 79,50%, Recall 95,57%, dan AUC 0,801.

Hasil ini menunjukkan bahwa kombinasi Support Vector Machine (SVM) dengan teknik Forward Selection dapat mengoptimalkan prediksi gagal jantung secara lebih akurat dan efisien. Peningkatan Recall dan AUC menunjukkan potensi besar model ini dalam aplikasi klinis untuk mendeteksi kondisi gagal jantung secara dini dan lebih tepat. Radial Kernel dengan FS memiliki performa terbaik karena Radial Basis Function dapat menangkap hubungan non-linear yang kompleks dalam data gagal jantung. Data klinis sering kali memiliki distribusi yang tidak linear, sehingga kernel ini mampu menciptakan pemisahan kelas yang lebih efektif dibandingkan kernel Dot dan Polynomial. Selain itu, kombinasi dengan Forward Selection (FS) memungkinkan model untuk menggunakan hanya fitur yang paling signifikan, sehingga mengurangi dimensi data, meningkatkan akurasi, serta mengoptimalkan Recall dan AUC, yang sangat penting dalam mendeteksi gagal jantung secara dini. Dibandingkan dengan penelitian sebelumnya menggunakan SVM+PSO dengan hasil akurasi 83,97%, penelitian ini menunjukkan bahwa kombinasi SVM dengan FS memberikan peningkatan signifikan dalam deteksi gagal jantung.

4. KESIMPULAN DAN SARAN

Kombinasi SVM dengan FS berhasil mengoptimalkan performa model dalam prediksi gagal jantung. Kernel Radial menunjukkan performa terbaik setelah penerapan FS dengan AUC tertinggi sebesar 0.881, Accuracy 84,64%, Precision 86,51%, dan Recall 92,55%. Kernel Dot memberikan peningkatan signifikan pada Accuracy (84,30%) dan Recall (95,12%) setelah penerapan FS, meskipun AUC mengalami sedikit penurunan menjadi 0.837. Kernel Polynomial mengalami peningkatan moderat dengan Accuracy 79,93% dan Recall 95,57%, tetapi memiliki AUC lebih rendah (0.801) dibandingkan kernel lainnya. Dari hasil keseluruhan teknik FS terbukti efektif dalam memilih fitur signifikan, seperti *time* dan *serum_creatinine*, yang berkontribusi terhadap peningkatan akurasi model.

Implikasi akademik dari penelitian ini adalah kontribusi dalam pengembangan seleksi fitur yang lebih efisien pada model SVM, terutama dalam menangani data klinis yang kompleks dan non-linear. Dengan menunjukkan keunggulan Radial Kernel dalam menangkap pola hubungan non-linear, penelitian ini memberikan wawasan

penting bagi pengembangan metode machine learning dalam analisis data kesehatan. Implikasi klinis, model ini memiliki potensi besar dalam diagnosis dini gagal jantung, terutama karena Recall yang tinggi meningkatkan deteksi pasien berisiko, sehingga memungkinkan intervensi lebih cepat dan perawatan lebih efektif.

Penelitian lanjutan dapat mengeksplorasi metode seleksi fitur yang lebih optimal, seperti Recursive Feature Elimination (RFE) atau Principal Component Analysis (PCA), untuk membandingkan efektivitasnya dengan Forward Selection dalam meningkatkan akurasi model SVM. Selain itu, integrasi teknik deep learning atau hybrid machine learning dapat dipertimbangkan untuk mengembangkan model prediksi yang lebih adaptif dan akurat, terutama dalam menghadapi kompleksitas data klinis yang lebih besar.

Daftar Pustaka:

- [1] A. Ishaq *et al.*, "Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques," *IEEE Access*, vol. 9, pp. 39707–39716, 2021, doi: 10.1109/ACCESS.2021.3064084.
- [2] World Health Organization (WHO), "Cardiovascular diseases," *Cardiovascular diseases (CVDs)*, 2021. <https://www.who.int/health-topics/cardiovascular-diseases>.
- [3] K. K. R. Indonesia, "Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer," *Kementerian Kesehatan RI*, 2022. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220929/0541166/penyakit-jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer/>.
- [4] R. K. Sachdeva, K. D. Singh, S. Sharma, P. Bathla, and V. Solanki, "An Organized Method for Heart Failure Classification," 2023, doi: 10.1109/ESCI56872.2023.10099809.
- [5] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.
- [6] A. Setiawan, R. Febrio Waleska, M. Adji Purnama, Rahmaddeni, and L. Efrizoni, "Komparasi algoritma K-Nearest Neighbor(K-Nn), Support Vector Machine(Svm), Dan Decision Treedalam Klasifikasi Penyakit Stroke," *J. Inform. Rekayasa Elektronik*, vol. 7, no. 1, pp. 107–114, 2024, [Online]. Available: <http://e-journal.stmiklombok.ac.id/index.php/jire> ISSN.2620-6900.
- [7] M. Mamun, A. Farjana, M. Al Mamun, M. S. Ahammed, and M. M. Rahman, "Heart failure survival prediction using machine learning algorithm: am I safe from heart failure?," 2022, doi: 10.1109/AIIoT54504.2022.9817303.
- [8] D. Chicco and L. Oneto, "An Enhanced Random Forests Approach to Predict Heart Failure From Small Imbalanced Gene Expression Data," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 18, no. 6, 2021, doi: 10.1109/TCBB.2020.3041527.
- [9] M. Pal and S. Parija, "Prediction of Heart Diseases using Random Forest," *J. Phys. Conf. Ser.*, vol. 1817, no. 1, 2021, doi: 10.1088/1742-6596/1817/1/012009.
- [10] R. A. M. Elghalid, A. Alwirshiffani, A. A. I. Mohamed, F. H. A. Aldeeb, and A. Andiasha, "Comparison of Some Machine Learning Algorithms for Predicting Heart Failure," 2022, doi: 10.1109/ICEMIS56295.2022.9914325.
- [11] D. Yulindari, A. Wuryanto, F. A. Sariasih, Sidik, and F. A. Sariasih, "Improving the Accuracy of Heart Failure Prediction Using the Particle Swarm Optimization Method," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 1, pp. 210–220, 2024, doi: <https://doi.org/10.33395/sinkron.v9i1.13017> e-ISSN.
- [12] I. Nawawi, "OPTIMISASI PEMILIHAN FITUR UNTUK PREDIKSI GAGAL JANTUNG: FUSION RANDOM FOREST DAN PARTICLE SWARM OPTIMIZATION," *INTI NUSA MANDIRI*, vol. 18, no. 2, pp. 122–128, 2024, doi: DOI: <https://doi.org/10.33480/inti.v18i2.5031>.
- [13] Sumarna, Sartini, W. E. Pangesti, R. Suryadithia, and V. Riyanto, "Decision Tree Optimization in Heart Failure Diagnostics: a Particle Swarm Optimization Approach," *J. Tek. Inform.*, vol. 5, no. 3, pp. 739–746, 2024, doi: <https://doi.org/10.52436/1.jutif.2024.5.3.1815>.
- [14] A. Hamid and Ridwansyah, "Optimizing Heart Failure Detection: A Comparison between Naive Bayes and Particle Swarm Optimization," *Paradigma*, vol. 26, no. 1, pp. 30–36, 2024, doi: <https://doi.org/10.31294/p.v26i1.3284>.

- [15] C. M. Zakariya *et al.*, "ANALISIS SENTIMEN TERKAIT PROGRAM KARTU PRAKERJA MENGGUNAKAN METODE K-NEAREST NEIGHBORS," *JIRE (Jurnal Inform. Rekayasa Elektron.*, vol. 7, no. 2, pp. 183–192, 2024.
- [16] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risandar, "Klasifikasi Mahasiswa Her Berbasis Algortima Svm Dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2020, doi: 10.25126/jtiik.202073080.
- [17] G. Wijaya, "Improvement of Kernel SVM to Enhance Accuracy in Chronic Kidney Disease," vol. 9, no. 1, pp. 136–144, 2024, doi: <https://doi.org/10.33395/sinkron.v9i1.13112> e-ISSN.
- [18] I. Iswanto, T. Tulus, and P. Poltak, "Comparison of Feature Selection To Performance Improvement of K-Nearest Neighbor Algorithm in Data Classification," *J. Tek. Inform.*, vol. 3, no. 6, pp. 1709–1716, 2022, doi: 10.20884/1.jutif.2022.3.6.471.
- [19] R. Ridwansyah, G. Wijaya, and J. J. Purnama, "Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 53–58, 2020, doi: 10.33480/pilar.v16i1.1180.
- [20] H. Nurdin, I. Carolina, R. L. Andharsaputri, A. Wuryanto, and Ridwansyah, "Forward Selection as a Feature Selection Method in the SVM Kernel for Student Graduation Data," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. October, pp. 2531–2537, 2024, doi: 10.33395/sinkron.v8i4.14172.
- [21] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, pp. 1–16, 2020, doi: 10.1186/s12911-020-1023-5.
- [22] I. Ariyati, Ridwansyah, and Suhardjono, "Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen," *J. Inform. dan Komput.*, vol. 3, no. 2, pp. 70–75, 2018, [Online]. Available: <https://ejournal.akakom.ac.id/index.php/jiko/article/view/127/pdf>.
- [23] Ridwansyah, M. Iqbal, H. Destiana, Sugiono, and A. Hamid, "Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan," *semanTIK*, vol. 10, no. 2, pp. 1–10, 2024, doi: <https://doi.org/10.55679/semantik.v10i2.67>.
- [24] V. Riyanto, Imam Nawawi, R. Ridwansyah, Ganda Wijaya, and Toto Haryanto, "Optimization of the YOLOv7 object detection algorithm for estimating the amount of apple harvest," *Paradigma*, vol. 25, no. 1, pp. 31–34, 2023, doi: <https://doi.org/10.31294/p.v25i1.1809>.