

SELEKSI FITUR ALGORITMA GENETIKA DALAM KLASIFIKASI DATA REKAM MEDIS PCOS MENGGUNAKAN SVM

FEATURE SELECTION USING GENETIC ALGORITHM IN PCOS MEDICAL RECORD DATA CLASSIFICATION USING SVM

Fahriza Novianti¹⁾, Nurissaidah Ulinnuha²⁾

^{1), 2)} Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel

Jl Gn. Anyar, Surabaya, Jawa Timur, Indonesia

Email : rizanv19@gmail.com¹⁾, nuris.ulinnuha@uinsa.ac.id²⁾

Abstrak

Ketidakeimbangan hormon menyebabkan seorang wanita dengan sindrom ovarium polikistik (PCOS) memiliki sel telur atau sel telur yang tidak matang secara normal. Biasanya terjadi pada masa reproduksi, namun seringkali sulit dideteksi karena kurangnya kesadaran. Oleh karena itu, penting untuk mendeteksi kondisi ini secara dini agar dapat melakukan pengobatan atau pencegahan yang tepat. Salah satu cara untuk mendiagnosis PCOS adalah melalui penggunaan data medis. Dalam penelitian ini, 40 variabel digunakan, termasuk data hormonal, hasil ultrasonografi, dan informasi medis lainnya. Metode yang digunakan adalah Support Vector Machine (SVM), yang mampu menangani data yang tidak linear dengan kernel. Untuk meningkatkan akurasi, fitur-fitur dipilih menggunakan algoritma genetika, yang menghasilkan 19 variabel yang signifikan. Dengan menerapkan variabel hasil seleksi tersebut sebagai input, klasifikasi menghasilkan model terbaik dengan akurasi 94.26%, sensitivitas 87.57%, dan spesifisitas 97.52%. Tanpa proses seleksi fitur, klasifikasi SVM hanya memiliki akurasi 82.46%, sensitivitas 60.91%, dan spesifisitas 97.25%. Dari temuan penelitian ini, dapat diketahui bahwa metode seleksi fitur algoritma genetika dapat meningkatkan performa klasifikasi SVM.

Kata kunci: Algoritma Genetika, Klasifikasi, PCOS, Seleksi Fitur, SVM

Abstract

A hormonal imbalance causes a woman with polycystic ovarian syndrome (PCOS) to have an ovum or egg that does not mature normally. It usually occurs during the reproductive period, but is often difficult to detect due to lack of awareness. Therefore, it is important to detect this condition early so that proper treatment or prevention can be done. One way to diagnose PCOS is through the use of medical data. In this study, 40 variables were used, including hormonal data, ultrasound results, and other medical information. The method used was Support Vector Machine (SVM), which is able to handle non-linear data with a kernel. To improve accuracy, features were selected using a genetic algorithm, which resulted in 19 significant variables. By applying the selected variables as input, the classification produced the best model with 94.26% accuracy, 87.57% sensitivity, and 97.52% specificity. Without the feature selection process, SVM classification only has an accuracy of 82.46%, sensitivity of 60.91%, and specificity of 97.25%. From the findings of this research, it can be seen that the genetic algorithm feature selection method can improve SVM classification performance.

Keywords: Genetic Algorithm, Classification, PCOS, Feature Selection, SVM.

1. PENDAHULUAN

Polycystic Ovary Syndrome (PCOS) adalah penyakit medis yang mempengaruhi wanita yang ditandai dengan ketidakseimbangan hormon, menstruasi yang tidak teratur, dan terjadinya polikistik ovarium (PCO), yaitu ovum yang tidak matang yang terbentuk dalam ovarium. PCOS juga dapat menyebabkan masalah kesehatan lain, seperti obesitas, diabetes tipe 2, dan penyakit kardiovaskular [1].

Dalam diagnosis dan pengobatan PCOS, analisis data rekam medis dapat menjadi sangat penting. Namun, data medis biasanya terdiri dari banyak fitur atau atribut yang mungkin tidak semuanya relevan dalam diagnosis PCOS. Oleh karena itu, teknik seleksi fitur dapat diimplementasikan untuk mereduksi ukuran data dengan memilih fitur yang paling penting dan relevan [2].

Salah satu teknik seleksi fitur yang populer adalah Algoritma Genetika. Ide dasar di balik algoritma genetika adalah proses seleksi alam yang diperlukan untuk evolusi. Algoritma ini dapat dipakai untuk mencari solusi terbaik dari suatu masalah optimasi dengan cara memanipulasi populasi genetik dengan mempertahankan individu-individu terbaik dan memperkenalkan variasi genetik baru melalui proses seleksi, crossover, dan mutasi [3].

Dalam klasifikasi data rekam medis PCOS, *Support Vector Machine* (SVM) dapat digunakan sebagai algoritma klasifikasi. SVM adalah algoritma pembelajaran mesin yang dapat membangun model untuk memprediksi kelas target dari data yang diberikan [4]. SVM mencari hyperplane yang memaksimalkan margin antara dua kelas target, sehingga dapat mengklasifikasikan data dengan lebih akurat [5].

Penelitian yang pernah dilakukan dengan mengimplementasikan seleksi fitur algoritma genetika dengan SVM dan dibandingkan dengan Naive Bayes dalam menganalisis sentimen Calon Gubernur Jawa Barat periode 2018-2023. Hasilnya menunjukkan bahwa SVM dan algoritma genetika mencapai akurasi 93.03%, sedangkan metode gabungan Naive Bayes dengan algoritma genetika mencapai akurasi 92% [6]. Terbukti kedua metode tersebut mampu menghasilkan akurasi yang baik dalam melakukan klasifikasi. Namun, belum pernah dilakukan penelitian sebelumnya yang menggunakan algoritma genetika sebagai seleksi fitur pada metode SVM untuk mendiagnosis PCOS.

Telah dilakukan penelitian untuk menganalisis sentimen calon gubernur Jawa Barat periode 2018-2023 menggunakan metode seleksi fitur algoritma genetika dengan SVM yang dibandingkan dengan Naive Bayes. Menurut temuannya, SVM dan algoritma genetika menghasilkan akurasi 93,03%, sementara Naive Bayes dan algoritma genetika yang menghasilkan akurasi 92% [6]. Telah dibuktikan bahwa kedua pendekatan tersebut dapat memberikan hasil klasifikasi dengan akurasi yang baik. Sayangnya, belum ada penelitian sebelumnya yang menggunakan algoritma genetika sebagai pengganti seleksi fitur dalam pendekatan SVM untuk mendiagnosis PCOS.

Dengan melakukan seleksi fitur algoritma genetika pada metode SVM, penulis berharap dapat mencapai tingkat akurasi yang lebih baik dalam mendiagnosis PCOS berdasarkan informasi kesehatan pasien. Tujuan dari penelitian ini yaitu meningkatkan kinerja metode SVM dalam mendiagnosis PCOS dengan digunakannya teknik seleksi fitur algoritma genetika. Diharapkan bahwa hasil penelitian ini akan memberikan kontribusi dalam pengembangan metode diagnostik yang lebih efektif dan akurat untuk penyakit PCOS.

2. DASAR TEORI

2.1 Algoritma Genetika

Untuk memecahkan masalah optimasi, algoritma genetika memanfaatkan seleksi alam, dimana individu terbaik dipilih untuk menciptakan generasi yang lebih unggul [7]. Solusi terbaik algoritma genetika didapatkan dengan menggunakan nilai *fitness* yang dievaluasi melalui serangkaian tahapan yaitu seleksi, mutasi, crossover, dan elitisme, yang diulang hingga mencapai nilai fitness yang diharapkan [8]. Penjelasan mengenai proses algoritma genetika yaitu:

a. Inisialisasi Populasi

Teknik inisialisasi adalah pendefinisian salah satu solusi dari suatu permasalahan. Teknik ini melibatkan penentuan tipe dan jumlah gen yang akan digunakan untuk mewakili solusi dari permasalahan yang diteliti. Gen, yang membentuk sebuah kromosom, dapat memiliki nilai biner, simbolik, numerik, atau karakteristik, tergantung pada masalah yang harus diselesaikan. Kromosom akan terus berubah, sebuah proses yang dikenal sebagai generasi [9].

b. Pembangkitan Populasi Awal

Populasi di generasi pertama algoritma genetika dihasilkan secara acak. Populasi tersusun atas individu yang mewakili solusi yang diharapkan [10]. Individu tersusun atas sekumpulan kromosom, sedangkan kromosom tersusun atas sekumpulan gen, dan nilai dari sebuah gen disebut alel.

c. Perhitungan Nilai Fitness

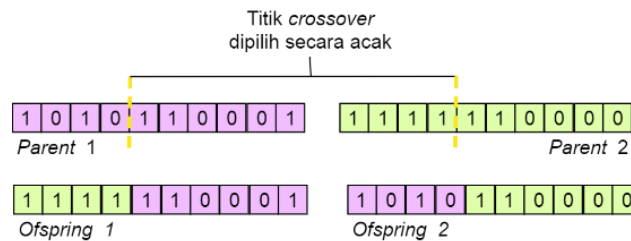
Fungsi fitness dipergunakan untuk memberi penilaian pada suatu permasalahan agar tercapai nilai optimal dalam algoritma genetika [11]. Fungsi fitness yang digunakan adalah nilai akurasi dalam metode klasifikasi. Semakin tinggi nilai fitness atau mendekati 100, maka performa algoritma genetika disebut semakin optimal.

d. Seleksi

Seleksi adalah suatu teknik yang digunakan untuk memilih orangtua dalam proses reproduksi. Pemilihan ini didasarkan pada nilai kebugaran individu. Tujuan dari seleksi adalah memberikan peluang terbesar kepada anggota populasi terbaik untuk melakukan reproduksi [11]. Salah satu teknik seleksi yang sering digunakan adalah metode mesin *roulette*.

e. *Crossover*

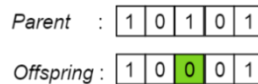
Selama proses *crossover*, dua kromosom induk digabungkan agar terbentuk suatu kromosom baru. Terbentuknya kromosom baru ini melibatkan penggabungan informasi oleh setiap induk kromosom. Hasil dari *crossover* ialah kromosom baru yang disebut dengan *offspring* (kromosom anak). Tujuan dari *crossover* adalah untuk menghasilkan replika individu yang memiliki tingkat signifikansi yang lebih tinggi dalam populasi tanpa menghasilkan variasi yang baru [12]. Ilustrasi proses *crossover* diperlihatkan pada Gambar 1.



Gambar 1. Ilustrasi Proses *Crossover*

f. Mutasi

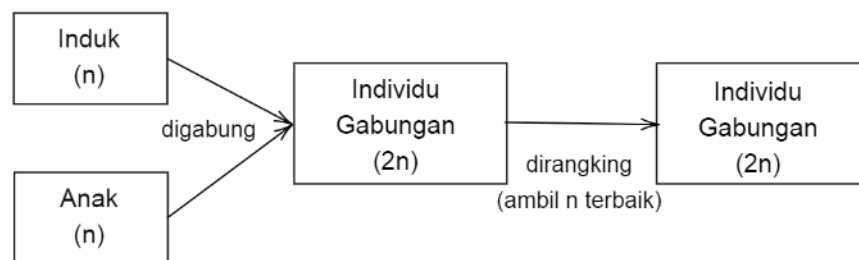
Dengan menggunakan pendekatan mutasi, nilai dari satu atau lebih gen kromosom diubah secara acak. Individu dengan kualitas di atas rata-rata menjadi tujuan dari mutasi. Selain itu, kerusakan yang diakibatkan oleh *crossover* pada materi genetik dapat diperbaiki dengan proses mutasi. Ilustrasi proses mutasi diperlihatkan pada Gambar 2.



Gambar 2. Ilustrasi Proses Mutasi

g. *Elitism*

Sebelum melanjutkan ke langkah iterasi berikutnya, dilakukan proses elitisme. Dalam langkah ini, beberapa kromosom paling baik dari populasi sekarang (awal) disalin ke populasi baru dengan mempergunakan sistem pemilihan. [11]. Sistem ini akan menyeleksi sebanyak n individu yang memiliki nilai *fitness* terbaik dalam suatu populasi untuk dilanjutkan ke iterasi berikutnya [13]. Hal tersebut mengakibatkan populasi baru dalam pengulangan selanjutnya tetap sama meskipun proses reproduksi individu baru sudah dilakukan. Gambar 3 mengilustrasikan proses *elitism*.



Gambar 3. Ilustrasi Proses *Elitism*

2.2 Support Vector Machine (SVM)

Ide di balik SVM adalah menemukan *hyperplane* yang optimal dalam ruang input untuk bertindak sebagai pemisah antara dua kelas data [14]. *Hyperplane*, yang ditentukan dengan menghitung margin dan mencari titik maksimumnya, adalah pemisah yang ideal antara dua kelas. Margin mengukur jarak antara titik data terdekat dari setiap kelas dengan *hyperplane* (*support vector*).

Data yang terdapat pada training data set dilambangkan dengan $x_u \in R^D$ sedangkan label kelas dinyatakan dengan $y_i = -1, +1$ untuk $i = 1, 2, 3, \dots, N$ dimana N adalah jumlah data training. Kelas -1 dan +1 dipisahkan dengan sempurna oleh *hyperplane* berdimensi D . *hyperplane* pada SVM dilambangkan dengan persamaan (1).

$$w \cdot x_u + b = 0 \tag{1}$$

Di mana w dan b adalah parameter model, dan $w \cdot x_u$, adalah hasil kali dalam antara w dan x . Secara matematis, formulasi optimasi SVM untuk kasus klasifikasi linear pada ruang primal menggunakan Persamaan (2).

$$\min \tau(w) = \frac{1}{2} \|w\|^2 \tag{2}$$

Persamaan (2) terbatas pada pertidaksamaan berikut,

$$y_u (w \cdot x_u + b) \geq 1, \quad i = 1, 2, 3, \dots, N$$

Sulit secara komputasi untuk menyelesaikan klasifikasi nonlinear serta waktu yang diperlukan lebih lama. SVM nonlinier umumnya menggunakan pendekatan kernel untuk fitur dataset. Konsep dari fungsi kernel ialah mentransformasi data menjadi berdimensi lebih tinggi [15]. Penelitian ini menggunakan kernel pada SVM yang diperlihatkan pada Tabel 1.

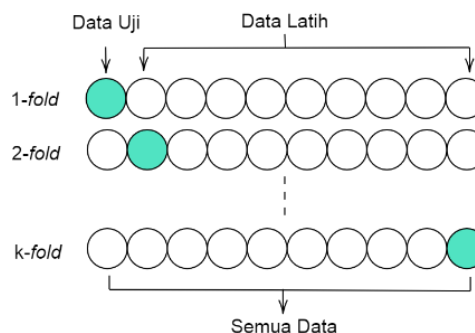
Tabel 1. Kernel pada SVM

Jenis Kernel	Persamaan
Linier	$x \cdot x^T$
Polynomial	$(x \cdot x^T + r)^d$
RBF	$\exp(-\gamma \ x - x_i\ ^2)$

2.3 K-Fold Cross Validation

Salah satu cara untuk memisahkan data adalah dengan menggunakan *K-fold cross validation*. Dengan menilai efisiensi dan akurasi model, metode ini berusaha untuk menentukan akurasi model klasifikasi [16]. Data dipisahkan menjadi data pelatihan dan pengujian selama fase pelatihan. Setelah melakukan pelatihan pada data training, model diverifikasi sebanyak k kali pada data testing, di mana k adalah jumlah fold. Berikut ini adalah langkah-langkah yang terlibat dalam proses ini:

1. Dilakukan pembagian data menjadi k subset berukuran sama, dengan satu subset digunakan untuk pengujian dan subset lainnya untuk pelatihan. Gambar 4 menunjukkan ilustrasi dari *k-fold cross validation*.



Gambar 4. Gambaran pembagian data pada *k-fold cross validation*

2. Pada fold pertama, data pertama digunakan sebagai data pengujian, sementara bagian yang lain digunakan sebagai data pelatihan. Setelah itu, akurasi dihitung pada fold tersebut.
3. Langkah 1 dan 2 diulangi sejumlah k. Skor model diperoleh dengan mengkalkulasi rata-rata akurasi dari masing-masing fold.

2.4 Confussion Matrix

Confusion matrix digunakan untuk menampilkan tingkat akurasi yang dihasilkan proses klasifikasi. Prinsip dasar dari metode ini yakni membandingkan hasil prediksi model dengan data klasifikasi sebenarnya. Gambaran *confusion matrix* dapat dilihat pada Tabel 2.

Tabel 2. *Confusion Matrix*

Kelas Sebenarnya	y_u	Kelas Prediksi	
	Normal	TN	FP
	PCOS	FN	TP

True Positive (TP) adalah ketika model secara akurat mengidentifikasi kelas positif sebagai kelas positif. Sebaliknya, *True Negative* (TN) adalah ketika model secara akurat mengidentifikasi kelas negatif sebagai kelas negatif. *False Positive* (FP) terjadi ketika model secara salah mengidentifikasi kelas negatif menjadi kelas positif. Sementara itu, *False Negative* (FN) terjadi ketika model salah mengidentifikasi kelas positif menjadi kelas negatif [17]. Indikator yang digunakan untuk menilai hasil klasifikasi yakni:

1. Akurasi,

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3}$$

2. Sensitivitas,

$$Sen = \frac{TP}{TP+FN} \tag{4}$$

3. Spesifisitas,

$$Spe = \frac{TN}{TN+FP} \tag{5}$$

3. METODOLOGI PENELITIAN

Laman website Kaggle Dataset menjadi penyedia data yang dipakai pada penelitian ini. Data yang dipakai diambil dari sepuluh rumah sakit Kerala yang berada di India. Data tersebut mencakup karakteristik klinis dan fisik dari 540 orang dengan dan tanpa masalah reproduksi, yang berkaitan dengan PCOS (*Polycystic Ovary Syndrome*). Terdapat 40 variabel independen dalam data tersebut. Tabel 3 mencantumkan variabel input pada penelitian ini.

Tabel 3. Variabel Penelitian

No.	Fitur	Deskripsi
1	Umur	Umur pasien (Tahun)
2	Berat	Berat pasien (kilogram)
3	Tinggi	Tinggi pasien (cm)
4	BMI	Indeks Masa Tubuh (Berat /Tinggi)
5	Denyut nadi	Detak jantung dalam denyut per menit (bpm)
6	RR	Tingkat pernapasan dalam napas per menit
7	HB	Hemoglobin (gram/desiliter)
8	Panjang siklus mens	Durasi hari siklus menstruasi
9	Status nikah	Lama menikah
10	Status hamil	Hamil atau tidak
11	Jumlah aborsi	Berapa kali pasien melakukan aborsi
12	beta-HCG	Jumlah <i>Beta Human Chorionic Gonadotropin</i>
13	FSH	Kadar hormon perangsang folikel (mIU/mL)
14	LH	Kadar hormon <i>luteinizing</i> (mIU/mL)
15	FSH/LH	Rasio FSH terhadap LH

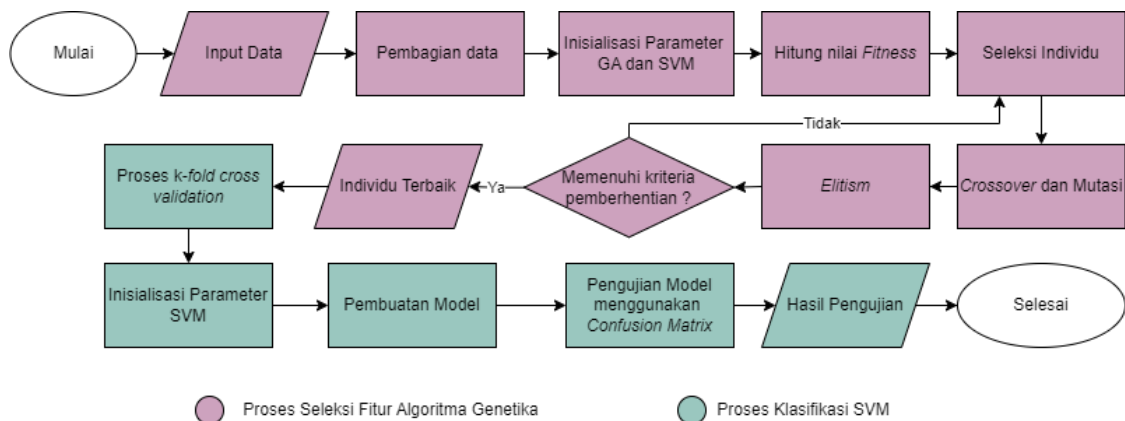
No.	Fitur	Deskripsi
16	Ukuran Pinggul	Ukuran pinggul pasien (inci)
17	Ukuran Pinggang	Ukuran pinggang pasien (nci)
18	WHRatio	Rasio antara pinggang dan pinggul
19	TSH	Kadar hormon perangsang tiroid (mIU/mL)
20	AMH	Kadar hormon <i>anti-Mullerian</i> (ng/mL)
21	PRL	Kadar <i>prolactin</i> (ng/mL)
22	Vit D3	Kadar Vitamin D3 (ng/mL)
23	PRG	Kadar hormon <i>progesteron</i> (ng/mL)
24	RBS	Kadar gula darah acak (mg/dL)
25	Weight Gain (WG)	Berat badan bertambah atau tidak
26	Skin Darkening	Kulit menggelap atau tidak
27	Hair Loss(HL)	Mengalami kerontokan rambut atau tidak
28	Pimples	Muncul jerawat atau tidak
29	Fast Food (FF)	Mengonsumsi <i>fastfood</i> atau tidak
30	Reg. Exercise (RE)	Berolahraga secara teratur atau tidak
31	BPSystolic	Tekanan sistolik (mmHg)
32	BPDiastolic	Tekanan diastolik (mmHg)
33	Follicle No. L	Jumlah folikel pada ovarium kiri
34	Follicle No. R	Jumlah folikel pada ovarium kanan
35	Avg F. Size L	Ukuran rata-rata folikel di ovarium kiri (mm)
36	Avg F. Size R	Ukuran rata-rata folikel di ovarium kiri (mm)
37	Endometrium	Ketebalan <i>endometrium</i> (mm)
38	Blood Group (BG)	Golongan darah (terdiri dari 8 kategori)
39	Cycle	Siklus menstruasi teratur atau tidak teratur
40	Hair Growth (HG)	Mengalami pertumbuhan rambut atau tidak

363 dari 540 pasien memiliki hasil tes PCOS negatif, sisanya 177 orang dengan hasil tes PCOS positif. Tabel 4 menjelaskan ringkasan data sampel.

Tabel 4. Data Sampel Pada Penelitian

Data Ke-	Umur	Tinggi	Berat	...	Cycle	BG	HG	Kelas
1	41	152	63	...	2	12	0	Negatif
2	20	152	42	...	4	17	0	Negatif
3	25	158	76	...	2	15	0	Negatif
4	32	155	68	...	2	17	0	Negatif
5	39	158.5	62	...	4	12	0	Negatif
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
539	28	152	13	...	4	13	0	Positif
540	26	161.5	14	...	2	14	0	Positif

Data Tabel 4 akan dikerjakan sesuai dengan metode yang dipakai. Prosedur ini disajikan pada *flowchart* Gambar 5.



Gambar 5. Flowchart Proses Seleksi Fitur Algoritma Gentika dan Klasifikasi SVM

Uji parameter bertujuan mendapatkan parameter optimal yang menghasilkan nilai evaluasi model terbaik. Ujicoba pada algoritma gentika ditentukan 100 generasi, probabilitas *crossover* 0.8 dan probabilitas mutasi 0.1. Selanjutnya ukuran populasi (*popsize*) diujicoba pada nilai 40, 60, dan 80. Pengujian juga dilakukan pada kernel linear, polynomial, dan *radial basis function* (RBF) di tahap pembentukan model SVM.

4. PENGUJIAN DAN PEMBAHASAN

4.1 Seleksi Fitur Algoritma Genetika

Terdapat tiga jenis parameter ujicoba model pada algoritma gentika yakni ukuran populasi (*popsize* = 40, 60, 80), jumlah generasi ($I = 100$), probabilitas *crossover* ($Cr = 0.8$), dan probabilitas mutasi ($Mr = 0.1$). Kumpulan karakteristik yang paling akurat akan dihasilkan dari pengujian parameter-parameter ini.

Pada Support Vector Machines (SVM), tiga jenis kernel-linear, polynomial, dan kernel RBF-digunakan dalam proses seleksi fitur. Parameter yang digunakan secara default pada kernel linear adalah $C = 1$. Nilai default untuk kernel polinomial adalah $C = 1, \gamma = 0,143, r = 0$, dan $d = 3$. Kernel RBF menggunakan $C = 1$ dan $\gamma = 0.143$ sebagai parameter default. Tabel 5 menampilkan hasil evaluasi parameter algoritma gentika dengan ketiga kernel tersebut.

Pada seleksi fitur menggunakan algoritma gentika, hasil terbaik diperoleh di pembagian data 70:30, dengan *popsize* 80 dan kernel RBF pada SVM. 19 fitur diantaranya: BMI; BG; HB; RR; Status nikah; FSH; Ukuran pinggul; RBS; PRG; TSH; WG; HG; FF; RE; BPDiastolic; Follicle No (L); Follicle No (R); Cycle; dan Endometrium-diidentifikasi sebagai fitur terbaik.

Tabel 5. Pengujian Parameter GA

Data	Kernel	Popsize	Fitur	Fitness	Waktu (s)
70:30	RBF	80	19	89.29	102
		60	16	88.27	92
		40	18	86.11	77
	Polynomial	80	16	85.18	110
		60	23	84.55	94
		40	21	82.71	78
	Linear	80	13	80.57	98
		60	20	79.67	84
		40	25	78.19	69
	Rata-Rata				83.84
80:20	RBF	80	19	88.54	106
		60	13	88.14	99
		40	18	85.19	94
	Polynomial	80	24	81.45	111
		60	25	80.65	90
		40	21	80.13	83
	Linear	80	19	80.71	110
		60	25	77.71	99
		40	24	75.90	79
	Rata-Rata				82.05
90:10	RBF	80	16	86.12	127
		60	19	85.64	110
		40	21	84.41	84
	Polynomial	80	27	80.11	119
		60	23	78.11	99
		40	25	75.05	83
	Linear	80	20	80.42	109

Data	Kernel	Popsi	Fitur	Fitness	Waktu (s)
		60	18	77.64	88
		40	19	73.13	77
Rata-Rata				80.07	99.56

4.2 Klasifikasi SVM dengan Seleksi Fitur Algoritma Genetika

Pembagian data ke dalam kelompok pelatihan dan pengujian dilakukan dengan k-fold cross validation (k = 5 dan 10), klasifikasi SVM mengambil 19 fitur terpilih dari algoritma genetika. Fungsi kernel SVM dan nilai Cost (C) = [0.1,1,10,100] diuji. Tabel 6 menampilkan hasil model GA-SVM.

Tabel 6. Hasil Klasifikasi GA-SVM

K-Fold	Kernel	Cost	Akurasi(%)	Sensitivitas(%)	Spesifisitas(%)	Waktu (s)
5	Linear	0.1	91.11	84.75	94.22	14
		1	91.80	83.05	93.66	
		10	95.50	84.75	93.38	
		100	91.09	84.25	94.49	
	Polynomia 1	0.1	99.30	74.92	98.95	8
		1	91.83	76.49	99.14	
		10	91.74	76.44	99.11	
		100	91.67	76.44	99.09	
	RBF	0.1	91.29	79.09	97.25	12
		1	93.07	86.01	96.52	
		10	92.22	81.36	97.52	
		100	91.11	78.53	97.25	
Rata-rata			91.40	80.51	96.72	11.33
10	Linear	0.1	91.33	85.01	94.38	15
		1	91.04	84.63	94.16	
		10	99.30	84.86	93.89	
		100	91.67	86.44	94.21	
	Polynomia 1	0.1	99.30	81.92	95.31	10
		1	92.04	81.92	96.96	
		10	91.67	82.49	96.14	
		100	91.48	81.36	96.42	
	RBF	0.1	93.81	86.56	97.36	13
		1	94.26	87.57	97.52	
		10	93.89	87.01	97.24	
		100	92.59	81.92	97.79	
Rata-rata			92.14	84.31	95.95	12.67

Berdasarkan hasil klasifikasi yang tertera di Tabel 6, ditemukan model terbaik GA-SVM pada parameter kernel RBF, C=1, dan k-fold = 10. Skor evaluasi model terbaik yang tercantum dalam Tabel 6 diperoleh melalui *confusion matrix*. *Confusion matrix* yang diperoleh dalam klasifikasi GA-SVM ditunjukkan dalam Tabel 7.

Hasil terbaik terlihat pada proses SVM dengan akurasi 94.26%, sensitivitas 87.57% dan spesifitas 97.52%, berdasarkan hasil yang disajikan pada Tabel 6. Tabel 7 menampilkan *confusion matrix* yang dihasilkan dari model terbaik SVM-GA. Berdasarkan data yang diberikan, terdapat 354 kasus normal yang diklasifikasikan dengan benar, 9 kasus normal yang salah diklasifikasikan sebagai PCOS, 22 kasus PCOS yang salah diklasifikasikan sebagai normal, dan 155 kasus PCOS yang diklasifikasikan dengan benar.

Tabel 7. Confusion matrix GA-SVM

Kelas Sebenarnya	y_u	Kelas Prediksi	
	Normal	354	9
	PCOS	22	155

4.3 Klasifikasi SVM tanpa seleksi fitur

Klasifikasi SVM dilakukan dengan menggunakan semua 40 fitur untuk mengevaluasi kinerja model dengan menggunakan semua fitur yang tersedia. Hasil evaluasi model SVM ditampilkan dalam Tabel 8. Untuk menilai apakah algoritma genetika meningkatkan akurasi SVM, klasifikasi SVM dilakukan dengan menggunakan semua 40 fitur. Tabel 8 menampilkan nilai dari evaluasi model SVM.

Tabel 8. Hasil Klasifikasi SVM

K-Fold	Kernel	Cost	Akurasi (%)	Sensitivitas (%)	Spesifisitas (%)	Waktu (s)
5	Linear	1	81.60	64.02	90.77	16
		1	82.39	65.56	90.84	
		10	81.44	62.50	90.77	
		100	81.36	61.65	90.90	
	Polynomial	1	80.28	52.72	92.97	9
		1	82.31	53.52	96.34	
		10	81.28	54.74	94.23	
		100	80.45	49.66	99.09	
	RBF	1	82.80	68.66	90.23	13
		1	83.46	60.53	94.37	
		10	83.12	61.42	93.89	
		100	82.52	66.70	90.72	
Rata-rata			81.92	60.14	92.93	12.67
10	Linear	1	83.11	66.52	91.06	17
		1	83.67	62.50	93.94	
		10	81.63	68.75	87.88	
		100	79.59	62.51	87.87	
	Polynomial	1	81.52	53.18	95.12	11
		1	82.36	54.97	95.37	
		10	82.09	53.54	96.02	
		100	82.00	53.25	96.01	
	RBF	1	83.29	62.83	93.26	14
		1	85.19	60.45	97.25	
		10	82.10	62.88	92.68	
		100	82.93	69.57	92.97	
Rata-rata			82.46	60.91	93.29	14

Serupa dengan klasifikasi seleksi fitur, hasil klasifikasi dengan semua fitur yang ditunjukkan pada Tabel 8 menunjukkan bahwa kernel RBF, C = 1, dan 10-fold cross validation memberikan hasil yang terbaik. Tabel 9 menampilkan *confusion matrix* dari model terbaik klasifikasi SVM.

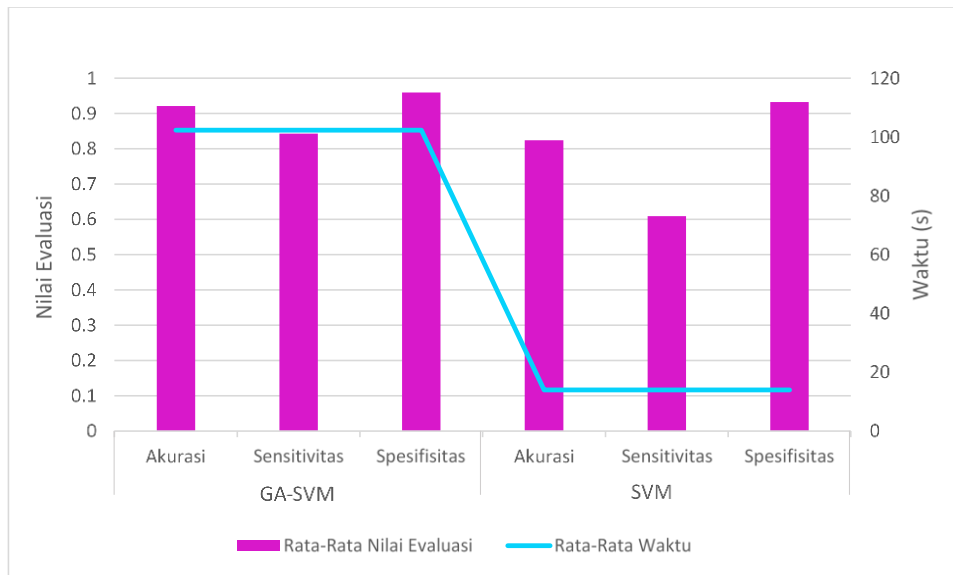
Tabel 9. *Confusion matrix* SVM

Kelas Sebenarnya	y_u	Kelas Prediksi	
	Normal	353	10
PCOS	70	107	

Berdasarkan Tabel 7 dan 9, dari hasil perbandingan *confusion matrix* SVM dan GA-SVM, terlihat bahwa meskipun terdapat sedikit peningkatan pada jumlah *True Negative* pada GA-SVM yang berakibat naiknya akurasi sebesar 9.07% dan spesifitas sebesar 0.27%, namun terdapat peningkatan yang signifikan pada jumlah *True Positive*. Dalam hal ini menyebabkan nilai sensitifitas pada GA-SVM naik sebesar 27.12%, menunjukkan bahwa GA-SVM lebih unggul dalam memprediksi kelas dengan lebih akurat daripada SVM biasa.

4.4 Analisis Hasil

Berdasarkan hasil evaluasi menggunakan metode GA-SVM dan SVM, pembagian data *cross validation* dengan 10-fold memiliki akurasi rata-rata yang lebih tinggi daripada 5-fold. Oleh karena itu, dibandingkan akurasi dan waktu komputasi dari GA-SVM dan SVM dengan menggunakan 10-fold *cross validation*.



Gambar 3. Perbandingan GA-SVM dan SVM

Gambar 3 mengilustrasikan bagaimana penggunaan algoritma genetika untuk seleksi fitur meningkatkan waktu komputasi. Klasifikasi tanpa seleksi fitur memiliki *running time* tidak sampai 20 detik, sedangkan kombinasi klasifikasi-seleksi fitur memiliki *running time* lebih lama sebanyak 100 detik. Rata-rata akurasi, sensitivitas, dan spesifisitas dari pendekatan GA-SVM adalah 92.14%, 84.31%, dan 95.95%. Sementara itu, akurasi, sensitivitas, dan spesifisitas rata-rata dari pendekatan SVM tanpa seleksi fitur adalah 82.46%, 60.91%, dan 93.29%. Akurasi pendekatan GA-SVM meningkat secara signifikan sekitar $\pm 0,1$, atau 10%. Hal ini menunjukkan kinerja yang baik dari algoritma genetika dan SVM.

5. KESIMPULAN

Berdasarkan hasil pemilihan variabel dari algoritma genetika, terdapat 475% variabel terpilih atau sembilan belas variabel yang signifikan yakni: BMI; BG; HB; RR; Status nikah; FSH; Ukuran pinggul; RBS; PRG; TSH; WG; HG; FF; RE; BPDiastolic; Follicle No (L); Follicle No (R); Cycle; dan Endometrium. Setelah dilakukan pengujian, didapatkan bahwa 10-fold, kernel RBF, dan nilai C=1 merupakan parameter model yang paling baik untuk digunakan dalam klasifikasi. Hasil klasifikasi GA-SVM yakni akurasi = 94.26%, sensitivitas = 87.57%, dan spesifisitas = 97.52%, sedangkan hasil klasifikasi SVM adalah akurasi = 82.46%, sensitivitas = 60.91%, dan spesifisitas = 97.25%. Menurut temuan penelitian ini, menggunakan algoritma genetika untuk seleksi fitur dapat meningkatkan nilai evaluasi klasifikasi SVM.

DAFTAR PUSTAKA

- [1] W. Chen and Y. Pang, “Metabolic Syndrome and PCOS: Pathogenesis and the Role of Metabolites,” *Metabolites*, vol. 11, no. 12, 2021, doi: 13390/metabo11120869.
- [2] W. Jia, M. Sun, J. Lian, and S. Hou, “Feature dimensionality reduction: a review,” *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2663–2693, 2022, doi: 11007/s40747-021-00637-x.
- [3] E. P. A. Akhmad, “Evaluasi Telemarketing Kartu Kredit Bank Menggunakan Algoritma Genetika untuk Seleksi Fitur dan Naive Bayes (Evaluation of Bank Credit Card

- Telemarketing Using Genetic Algorithms for Feature Selection and Naive Bayes),” *J. Apl. Pelayaran dan Kepelabuhanan*, vol. 10, no. 1, pp. 12–22, 2019.
- [4] R. Resmiati and T. Arifin, “Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination,” *Sistemasi*, vol. 10, no. 2, p. 381, 2021, doi: 132520/stmsi.v10i2.1238.
- [5] W. Widayani and H. Harliana, “Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa,” *J. Sains dan Inform.*, vol. 7, no. 1, pp. 20–27, 2021, doi: 134128/jsi.v7i1.268.
- [6] D. Gunawan, D. Riana, D. Ardiansyah, F. Akbar, and S. Alfarizi, “Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023,” *J. Tek. Komput. AMIK BSI*, vol. 6, no. 2, pp. 174–180, 2020, doi: 131294/jtk.v4i2.
- [7] H. Harafani and A. Maulana, “Penerapan Algoritma Genetika pada Support Vector Machine Sebagai Pengoptimasi Parameter untuk Memprediksi Kesuburan,” *J. Tek. Inform. STMIK Antar Bangsa*, vol. V, no. 1, pp. 51–59, 2019.
- [8] S. F. Pane, R. Maulana Awangga, E. V. Rahmadani, and S. Permana, “Implementasi Algoritma Genetika Untuk Optimalisasi Pelayanan Kependudukan,” *J. Tekno Insentif*, vol. 13, no. 2, pp. 36–43, 2019, doi: 136787/jti.v13i2.13
- [9] K. Krisnandi and H. Agung, “Implementasi Algoritma Genetika Untuk Memprediksi Waktu Dan Biaya Pengerjaan Proyek Konstruksi,” *J. Ilm. FIFO*, vol. 9, no. 2, p. 90, 2017, doi: 122441/fifo.2017.v9i2.001.
- [10] R. Iriany, A. W. Widodo, and W. F. Mahmudy, “Pemanfaatan Algoritma Genetika Untuk Optimasi 0 / 1 Multi-Dimensional Knapsack Problem Dalam Pendistribusian Produk (Studi Kasus UD . TOSA),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 4, pp. 257–269, 2017.
- [11] I. H. Sugeha, R. L. Inkiriwang, and P. A. K. Pratisis, “Optimasi Penjadwalan Menggunakan Metode Algoritma Genetika Pada Proyek Rehabilitasi Puskesmas Minanga,” *J. Sipil Statik*, vol. 7, no. 12, pp. 1669–1680, 2019, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/jss/article/view/26145>.
- [12] C. B. Gokulnath and S. P. Shantharajah, “An optimized feature selection based on genetic approach and support vector machine for heart disease,” *Cluster Comput.*, vol. 22, pp. 14777–14787, 2019, doi: 11007/s10586-018-2416-4.
- [13] R. T. Prasetyo, A. A. Rismayadi, and I. F. Anshori, “Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang,” *J. Inform.*, vol. 5, no. 2, pp. 186–194, 2018, doi: 131311/ji.v5i2.4123.
- [14] B. K. Khotimah, E. Setiawan, V. Sasmeka, A. Fridayanti, I. Maulana, and A. M. Zulfida, “Identifikasi Hama Dan Penyakit Tanaman Jagung Dengan Menggunakan Metode Klasifikasi Support Vector Machine (Svm),” *J. Ilm. NERO*, vol. 7, no. 1, pp. 69–74, 2022.
- [15] F. Novianti *et al.*, “Prediksi Penggunaan Bahan Bakar pada PLTGU menggunakan Metode Support Vector Regression (SVR),” vol. 21, no. 2, pp. 249–255, 2022.
- [16] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 577, 2018, doi: 125126/jtiik.201855983.
- [17] Y. N. Fuadah, I. D. Ubaidullah, N. Ibrahim, F. F. Taliningsing, N. K. Sy, and M. A. Pramuditho, “Optimasi Convolutional Neural Network dan K-Fold Cross Validation pada Sistem Klasifikasi Glaukoma,” *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. dan Tek. Elektron.*, vol. 10, no. 3, p. 728, 2022, doi: 126760/elkomika.v10i3.728.