

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

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
FAHRIZA NOVIANTI
H02219008

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2023

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : FAHRIZA NOVIANTI

NIM : H02219008

Program Studi : Matematika

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "PENERAPAN ALGORITMA GENETIKA UNTUK SELEKSI FITUR DALAM KLASIFIKASI DATA REKAM MEDIS PCOS MENGGUNAKAN SVM". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 13/06/2023



FAHRIZA NOVIANTI

NIM. H02219008


LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh


Nama : FAHRIZA NOVIANTI
NIM : H02219008
Judul Skripsi : PENERAPAN ALGORITMA GENETIKA UNTUK
SELEKSI FITUR DALAM KLASIFIKASI DATA
REKAM MEDIS PCOS MENGGUNAKAN SVM

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I


Nurissaidah Ulinnuha, M.Kom
NIP. 199011022014032004

Pembimbing II


Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya


Yurniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : FAHRIZA NOVIANTI
NIM : H02219008
Judul Skripsi : PENERAPAN ALGORITMA GENETIKA UNTUK
SELEKSI FITUR DALAM KLASIFIKASI DATA
REKAM MEDIS PCOS MENGGUNAKAN SVM

Telah dipertahankan di depan Tim Penguji
pada tanggal 13/06/2023

Mengesahkan,
Tim Penguji

Penguji I



Dian Candra Rini Novitasari, M.Kom
NIP. 198511242014032001

Penguji II



Aris Fanani, M.Kom
NIP. 198701272014031002

Penguji III -



Nurissaldah Ulinuha, M.Kom
NIP. 199011022014032004


Penguji IV



Lutfi Hakim, M.Ag
NIP. 197312252006041001



Mengetahui,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya


Saepul Hamdani, M.Pd
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpustakaan@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : FAHRIZA NOVIANTI
NIM : H02219008
Fakultas/Jurusan : FST / MATEMATIKA
E-mail address : fahriza19.fn@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

PENERAPAN ALGORITMA GENETIKA UNTUK SELEKSI FITUR

DALAM KLASIFIKASI DATA REKAM MEDIS PCOS MENGGUNAKAN

SVM

berserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 13 JUNI 2023

Penulis

(FAHRIZA NOVIANTI)

ABSTRAK

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

PCOS (*Polycystic Ovary Syndrome*) adalah gangguan hormonal pada wanita yang dapat menyebabkan infertilitas. PCOS kebanyakan menyerang wanita pada usia reproduksi, tetapi sering tidak terdeteksi karena kurangnya kesadaran akan kondisi tersebut. Oleh karena itu, deteksi dini perlu dilakukan untuk tindakan pengobatan atau pencegahan PCOS. Diagnosa dapat dilakukan dengan melakukan klasifikasi pada data rekam medis pasien. Dalam penelitian ini, digunakan 40 variabel yang mencakup data hormonal, USG, dan data kesehatan lainnya. Metode yang digunakan untuk klasifikasi adalah *Support Vector Machine* (SVM). SVM dapat mengatasi masalah data nonlinier dengan menggunakan konsep kernel. Untuk meningkatkan akurasi, seleksi fitur dilakukan menggunakan algoritma genetika. Seleksi fitur algoritma genetika menghasilkan 19 fitur yang signifikan. Dengan menggunakan fitur yang terpilih, hasil klasifikasi menghasilkan akurasi= 0.9426, sensitivitas= 0.8757, dan spesifisitas= 0.9752. Hasil klasifikasi SVM tanpa seleksi fitur dengan menggunakan kernel yang sama yaitu akurasi= 0.8246, sensitivitas= 0.6091, dan spesifisitas= 0.9725. Penelitian ini menyimpulkan bahwa implementasi kedua metode tersebut menghasilkan kinerja yang baik.

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

ABSTRACT

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

PCOS (Polycystic Ovary Syndrome) is a hormonal disorder in women that can lead to infertility. PCOS mainly affects women of reproductive age, but it often goes undetected due to a lack of awareness about the condition. Therefore, early detection is necessary for treatment or prevention of PCOS. Diagnosis can be done by classifying patient's medical record data. In this study, 40 variables including hormonal data, ultrasound data, and other health data were used. The method used for classification is Support Vector Machine (SVM). SVM can handle non-linear data problems by utilizing kernel concepts. To improve accuracy, feature selection was performed using a genetic algorithm. The genetic algorithm feature selection resulted in 19 significant features. By using the selected features, the classification results achieved an accuracy= 0.9426, sensitivity= 0.8757, and specificity= 0.9752. The SVM classification results without feature selection using the same kernel were accuracy= 0.8246, sensitivity= 0.6091, and specificity= 0.9725. This study concludes that the implementation of both methods yields good performance.

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

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR LAMBANG	ix
ABSTRAK	x
ABSTRACT	xi
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	6
1.3. Tujuan Penelitian	7
1.4. Manfaat Penelitian	7
1.5. Batasan Masalah	8
1.6. Sistematika Penulisan	8
II TINJAUAN PUSTAKA	10
2.1. <i>Polycystic Ovary Syndrome (PCOS)</i>	10
2.1.1. Diagnosis	10
2.1.2. Ciri-Ciri Penderita PCOS	13
2.1.3. Cara Mengatasi dan Perawatan PCOS	13
2.2. Klasifikasi	14
2.3. Seleksi Fitur	14
2.4. <i>Support Vector Machine (SVM)</i>	15
2.4.1. Kernel	25
2.5. Algoritma Genetika	26

2.6. <i>K-fold Cross Validation</i>	32
2.7. Confusion Matrix	33
2.8. Integrasi Keislaman	35
III METODE PENELITIAN	41
3.1. Jenis Penelitian	41
3.2. Sumber Data	41
3.3. Variabel Penelitian	41
3.4. Diagram Alir Penelitian	43
3.5. Tahapan Penelitian	44
3.6. Uji Coba Parameter	46
IV HASIL DAN PEMBAHASAN	47
4.1. Deskripsi Data	47
4.2. Seleksi Fitur dengan Algoritma Genetika	49
4.2.1. Inisialisasi Fitur	50
4.2.2. Pembangkitan Populasi Awal	50
4.2.3. Perhitungan Nilai <i>Fitness</i>	51
4.2.4. Seleksi Individu	55
4.2.5. <i>Crossover</i>	57
4.2.6. Mutasi	59
4.2.7. <i>Elitism</i>	60
4.2.8. Hasil Seleksi Fitur Algoritma Genetika	63
4.3. Klasifikasi SVM dengan Seleksi Fitur Algoritma Genetika	65
4.4. Klasifikasi SVM Tanpa Seleksi Fitur	66
4.5. Analisis Hasil	68
4.6. Integrasi Keilmuan	70
V PENUTUP	73
5.1. Kesimpulan	73
5.2. Saran	74
DAFTAR PUSTAKA	74

DAFTAR TABEL

2.1 Kernel SVM	26
2.2 <i>Confusion Matrix</i>	34
4.1 Sampel Data Penelitian	48
4.2 Pembangkitan Populasi	51
4.3 <i>Confusion Matrix</i> Individu ke-1	54
4.4 Nilai <i>Fitness</i>	55
4.5 Hasil Perhitungan Probabilitas dan Kumulatif	56
4.6 Seleksi Individu	57
4.7 Individu Terbaik pada Generasi ke-10	63
4.8 Hasil Seleksi Fitur Algoritma Genetika	64
4.9 Hasil Klasifikasi SVM dengan Fitur Terseleksi	65
4.10 <i>Confusion Matrix GA-SVM</i>	66
4.11 Hasil Klasifikasi SVM	67
4.12 <i>Confusion Matrix SVM</i>	67

DAFTAR GAMBAR

2.1 Hasil USG penderita PCOS	12
2.2 SVM pada <i>Linear Separable Data</i>	16
2.3 Transformasi Data menggunakan Fungsi Kernel	25
2.4 Konsep Algoritma Genetika	27
2.5 Mesin Roulette	29
2.6 Ilustrasi <i>Crossover</i> 1 Titik	31
2.7 Ilustrasi Mutasi	32
2.8 Proses Elitism	32
2.9 Ilustrasi Proses <i>k-fold Cross Validation</i>	33
3.1 Variabel Prediktor Penelitian	42
3.2 Diagram Alir Seleksi Fitur Algoritma Genetika dan Klasifikasi SVM	43
4.1 Pola Data dari Beberapa Variabel	49
4.2 <i>Crossover</i> Iterasi-1	58
4.3 <i>Crossover</i> Iterasi-2	59
4.4 <i>Crossover</i> Iterasi-3	59
4.5 <i>Crossover</i> Iterasi-4	59
4.6 Proses Mutasi	60
4.7 Nilai <i>Fitness</i> Seluruh Individu	61
4.8 Hasil Elitism	62
4.9 Perbandingan Hasil	69

DAFTAR LAMBANG

b	: bias
w	: bobot SVM berupa vektor kolom
d^+	: jarak antara <i>hyperplane</i> utama dengan <i>hyperplane</i> ke-2 yang melewati kelas positif
d^-	: jarak antara <i>hyperplane</i> utama dengan <i>hyperplane</i> ke-3 yang melewati kelas negatif
\forall_u	: untuk semua u
x'	: transpose dari x
$\frac{\partial L_p}{\partial b}$: turunan fungsi L_p terhadap b
α	: pengali <i>lagrange</i>
$\sum_{u=1}^N \alpha_u y_u x_u$: penjumlahan $\alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \dots + \alpha_N y_N x_N$
$\ w\ $: panjang euclidean dari vektor w
ε	: <i>slack</i> variabel (epsilon)
\exp	: eksponensial
γ	: gamma
C	: cost
r	: koefisien pada kernel polinomial

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Polycystic Ovary Syndrome (PCOS) ialah suatu penyebab utama infertilitas pada wanita. Penderita PCOS memiliki gangguan pada sistem endokrin saat masa reproduksi yang mengakibatkan keseimbangan hormonal menjadi terganggu (Cunha and Póvoa, 2021). PCOS juga dapat ditandai dengan munculnya kista ovarium dan terjadinya anovulasi. Gangguan pada hormon endokrin dapat memengaruhi hormon yang lainnya seperti pada hormon reproduksi yaitu LH, FSH, estrogen dan testosteron. Hal tersebut akan menyebabkan tidak teraturnya siklus menstruasi sehingga dapat mengganggu sistem reproduksi (Bulsara et al., 2021).

Menurut perkiraan *World Health Organization* (WHO) lebih dari 116 juta wanita atau sekitar 3,4% wanita di seluruh dunia terkena PCOS (Jabeen et al., 2022). Jumlah penderita ini terus meningkat pada tiap tahunnya sehingga membuat para wanita perlu waspada terhadap penyakit ini. Gejala yang dialami oleh penderita PCOS merupakan gejala umum yang sering dialami oleh para wanita sehingga sulit untuk mengetahui apakah gejala tersebut menunjukkan wanita tersebut menderita PCOS atau tidak dan gejala ini dapat berubah dari waktu ke waktu (Ji et al., 2018). Beberapa gejala yang dialami oleh penderita PCOS yaitu kelebihan rambut di tubuh (*hirsutisme*), jerawat dan masalah kulit lainnya, rambut rontok, siklus menstruasi tidak teratur bahkan tidak mensturasi, volume menstruasi

baginya. Apabila sebagai manusia kita dapat bersabar dan mengharapkan suatu pahala, maka akan diberikan suatu pahala padanya.

Wanita pada usia reproduksi rentan terkena PCOS. Namun sejatinya PCOS ialah kelainan genetik yang dapat menyerang wanita pada berbagai usia. Terdapat beberapa penelitian yang menunjukkan penyebaran penyakit PCOS di berbagai wilayah. Penyebaran di Amerika Serikat, Inggris Raya, Spanyol, Yunani, Australia, dan Meksiko menunjukkan tingkat penyebaran penyakit berkisar antara 6% hingga 9% (Wolf et al., 2018). Kemudian tingkat penyebaran PCOS di Eropa berkisar 258.5 hingga 440 per 100.000 penduduk (Miazgowski et al., 2021). Sedangkan di Indonesia sendiri penyebaran PCOS diperkirakan antara 4-6% wanita usia subur (Hestiantoro and Pamungkas, 2020).

Faktor risiko pemicu PCOS antara lain yaitu usia, resistensi insulin, gaya hidup, faktor genetik, body mass index (BMI) serta faktor hormonal lain (Armanini et al., 2022). Namun, PCOS seringkali tidak terdiagnosis dan salah penanganan pada kebanyakan wanita dikarenakan kurangnya kesadaran mengenai kondisi tersebut dan seringkali tidak terdeteksi selama bertahun-tahun (Subha et al., 2022).

Diagnosis PCOS berdasarkan kriteria Rotterdam terdapat 3 kriteria yaitu: hiperandrogen yang diamati secara klinis ataupun pemeriksaan biokimia, anovulasi, serta gambar dari *ovarium polycystic* melalui hasil USG. Dua dari tiga kriteria tersebut sudah cukup untuk mendiagnosis PCOS (Koneru and Priyanka, 2019).

Dikarenakan PCOS diakibatkan oleh kelainan sistem endokrin, maka penyakit ini juga dapat memicu timbulnya penyakit metabolik seperti hipertensi, penyakit jantung, diabetes melitus dan yang lainnya (Chen and Pang, 2021). Oleh karena itu, perlu dilakukan diagnosis dini terhadap PCOS untuk menghindari

risiko penyakit lain yang ditimbulkan serta diagnosis ini juga diperlukan dalam mengatasi masalah infertilitas. Diagnosis dapat dilakukan dengan melakukan klasifikasi pada data rekam medis pasien yang mempunyai gejala, tanda klinis, dan faktor risiko yang berkaitan dengan PCOS (Koneru and Priyanka, 2019).

Metode yang dapat dipergunakan dalam melakukan klasifikasi penyakit salah satunya ialah *Support Vector Machine* (SVM). SVM ialah metode yang biasa dipergunakan untuk menyelesaikan permasalahan regresi dan klasifikasi. Cara kerja SVM ialah menemukan *hyperplane* terbaik dengan memaksimalkan margin (Agustina et al., 2018). SVM bekerja dengan baik pada kumpulan data berdimensi tinggi dengan menggunakan teknik kernel (Saikin et al., 2021). Kernel yang biasa dipergunakan dalam menyelesaikan permasalahan pada SVM ada 3, yaitu kernel linier, kernel polinomial, dan kernel *Radial Basis Function* (RBF) (Widayani and Harliana, 2021).

Penelitian yang dilakukan (Anggoro and Kurnia, 2020) dengan membandingkan metode SVM dengan *K-Nearest Neighbors* (KNN) dalam melakukan diagnosis pada penyakit jantung diperoleh hasil akurasi pada metode SVM lebih baik yaitu sebesar 90.1% dibandingkan hasil akurasi metode KNN yaitu sebesar 81.31%. Pada penelitian lain yang dilakukan (Pahlevi et al., 2021) dalam melakukan diagnosis penyakit Diabetes Melitus tipe 2 dengan membandingkan algoritma SVM dan *Iterative Dichotomiser 3* (ID3), hasil yang didapatkan juga menunjukkan bahwa akurasi SVM lebih baik. Akurasi yang didapatkan dengan menggunakan metode SVM ialah sebesar 78.5% sedangkan akurasi yang didapatkan dengan metode ID3 ialah sebesar 74%.

Beberapa penelitian sebelumnya telah menunjukkan bahwa SVM merupakan metode yang baik dalam melakukan klasifikasi. Kelebihan metode SVM ialah mampu mengatasi permasalahan data dengan dimensi yang tinggi dan metode ini

tergolong mudah dalam penerapannya kerana dalam pencarian *support vector* dapat dirumuskan dengan solusi *quadratic programming* (Puspitasari et al., 2018). Di sisi lain, SVM memiliki kekurangan yaitu sulit dalam memilih fitur yang signifikan dari data yang digunakan sehingga berakibat nilai akurasi cenderung lebih kecil (Resmiati and Arifin, 2021). Dalam mengatasi permasalahan tersebut, perlu diterapkan seleksi fitur untuk meningkatkan tingkat akurasi dalam melakukan klasifikasi menggunakan metode SVM.

Seleksi fitur yang dipergunakan pada penelitian ini ialah algoritma genetika. Algoritma ini bekerja dengan menerapkan seleksi alam hingga tersisa individu paling baik yang mampu bertahan untuk dipergunakan lagi dalam seleksi lanjutan (Akhmad and Marsuni, 2019). Dipilihnya algoritma genetika untuk seleksi fitur dikarenakan algoritma ini dinilai amat baik dalam mencari solusi masalah yang kompleks dengan banyaknya variabel (Santoso et al., 2017). Dan juga, algoritma ini mampu mengoptimasi berbagai permasalahan bergantung pada nilai *fitness* dan mampu dalam memelajari ruang penelitian yang cukup besar (Prasetio et al., 2018).

Pada penelitian sebelumnya terkait membandingkan algoritma genetika dengan algoritma *greedy* dalam memprediksi rentet waktu harga *crude palm oil* didapatkan hasil terbaik menggunakan algoritma genetika dengan nilai RMSE sebesar 0.096 dibandingkan dengan algoritma *greedy* dengan nilai RMSE 0.111 (Puspitasari, 2017). Pada permasalahan lain, mengenai identifikasi ketidaknormalan pada payudara dengan membandingkan seleksi fitur algoritma genetika dan *particle swarm optimization* (PSO) didapatkan hasil seleksi fitur menggunakan algoritma genetika lebih baik dengan akurasi sebesar 87.08% dibandingkan dengan PSO yang akurasi sebesar 86.16% (Rodrigues et al., 2019).

Kebaikan seleksi fitur algoritma genetika pada SVM ditunjukkan oleh penelitian yang dilakukan (Gunawan et al., 2020) dengan membandingkan kombinasi seleksi fitur algoritma genetika dengan naive bayes dalam menganalisis sentimen Calon Gubernur Jawa Barat Tahun 2018-2023. Hasil akurasi SVM dengan algoritma genetika lebih baik yaitu dengan 93.03% sedangkan Naive Bayes dengan algoritma genetika sebesar 92%.

Dikarenakan implementasi kedua metode tersebut dapat menghasilkan akurasi yang cukup baik dalam melakukan klasifikasi, maka penulis ingin menggunakan seleksi fitur algoritma genetika terhadap metode SVM untuk mendeteksi penyakit PCOS berdasarkan data rekam medis. Pada penelitian sebelumnya, belum pernah dilakukan seleksi fitur dengan menggunakan algoritma genetika pada metode SVM dalam mendiagnosis PCOS. Harapan dari penelitian ini adalah diperoleh nilai akurasi yang tinggi dalam melakukan diagnosis PCOS berdasarkan data rekam medis pasien menggunakan metode SVM dan seleksi fitur algoritma genetika.

1.2. Rumusan Masalah

Dari yang sudah dijabarkan pada latar belakang, maka rumusan masalah yang akan dibahas yaitu:

1. Bagaimana hasil seleksi fitur algoritma genetika pada data rekam medis pasien PCOS ?
2. Bagaimana hasil evaluasi sistem dari klasifikasi Support Vector Machine dengan Algoritma Genetika sebagai seleksi fitur pada data rekam medis PCOS ?
3. Bagaimana perbandingan hasil klasifikasi antara GA-SVM dan SVM pada

data rekam medis PCOS ?

1.3. Tujuan Penelitian

Didasarkan dari rumusan masalah di atas, maka tujuan dilakukannya penelitian ini ialah:

1. Mengetahui hasil seleksi fitur Algoritma Genetika pada data rekam medis pasien PCOS.
2. Mengetahui hasil evaluasi sistem dari klasifikasi Support Vector Machine dengan Algoritma Genetika sebagai seleksi fitur pada data rekam medis PCOS.
3. Mengetahui perbandingan hasil klasifikasi antara GA-SVM dan SVM pada data rekam medis PCOS.

1.4. Manfaat Penelitian

Dari pemaparan tujuan penelitian di atas, harapan dari penelitian ini ialah dapat memberi manfaat yaitu:

1. Manfaat Teoritis

Dalam bidang keilmuan matematika dan kesehatan, kajian ini diharapkan mampu memperkaya wawasan mengenai analisis risiko penyakit PCOS menggunakan metode GA-SVM.

2. Manfaat Praktis

- (a) Bagi Universitas, penelitian ini dapat memperbanyak koleksi bahan bacaan dan referensi belajar yang bermanfaat bagi mahasiswa dan mahasiswi Prodi Matematika UINSA Surabaya.

- (b) Bagi Penulis, penelitian ini dapat memperluas pengetahuan dan mengasah kemampuan menulis serta dapat mengenali penyakit PCOS dan metode yang dipergunakan dalam olah data rekam medis.
- (c) Bagi Tenaga Medis, penelitian ini dapat mempermudah tenaga medis dalam menganalisa hasil rekam medis pasien.

1.5. Batasan Masalah

Ruang lingkup yang luas dapat mengakibatkan sasaran pembahasan berkurang. Batasan terhadap permasalahan dapat dilakukan agar pembahasan yang dilakukan dapat sesuai dengan sasaran yang dituju. Batasan permasalahan pada penelitian ini diantaranya:

1. Data yang dipergunakan dalam penelitian ini adalah *Polycystic ovary syndrome* (PCOS) dataset yang diambil dari *Kaggle* database.
2. Parameter yang digunakan pada sistem klasifikasi penelitian ini yaitu kernel *Radial Basis Function* (RBF), kernel Linear, dan kernel Polynomial.
3. Penelitian ini menggunakan 5 dan 10-*fold cross validation*

1.6. Sistematika Penulisan

Penyusunan sistematika penulisan ialah dengan membagi penulisan menjadi beberapa bab. Hal ini berguna untuk memudahkan dalam proses penulisan. Berikut sistematika penulisan pada penelitian ini:

1. BAB 1: PENDAHULUAN

Bab ini berisikan latar belakang yang menjelaskan alasan penulisan penelitian, kemudian rumusan masalah yang dijawab pada hasil penelitian

dan pembahasan, lalu tujuan, manfaat dari penelitian, batasan masalah yang berguna untuk mempertegas masalah yang dipergunakan dalam penulisan, serta sistematika penulisan.

2. Bab II: TINJAUAN PUSTAKA

Bab ini berisikan penjelasan terkait teori yang dipergunakan dalam penelitian yang dilakukan.

3. Bab III: METODOLOGI PENELITIAN

Bab ini berisikan penjelasan terkait tahapan penelitian yang berupa jenis penelitian, sumber data, variabel, tahapan dan uji coba parameter yang dipergunakan dalam penelitian.

4. Bab IV: HASIL DAN PEMBAHASAN

Bab ini berisikan deskripsi data, implementasi algoritma genetika dalam seleksi fitur dan klasifikasi dengan Support Vector Machine, hasil uji coba parameter, dan analisis hasil klasifikasi.

5. Bab V: PENUTUP

Bab ini berisikan simpulan yang dirangkum dari hasil yang telah didapatkan sebelumnya dan juga berisikan saran yang ditujukan untuk peneliti selanjutnya apabila ingin mengembangkan penelitian ini.

BAB II

TINJAUAN PUSTAKA

2.1. *Polycystic Ovary Syndrome (PCOS)*

PCOS menyerang sekitar 8-20% wanita pada usia reproduktif di dunia dan terdapat sekitar 70% wanita yang terjangkit PCOS tidak terdiagnosis (Teede et al., 2018). Penyebab PCOS hingga saat ini belum diketahui secara pasti, namun diduga buruknya kebiasaan hidup serta faktor genetik dapat memicu wanita terserang PCOS (Fritz and Speroff, 2011). Akibat dari PCOS apabila tidak segera dilakukan pengobatan ialah gangguan kesuburan serta penyakit degeneratif lainnya (Chang and Dumesic, 2019).

Wanita seringkali baru menyadari dia terserang PCOS saat dia telah berumah tangga dan ingin segera mendapatkan keturunan. Ciri-ciri PCOS seringkali disepelekan, misalnya siklus menstruasi tidak teratur (Armanini et al., 2022). Hal tersebut banyak wanita yang kurang menyadari jika dia terserang PCOS. Sebenarnya pada usia remaja penyakit PCOS sudah bisa terdeteksi. Namun, keraguan mengenai permasalahan hormonal atau sekedar menyepelekan gejala yang ada pada tubuh membuat gejala ini akan semakin parah apabila tidak segera dilakukan pengobatan dan diagnosis dini (Chen and Pang, 2021).

2.1.1. **Diagnosis**

Dalam melakukan diagnosa terkait dengan PCOS, terdapat suatu kriteria penilaian yang disebut sebagai kriteria Rotterdam. Kriteria ini ialah kriteria yang dipergunakan dalam penanganan infertilitas di Indonesia (Mustari et al., 2018).

Berdasarkan kriteria Rotterdam, ada 3 kriteria yang dipergunakan tetapi meskipun hanya dengan menggunakan 2 kriteria saja sudah dapat ditegaskan bahwa wanita tersebut mengidap PCOS (Koneru and Priyanka, 2019). Berikut merupakan 3 kriteria *Rotterdam*:

1. Anovulasi

Ovulasi atau pembuahan yang normal ditandai dengan normalnya siklus menstruasi. Sekitar 60-80% wanita yang mengidap PCOS memiliki gangguan pada siklus menstruasi dan jenis yang kebanyakan diderita ialah amenore dan oligomenorea. Selain itu, kadar LH serum penderita PCOS sering mengalami peningkatan (Fritz and Speroff, 2011).

2. Hiperandrogenisme

Pada kriteria *Rotterdam*, *hyperandrogenisme* meliputi tanda klinis dan biokimiawi (Fritz and Speroff, 2011). Tanda-tanda tersebut dijelaskan sebagai berikut:

(a) Biokimiawi

Berupa peningkatan kadar androgen pada tubuh. Androgen terpenting yang kerap kali dipergunakan dalam diagnosa ialah testosteron. Peningkatan androgen lain meliputi DHEA, DHEA-S, dan androstenedion (Vidya Bharathi et al., 2017).

(b) Klinis

Hirsutisme ialah tanda berlebihnya androgen yang paling jelas serta merupakan gejala penting pada PCOS. *Hirsutisme* ialah bertumbuhnya rambut tebal pada wanita di bagian tubuh yang pada dasarnya tumbuh rambut pada pria. Selain itu, tanda klinis lainnya

adalah munculnya jerawat serta Kerontokan permanen rambut yang menyebabkan kebotakan (Vidya Bharathi et al., 2017).

3. Gambaran *Polycystic Ovarium* pada Pemeriksaan USG

Ciri-ciri yang kerap dialami oleh penderita PCOS berdasarkan pada hasil USG adalah memiliki 12 folikel atau lebih yang berdiameter 2-9 mm pada tiap-tiap ovarium dan/ atau meningkatnya volume ovarium ($> 10\text{ml}$) (El Hayek et al., 2016). Gambaran Ovarium Polikistik penderita PCOS diperlihatkan pada Gambar 2.1.



Gambar 2.1 Hasil USG penderita PCOS

Hasil USG pada Gambar 2.1 dapat terlihat bahwa panjang ovarium (4,6 cm) dan lebar (2,2 cm) serta luas ovarium (8,9 cm²). Jumlah folikel dengan diameter antara 2 dan 5 mm lebih dari 10.

Wanita dengan gambaran ovarium polikistik tanpa hiperandrogenisme atau gangguan ovulasi tidak dapat belum bisa didiagnosis PCOS hingga didapatkannya informasi atau tanda klinis lainnya (Chang and Dumesic, 2019). Oleh karena itu, saat melakukan diagnosis perlu diperhatikan aspek lain selain hasil USG yaitu dengan pemeriksaan laboratorium terkait hormonal dan yang lainnya.

2.1.2. Ciri-Ciri Penderita PCOS

Ciri Fisik PCOS Seorang wanita yang mengalami PCOS pada umumnya terlihat dari ciri fisiknya yang memiliki kadar hormon androgen berlebih. Ciri-ciri fisik penderita PCOS yaitu (Norman and Teede, 2018):

1. Hirsutisme atau tumbuhnya rambut yang tidak diinginkan (kumis, jambang, rambut tangan, rambut dada dan rambut kaki)
2. Jerawat yang sulit hilang
3. Rambut rontok, apabila telah parah dapat menimbulkan kebutakan
4. Ketombe Rambut berminyak
5. Kulit menggelap
6. Sulit untuk menurunkan berat badan dan obesitas

2.1.3. Cara Mengatasi dan Perawatan PCOS

Pengobatan serta perawatan penderita PCOS berbeda tergantung pada gejala yang ada (Ji et al., 2018). Penanganan antara wanita telah menikah dan belum menikah juga berbeda. Pada wanita yang telah menikah dan ingin secepatnya mendapatkan anak, langkah pertama adalah memperbaiki pola hidup terlebih dahulu. Lalu dokter akan memberikan penilaian terhadap indung telur, kondisi rahim dan saluran telur pada wanita dan juga kondisi sel sperma pada pria. Apabila penilaian menunjukkan semuanya baik, maka program kehamilan dapat dilaksanakan secara alamiah. Namun apabila hasilnya kurang baik, dokter akan memberikan saran untuk bantuan program kehamilan (Suparman and Suparman, 2017).

Wanita yang belum menikah, fokus pengobatan akan pada perbaikan siklus menstruasi agar supaya normal kembali. Dokter biasanya memberi saran untuk melakukan perbaikan pada pola hidup agar berat badan turun sehingga gejala PCOS mereda. Apabila diperlukan, dokter akan memberi obat pemancing menstruasi yang akan diberikan selama 3 siklus berurutan. Apabila siklus menstruasi belum kembali normal, maka akan dilakukan pengulangan pengobatan sampai siklus menjadi normal (Suparman and Suparman, 2017).

2.2. Klasifikasi

Klasifikasi adalah suatu proses pengelompokkan suatu data berdasarkan karakteristik yang dimiliki oleh data tersebut. Dalam pengerjaannya, klasifikasi bisa dikerjakan dengan beraneka cara baik manual maupun dengan teknologi. Cara manual yaitu dikerjakan oleh manusia tanpa bantuan dari sistem komputer. Sedangkan apabila menggunakan teknologi, banyak metode yang dapat diterapkan diantaranya yaitu Support Vector Machine, Decision Tree, Fuzzy, Naïve Bayes dan Jaringan Saraf Tiruan (Wibawa et al., 2018)

Proses klasifikasi akan dimulai dengan dilakukannya pembagian data menjadi 2, yaitu data uji dan data latih. Pada tahap awal, data uji digunakan untuk membuat model. Selanjutnya yaitu model yang didapatkan akan dilakukan pengujian dengan menggunakan data latih untuk mendapatkan nilai keakuratan dari model. Pembagian data tersebut dapat berpengaruh dalam hasil akurasi dari suatu metode klasifikasi (Bianto et al., 2020).

2.3. Seleksi Fitur

Seleksi fitur memiliki peranan penting dalam melakukan pengolahan data. Teknik ini dipergunakan untuk mereduksi dimensi data serta fitur yang tidak

signifikan sehingga efisiensi dan efektifitas kinerja algoritma klasifikasi dapat meningkat (Chen et al., 2020).

Dalam pembelajaran *machine learning* terdapat 2 jenis metode dalam melakukan seleksi fitur yaitu *filter* dan *wrappers* (Anggraeni et al., 2021).

1. *Filters*

Filters berkerja dengan menggunakan fitur yang dipilih. Dalam melakukan evaluasi pada fitur, *filters* mempergunakan matriks evaluasi untuk mengukur kemampuan fitur dalam membedakan tiap kelas. Metode *filters* terdiri dari *correlation based feature selection* (CFS), *chi-square*, dan *information gain* (Moraes et al., 2013).

2. *Wrappers*

Dalam melakukan evaluasi, *Wrappers* menggunakan akurasi klasifikasi beberapa algoritma. *Wrappers* harus melakukan pengujian metode klasifikasi untuk tiap bagian fitur yang akan dievaluasi. Dikarenakan evaluasi fitur dilakukan secara berulang, maka tingkat akurasi klasifikasi yang dihasilkan tinggi. Salah satu metode *wrappers* dalam seleksi fitur adalah algoritma genetika (Abdilah et al., 2018).

2.4. *Support Vector Machine* (SVM)

SVM ialah salah satu penerapan *machine learning* yang dipergunakan dalam masalah regresi dan klasifikasi. Klasifikasi dengan SVM adalah dilakukan dengan ditemukannya fungsi pemisah yang memisahkan data menjadi dua kelas yang berbeda (Saikin et al., 2021). Prinsip kerja metode ini ialah menemukan *hyperplane* terbaik diantara banyak fungsi yang terbentuk dalam memisahkan dua buah kelas. Cara menemukan *hyperplane* terbaik ialah dengan dimaksimalkannya

turunan parsial, masalah dualnya menjadi:

$$\begin{aligned} \max L_p &= \frac{1}{2} \left(\sum_{u=1}^N \alpha_u y_u x_u \right) \left(\sum_{u=1}^N \alpha_u y_u x_u \right) + C \sum_{u=1}^N \varepsilon_u \\ &\quad - \sum_{u=1}^N \alpha_u y_u \left(\left(\sum_{u=1}^N \alpha_u y_u x_u \right) \cdot x_u + b \right) \\ &\quad - \sum_{u=1}^N \alpha_u - \sum_{u=1}^N \alpha_u \varepsilon_u - \sum_{u=1}^N q_u \varepsilon_u \\ \max L_p &= \frac{1}{2} \sum_{u=1}^N \sum_{v=1}^N \alpha_u y_u \alpha_v y_v (x_u \cdot x_v) - \sum_{u=1}^N \sum_{v=1}^N \alpha_u y_u \alpha_v y_v (x_u \cdot x_v) \\ &\quad - b \sum_{u=1}^N \alpha_u y_u + \sum_{u=1}^N \alpha_u + C \sum_{u=1}^N \varepsilon_u - \sum_{u=1}^N \alpha_u \varepsilon_u - \sum_{u=1}^N q_u \varepsilon_u \\ \max L_p &= \sum_{u=1}^N -\frac{1}{2} \sum_{u=1}^N \sum_{v=1}^N \alpha_u y_u \alpha_v y_v (x_u \cdot x_v) \end{aligned}$$

dengan syarat $\sum_{u=1}^N \alpha_u y_u = 0$ karena $\frac{\partial L_p}{\partial \varepsilon_u} = C - \alpha_u - q_u$.

Nilai q dalam fungsi tujuan diabaikan, sehingga muncul kendala baru dengan yaitu $0 \leq \alpha_u \leq C$. Definisi dari bias b menghasilkan sebuah kendala yang harus dipenuhi:

$$\sum_{u=1}^N \alpha_u y_u = 0$$

Data input $x_u = [x_{u1}, x_{u2}, \dots, x_{im}]$ diubah menjadi $x'_u = [x_{u1}, x_{u2}, \dots, x_{im} \lambda]$ serta vektor bobot w menjadi $w' = [w_1, w_2, \dots, w_m \frac{b}{\lambda}]$ dimana λ ialah suatu konstanta. Bentuk masalah menjadi:

$$\begin{aligned} L'_p = \min L_p &= \frac{1}{2} \|w'\|^2 + C \sum_{u=1}^N \varepsilon_u + \sum_{u=1}^N \alpha_u (1 - y_u (x'_u \cdot w')) - \varepsilon_u \\ &\quad + \sum_{u=1}^N q_u (-\varepsilon_u) \end{aligned}$$

1. Inisialisasi

Teknik inisialisasi menyatakan salah satu solusi dari sebuah permasalahan. Teknik ini dilakukan dengan memberi definisi pada tipe serta jumlah gen yang dipergunakan dan dipastikan dapat mewakili solusi dari masalah yang diangkat. Inisialisasi dilakukan dengan mengodekan suatu gen dalam kromosom. Satu gen menggambarkan satu variabel. Gen dapat berupa berupa simbol, karakter, bilangan numerik, ataupun biner. Hal tersebut bergantung pada masalah yang ingin diselesaikan (Krisnandi and Agung, 2017).

2. Pembangkitan Populasi Awal

Pada algoritma genetika, dibangkitkannya populasi awal dilakukan secara acak. Populasi terdiri atas individu-individu yang menggambarkan solusi yang diinginkan (Iriany et al., 2017). Individu terdiri atas kumpulan kromosom, kromosom terdiri atas kumpulan gen, serta nilai sebuah gen ialah alel.

3. Fungsi *Fitness*

Fungsi *fitness* pada algoritma genetika dipergunakan untuk memberikan penilaian pada suatu permasalahan atau biasa disebut *fitness value* yang dijadikan sebagai acuan dalam mendapatkan nilai yang optimal pada algoritma genetika (Sugeha et al., 2019). Pada penelitian ini, fungsi *fitness* yang dipergunakan ialah nilai akurasi pada metode klasifikasi (Binarwati et al., 2017). Pada penelitian ini digunakan nilai akurasi yang didapatkan dengan menggunakan metode SVM.

4. Seleksi

Seleksi ialah teknik dipilihnya *parents* untuk melakukan reproduksi.

terbentuk. Individu tersebut akan menjadi suatu *parent*.

- (d) Mengulangi langkah 2 dan 3 hingga sebanyak jumlah individu pada suatu populasi.

Setelah didapatkan individu terpilih pada proses seleksi ini, maka selanjutnya akan dilakukan proses reproduksi yaitu teknik *crossover* dan mutasi pada individu yang telah terpilih tersebut.

5. Pindah Silang (*Crossover*)

Pada saat *crossover*, dua kromosom induk akan membentuk kromosom baru. Pembentukan kromosom baru ini dilakukan dengan menggabungkan informasi dari tiap-tiap induk kromosom. Hasil *Crossover* adalah kromosom baru yang dikenal sebagai *offspring* (kromosom anak) (Arifudin, 2012). *Crossover* memiliki tujuan yaitu untuk mengembangkan replika individu menjadi lebih signifikan dalam suatu populasi tanpa menghasilkan yang baru (Gokulnath and Shantharajah, 2019).

Offspring yang dihasilkan memiliki jumlah yang bergantung oleh nilai probabilitas *crossover* (C_r) yang telah ditetapkan di awal (Taufiq et al., 2017). Umumnya, nilai probabilitas *crossover* berkisar $0 \leq C_r \leq 1$ dan untuk nilai yang baik berada pada kisaran 0,6 sampai 1 (Iriany et al., 2017).

Terdapat beberapa teknik yang dapat dipergunakan dalam melakukan *crossover*. Diantaranya adalah *crossover* n titik, *crossover* kromosom permutasi, *crossover* satu titik, dan *crossover* aritmatika. Dalam penelitian ini dipergunakan *crossover* 1 titik. Ilustrasi dari *crossover* diperlihatkan pada Gambar 2.6.

masing-masing.' Maka Rabbmu lebih mengetahui siapa yang lebih benar jalannya. (QS. Al-Isra': 84).

Dalam surat Al Isra ayat 84 tersebut dijelaskan bahwa Allah SWT mengetahui segala sesuatu yang telah kita lakukan dan akan memberikan balasan sesuai dengan amal perbuatan yang telah dilakukan. Pelajaran yang dapat diambil dari surat tersebut ialah mengajarkan kita untuk selalu berbuat kebaikan agar kebaikan tersebut dapat kembali kepada diri kita sendiri.

Ketika kita telah berusaha untuk merubah gaya hidup menjadi lebih baik namun belum juga mendapatkan kesembuhan, hendaknya kita tetap bersabar karena apabila kita masih diberikan penyakit oleh Allah SWT, maka penyakit itu merupakan suatu penanda bahwa Dia mengasihi kita dengan memberi kita penyakit agar senantiasa taat kepada-Nya. Sebagai umat manusia hendaknya harus menerima segala ujian yang diberikan agar selalu diridhoi oleh Allah SWT.

Penyakit PCOS yang diderita oleh kebanyakan wanita bukanlah penyakit yang ringan dikarenakan salah satu dampak dari PCOS adalah infertilitas yang berarti wanita tersebut tidak bisa memiliki seorang anak. Apabila seorang wanita yang menderita PCOS menerima dengan sabar dan ridho, maka Allah SWT juga akan menurunkan ridho kepada wanita tersebut dan atas ridho Allah SWT terdapat kemungkinan penderita PCOS bisa disembuhkan sehingga dapat terhindar dari kemandulan.

BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini termasuk dalam jenis penelitian kuantitatif. Data yang dipergunakan berupa data rekam medis pasien dengan berbagai macam kategori data yang akan dikonversikan ke dalam bentuk numerik untuk kemudian dilakukan dalam proses klasifikasi. Proses klasifikasi ini bertujuan untuk melakukan deteksi dini PCOS pada wanita sebagai suatu tindakan agar dapat mencegah terjadinya risiko infertilitas.

3.2. Sumber Data

Data yang dipergunakan dalam penelitian ini diambil secara daring pada halaman website Kaggle Dataset. Data yang dipergunakan terdiri dari data yang dikumpulkan dari sepuluh rumah sakit berbeda di Kerala, India yang tersedia di database *Kaggle*. Data terdiri dari informasi fisik dan klinis terkait dengan PCOS pada 540 pasien dengan dan tanpa masalah kesuburan. Dari 540 pasien tersebut didapatkan 177 pasien yang positif terkena PCOS sedangkan 363 lainnya negatif PCOS.

3.3. Variabel Penelitian

Variabel penelitian adalah sebuah parameter yang memiliki fungsi untuk objek penelitian sehingga didapatkan suatu informasi yang pada akhirnya dapat diambil suatu kesimpulan. Terdapat 2 jenis variabel pada penelitian ini yaitu:

3.5. Tahapan Penelitian

Penelitian ini dilakukan dengan menerapkan dua metode yaitu Algoritma Genetika sebagai seleksi fitur dan SVM sebagai metode klasifikasi. Alur penelitian dijelaskan sebagai berikut:

1. Seleksi Fitur Algoritma Genetika.

- (a) Masukkan data rekam medis pasien PCOS.
- (b) Melakukan pembagian data.

Uji coba pembagian data dilakukan sebanyak 3 kali yaitu dengan pembagian 90:10, 80:20, dan 70:30. Pembagian data 90:10 mengartikan 90% data dipergunakan untuk data latih sedangkan 10% sisanya dipergunakan sebagai data uji. Begitu pula dengan pembagian data 80:20 serta 70:30.

Uji coba pembagian data dilakukan untuk menguji dan membandingkan metode yang dilakukan secara adil, dengan menggunakan subset data yang sama untuk melihat mana yang memberikan hasil terbaik. Uji coba pembagian data membantu mengetahui seberapa baik metode yang dilakukan dalam suatu penelitian.

- (c) Tahap seleksi fitur.

Metode yang dipergunakan untuk seleksi fitur ialah algoritma genetika. Penentuan fitur terbaik dengan menggunakan algoritma ini ialah dengan terlebih dahulu dilakukan inisialisasi parameter. Parameter algoritma genetika berupa *Popsiz*e, probabilitas *crossover* (*Cr*) dan probabilitas mutasi (*Mr*). Pemilihan fitur dilakukan dengan

membangkitkan populasi secara acak. Individu yang terbentuk dalam satu populasi ialah sama dengan nilai *popsize* yang digunakan. Setiap individu yang terbentuk, dihitung nilai *fitness* dengan menggunakan nilai akurasi metode klasifikasi yaitu SVM.

Langkah selanjutnya setelah didapatkan nilai *fitness* pada tiap individu dalam suatu populasi ialah dilakukannya proses seleksi. Seleksi dilakukan dengan menggunakan konsep mesin roulette dengan harapan individu terbaik memiliki peluang lebih besar untuk melakukan proses reproduksi yaitu persilangan (*crossover*), dan mutasi.

Individu yang telah terseleksi kemudian dilakukan proses yaitu *crossover* dan mutasi. Pada proses *crossover* dan mutasi akan menghasilkan individu baru yang disebut dengan *offspring*. Setelah proses *crossover* dan mutasi dilakukan, jumlah individu dalam suatu populasi akan bertambah. Proses *elitism* dilakukan untuk mempertahankan jumlah individu dalam suatu populasi. Proses ini dilakukan dengan memilih individu teratas berdasarkan nilai *fitness* dalam jumlah yang sama dengan jumlah individu yang terbentuk (*popsize*). Hal ini dikerjakan berulang kali hingga mencapai batas yang dipergunakan (maksimal generasi). Pada tahapan ini akan didapatkan fitur-fitur yang terpilih dengan nilai akurasi terbaik.

2. Klasifikasi SVM dengan Fitur Terbaik dari Algoritma Genetika

- (a) Masukkan data dari hasil seleksi fitur.
- (b) Melakukan pembagian data dengan *K-fold cross validation*.

Uji coba pembagian data dilakukan sebanyak 2 kali yaitu dengan pembagian *5-fold* dan *10-fold*.

(c) Inisialisasi Parameter.

Pada SVM dilakukan inisialisasi parameter pada kernel dalam melakukan pembentukan model. Kernel RBF berupa nilai C dan γ . Kernel polinomial berupa nilai C , γ , r dan d . Kernel linear berupa nilai C .

(d) Tahap Evaluasi.

Tahap evaluasi hasil klasifikasi yaitu dengan menggunakan *confusion matrix* untuk mendapatkan nilai akurasi, spesifitas dan sensitivitas menggunakan Persamaan (2.20) - (2.21)

3. Analisis Hasil

Analisis Hasil dilakukan dengan membandingkan hasil nilai evaluasi dari model klasifikasi yang berupa nilai akurasi, sensitivitas, dan spesifitas berdasarkan dengan uji coba kernel linear, polinomial dan RBF yang telah dilakukan pada metode Klasifikasi SVM dengan seleksi fitur algoritma genetika dan tanpa seleksi fitur.

3.6. Uji Coba Parameter

Uji coba parameter dilakukan untuk menentukan parameter terbaik yang akan menghasilkan hasil akurasi terbaik. Uji coba pada tahap algoritma genetika meliputi ukuran populasi ($popsize = 40, 60, 80$), banyaknya generasi ($I = 100$), probabilitas *crossover* (Cr) = 0.8, dan probabilitas mutasi (Mr) = 0.1. Sedangkan pada tahap SVM, dilakukan uji coba terhadap semua kernel yang ada pada SVM.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data yang dipergunakan pada penelitian ini merupakan data sekunder yang didapatkan dari sepuluh rumah sakit berbeda di Kerala, India yang tersedia di database Kaggle. Data terdiri dari 540 pasien dengan 40 atribut penunjang keputusan yang tertera pada Tabel 3.1 dan kelas yang terdiri dari 177 kelas positif dan 363 kelas negatif.

Variabel prediktor pada penelitian ini ada 2 jenis yaitu berupa kategorik dan berupa numerik. Terdapat 10 variabel yang merupakan variabel kategorik, yaitu:

1. Golongan Darah → A+ = 11, A- = 12, B+ = 13, B- = 14, O+ = 15, O- = 16, AB+ = 17, AB- = 18
2. Siklus Menstruasi → Teratur = 2, Tidak Teratur = 4
3. Pernah Hamil → Iya = 0, Tidak = 1
4. Kenaikan Berat Badan → Tidak = 0, Iya = 1
5. Hirsutisme → Tidak = 0, Iya = 1
6. Kulit Menggelap → Tidak = 0, Iya = 1
7. Rambut Rontok → Tidak = 0, Iya = 1
8. Berjerawat → Tidak = 0, Iya = 1

program (Koncz and Paralic, 2011). Seleksi fitur dengan menggunakan algoritma genetika dilakukan melalui beberapa tahapan yaitu inisialisasi, pembangkitan populasi, perhitungan nilai *fitness*, seleksi, *crossover*, mutasi, dan elitism.

Uji coba perhitungan seleksi fitur Algoritma Genetika dilakukan dengan menggunakan uji parameter digunakan yaitu $popsiz = 10$, $Cr = 0.8$ dan $Mr = 0.1$, sedangkan untuk uji coba perhitungan nilai akurasi SVM (*fitness*) digunakan kernel RBF dan pembagian data 90:10.

4.2.1. Inisialisasi Fitur

Dalam proses seleksi fitur, tahapan inisialisasi fitur berarti memberikan definisi pada tipe gen yang akan digunakan digunakan. Gen menggambarkan suatu variabel atau fitur. Pada seleksi fitur ini digunakan tipe gen yaitu bilangan biner yaitu '1' yang melambangkan fitur yang akan dipilih atau digunakan sedangkan untuk fitur yang tidak dipergunakan akan dilambangkan dengan '0'.

4.2.2. Pembangkitan Populasi Awal

Setelah melakukan inisialisasi pada fitur, selanjutnya adalah melakukan pembentukan populasi awal. Populasi awal dibentuk secara acak. Pada perhitungan ini menggunakan $popsiz$ yaitu 10 yang mengartikan bahwa ukuran populasi atau populasi yang akan terbentuk terdiri dari 10 individu. Tabel pembangkitan populasi diperlihatkan pada Tabel 4.2.

Tabel 4.6

Tabel 4.6 Seleksi Individu

<i>Parent</i>	<i>Acak (p)</i>	<i>Individu ke-</i>	<i>Fitness</i>
1.	0.9301	Individu 10	0.7407
2.	0.2057	Individu 3	0.8148
3.	0.3308	Individu 4	0.7778
4.	0.8589	Individu 9	0.8148
5.	0.6407	Individu 7	0.8333
6.	0.5518	Individu 6	0.7963
7.	0.4460	Individu 5	0.8333
8.	0.0542	Individu 1	0.7963
9.	0.6628	Individu 7	0.8333
10.	0.2057	Individu 3	0.8148

Diamati pada Tabel 4.6, seleksi pertama individu 7 terpilih dikarenakan nilai $p = 0.7069$ yang berada pada rentang 0.60564 sampai dengan 0.71127. Pada seleksi kedua, individu 3 terpilih karena nilai $p = 0.2057$ yang berada pada rentang 0.19719 sampai dengan 0.30048. Hal serupa diperlakukan dengan cara yang sama hingga individu terpilih ke-10. Tiap individu yang telah terseleksi, akan menjadi calon *parent* untuk menghasilkan *offspring* ketika proses reproduksi gen.

4.2.5. Crossover

Setelah didapatkannya *parents* dari proses seleksi, yang akan dilakukan selanjutnya ialah melakukan proses reproduksi. Proses reproduksi yang pertama dilakukan yaitu *crossover*. Pada tahap ini, akan dipilih secara acak dua buah individu untuk dilakukan proses *crossover*. *Offspring* yang didapatkan bergantung pada jumlah individu serta probabilitas *crossover* yang dipergunakan ($Cr \times Popsiz$ e). Pada perhitungan ini akan dihasilkan sejumlah 8, hal tersebut dikarenakan karena Cr yang digunakan adalah 0.8 ($0.8 \times 10 = 8$). Metode *crossover* yang digunakan adalah *crossover* satu titik. Pada iterasi pertama, dipilih dua *parent* secara acak. *Parent* yang terpilih pada iterasi pertama adalah *parent* 4

Gambar 4.7 Nilai *Fitness* Seluruh Individu

Parent 1 (Individu 7) = 0.7407																				
0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1
Parent 2 (Individu 3) = 0.8148																				
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	1	0
Parent 3 (Individu 4) = 0.7778																				
0	0	0	1	1	1	1	0	0	0	1	1	0	0	1	0	0	0	0	1	1
Parent 4 (Individu 9) = 0.8148																				
1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0
Parent 5 (Individu 7) = 0.8333																				
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1
Parent 6 (Individu 6) = 0.7963																				
1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0
Parent 7 (Individu 5) = 0.8333																				
1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0
Parent 8 (Individu 1) = 0.7693																				
1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
Parent 9 (Individu 7) = 0.8333																				
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1
Parent 10 (Individu 3) = 0.8148																				
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	1	0
Offspring 1 = 0.7963																				
1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
Offspring 2 = 0.7407																				
1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
Offspring 3 = 0.7407																				
0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	1	0	0	0	1	1
Offspring 4 = 0.8333																				
0	0	0	1	1	1	1	0	0	0	1	1	0	0	0	0	1	0	0	0	1
Offspring 5 = 0.7778																				
1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	1
Offspring 6 = 0.8148																				
0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	1
Offspring 7 = 0.8333																				
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	1	1	0	0
Offspring 8 = 0.7963																				
1	1	1	0	0	1	1	1	0	0	0	0	0	0	0	1	0	0	1	1	0
Offspring 9 = 0.7778																				
0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	1	1	1	0

Setelah diperoleh hasil *fitness* seluruh individu, selanjutnya dilakukan pemilihan 10 individu dengan *fitness* terbaik untuk diteruskan ke generasi berikutnya dan untuk 9 individu lainnya tidak lagi digunakan untuk generasi berikutnya. Hasil pemilihan 10 individu terbaik diperlihatkan pada Tabel 4.8.

Gambar 4.8 Hasil Elitism

Parent 5 (Individu 7) = 0.8333																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	
Parent 9 (Individu 7) = 0.8333																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	
Parent 7 (Individu 5) = 0.8333																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	1	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	
Offspring 4 = 0.8333																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	1	1	1	1	1	0	0	0	0	1	1	0	0	0	0	1	1	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	
Offspring 7 = 0.8333																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	1	0	1	0	1	1	1	0	0	0	1	1	0	0	0	0	0	0	0	
Offspring 6 = 0.8148																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	
Parent 2 (Individu 3) = 0.8148																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	1	0	1	0	1	1	1	1	0	0	0	1	0	1	0	1	1	0	0	
Parent 10 (Individu 3) = 0.8148																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
0	1	0	1	0	1	1	1	1	0	0	0	1	0	1	0	1	1	0	0	
Parent 6 (Individu 6) = 0.7963																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
1	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	
Offspring 8 = 0.7963																				
Age	Weight	Height	BMI	Pulse	RR	Hb	CL	Marriage	Pregnant	Abortion	B-HCG	FSH	LH	FSHLH	Hip	Waist	WHRatio	TSH	AMH	
1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	
PRL	D3	PRG	RBS	WG	Skindark	HL	Pimples	FastFood	Exercise	Systolic	Diastolic	FollicleL	FollicleR	AvgFsizeL	AvgFsizeR	Endometrium	Blood	Cycle	HG	
1	0	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1	1	0	0	

S U R A B A Y A

Tabel 4.8 Hasil Seleksi Fitur Algoritma Genetika

Data	Kernel	Popsi	Fitur	Fitness	Time
90:10	Linear	40	19	0.7313	77
		60	18	0.7764	88
		80	20	0.8042	109
	Polynomial	40	25	0.7505	83
		60	23	0.7811	99
		80	27	0.8011	119
	RBF	40	21	0.8441	84
		60	19	0.8564	110
		80	14	0.8612	127
Rata-Rata				0.8007	99.56
80:20	Linear	40	24	0.759	79
		60	25	0.7771	99
		80	19	0.8071	110
	Polynomial	40	21	0.8013	83
		60	25	0.8065	90
		80	24	0.8145	111
	RBF	40	18	0.8519	94
		60	13	0.8814	99
		80	19	0.8854	106
Rata-Rata				0.8205	96.78
70:30	Linear	40	25	0.7819	69
		60	20	0.7967	84
		80	13	0.8057	98
	Polynomial	40	21	0.8271	78
		60	23	0.8455	94
		80	16	0.8518	110
	RBF	40	18	0.8611	77
		60	16	0.8827	92
		80	19	0.8929	102
Rata-Rata				0.8384	89.33

Hasil terbaik seleksi fitur algoritma genetika ialah pada saat pembagian data sebesar 70:30 dengan menggunakan Kernel RBF pada SVM dan *popsi* pada algoritma genetika sebesar 80. Fitur yang didapatkan dari parameter terbaik tersebut berjumlah 19 fitur yaitu: BMI; *Blood Group*; RR; HB; *Cycle*; *Marriage Status*; FSH; Hip; TSH; PRG; RBS; *Weight Gain*; *Hair Growth*; *Fast Food*; *Reg.Exercise*; *BPDiastolic*; *Follicle No (L)*; *Follicle No (R)*; dan Endometrium.

Tabel 4.11 Hasil Klasifikasi SVM

K-Fold	Kernel	Cost	Akurasi	Sensitivitas	Spesifisitas	Waktu (s)
5	Linear	0.1	0.8160	0.6402	0.9077	16
		1	0.8239	0.6556	0.9084	
		10	0.8144	0.6250	0.9077	
		100	0.8136	0.6165	0.9090	
	Polinomial	0.1	0.8028	0.5272	0.9297	9
		1	0.8231	0.5352	0.9634	
		10	0.8128	0.5474	0.9423	
		100	0.8045	0.4966	0.9909	
	RBF	0.1	0.8280	0.6866	0.9023	13
		1	0.8346	0.6053	0.9437	
		10	0.8312	0.6142	0.9389	
		100	0.8252	0.6670	0.9072	
Rata-rata			0.8192	0.6014	0.9293	12.67
10	Linear	0.1	0.8311	0.6652	0.9106	17
		1	0.8367	0.6250	0.9394	
		10	0.8163	0.6875	0.8788	
		100	0.7959	0.6251	0.8787	
	Polinomial	0.1	0.8152	0.5318	0.9512	11
		1	0.8236	0.5497	0.9537	
		10	0.8209	0.5354	0.9602	
		100	0.8200	0.5325	0.9601	
	RBF	0.1	0.8329	0.6283	0.9326	14
		1	0.8519	0.6045	0.9725	
		10	0.8210	0.6288	0.9268	
		100	0.8293	0.6957	0.9297	
Rata-rata			0.8246	0.6091	0.9329	14

Sama halnya dengan klasifikasi dengan menggunakan seleksi fitur, hasil klasifikasi menggunakan seluruh fitur yang diperlihatkan pada Tabel 4.11 menunjukkan bahwa hasil terbaik didapatkan menggunakan kernel RBF dengan nilai $C = 1$ dan pembagian data 10-fold cross validation. Confusion matrix yang pada klasifikasi SVM diperlihatkan pada Tabel 4.12.

Tabel 4.12 Confusion Matrix SVM

Kelas Sebenarnya	y_u	Kelas Prediksi	
		Normal	PCOS
	Normal	353	10
	PCOS	70	107

BAB V

PENUTUP

Dari hasil implementasi dan uji coba menggunakan GA-SVM untuk mengelompokkan data rekam medis PCOS menjadi kategori negatif dan positif, terdapat beberapa simpulan dan saran yang dapat diambil.

5.1. Kesimpulan

Dengan mempertimbangkan hasil pengolahan data sesuai dengan tujuan penelitian, dapat disimpulkan hasil penelitian sebagai berikut:

1. Dengan menggunakan metode seleksi fitur Algoritma Genetika, didapatkan hasil seleksi fitur dari 40 variabel awal yang digunakan diseleksi sehingga hanya menjadi 19 variabel yaitu: BMI; *Blood Group*; RR; HB; *Cycle*; *Marriage Status*; FSH; Hip; TSH; PRG; RBS; *Weight Gain*; *Hair Growth*; *Fast Food*; *Reg.Exercise*; *BPDiastolic*; *Follicle No (L)*; *Follicle No (R)*; dan Endometrium.
2. Klasifikasi SVM dengan seleksi fitur algoritma genetika menunjukkan hasil terbaik dengan menggunakan pembagian *10-fold cross validation* dengan menggunakan kernel RBF dan parameter $C = 1$. Hasil evaluasi yang didapatkan yaitu akurasi= 0.9426, sensitifitas= 0.8757, dan spesifisitas= 0.9752.
3. Model GA-SVM menunjukkan performa yang lebih baik dibandingkan dengan metode SVM dalam hal akurasi, sensitivitas, dan spesifisitas. Dalam

pengujian, model GA-SVM memiliki nilai rata-rata akurasi sebesar 0.9214, sensitivitas sebesar 0.8431, dan spesifisitas sebesar 0.9595. Di sisi lain, metode SVM memiliki nilai rata-rata akurasi sebesar 0.8246, sensitivitas sebesar 0.6091, dan spesifisitas sebesar 0.9329. Penggunaan seleksi fitur algoritma genetika terbukti dapat meningkatkan nilai evaluasi dalam klasifikasi SVM dibandingkan dengan klasifikasi SVM tanpa menggunakan seleksi fitur. Dikarenakan dilakukan penambahan suatu proses, maka menyebabkan waktu komputasi yang lebih lama dibandingkan klasifikasi tanpa seleksi fitur. Rata-rata Waktu yang diperlukan untuk klasifikasi dengan seleksi fitur ialah selama ± 105 detik, sedangkan tanpa seleksi fitur rata-rata waktu yang diperlukan hanya selama ± 14 detik.

5.2. Saran

Dalam penelitian ini, terdapat 540 data awal yang terdiri dari 363 data negatif dan 177 data positif PCOS. Keadaan ini dapat berdampak pada proses pembelajaran karena jumlah data pada kelas positif PCOS jauh lebih sedikit daripada data negatif. Hal ini yang dapat menyebabkan nilai spesifisitas lebih besar dibandingkan dengan nilai sensitivitas yang didapatkan dalam proses klasifikasi.

Pada penelitian selanjutnya, diharapkan peneliti dapat melakukan *balancing* pada data untuk mengatasi ketidakseimbangan data pada pasien positif dan negatif. Teknik yang dapat digunakan untuk melakukan *balancing* data diantaranya yaitu *undersampling*, *oversampling*, metode *hybrid* dan pembobotan (*weighting*).

DAFTAR PUSTAKA

- Abdilah, A., Mardiyani, E., and Safudin, M. (2018). Integrasi Algoritma Genetika Dan Information Gain Untuk Seleksi Fitur Pada Analisis Sentimen Review Film Menggunakan Algoritma Naïve Bayes Bayes. *Jurnal Teknik Komputer AMIK BSI (JTK)*, 4(1):186–193.
- Agustina, W., Furqon, M. T., and Rahayudi, B. (2018). Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2 No.(10):3366 – 3372.
- Akhmad and Marsuni, N. S. (2019). Causality Relationship of Poverty , Unemployment Rate , and Economic Growth in South Sulawesi. 15:231–241.
- Anggoro, D. A. and Kurnia, N. D. (2020). Comparison of Accuracy Level of Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Algorithms in Predicting Heart Disease. *International Journal of Emerging Trends in Engineering Research*, 8(5):1689–1694.
- Anggraeni, A. N., Mustofa, K., and Priyanta, S. (2021). Comparison of Filter and Wrapper Based Feature Selection Methods on Spam Comment Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3):245.
- Arifudin, R. (2012). Optimasi Penjadwalan Proyek Dengan Penyeimbangan Biaya Menggunakan Kombinasi Cpm Dan Algoritma Genetika. *Jurnal Masyarakat Informatika*, 2(4):1–14.

- Armanini, D., Boscaro, M., Bordin, L., and Sabbadin, C. (2022). Controversies in the Pathogenesis, Diagnosis and Treatment of PCOS: Focus on Insulin Resistance, Inflammation, and Hyperandrogenism. *International Journal of Molecular Sciences*, 23(8).
- Bianto, M. A., Kusrini, and Sudarmawan (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. *Creative Information Technology Journal*, 6(1):75.
- Binarwati, L., Mukhlash, I., and Soetrisno, S. (2017). Implementasi Algoritma Genetika untuk Optimalisasi Random Forest dalam Proses Klasifikasi Penerimaan Tenaga Kerja Baru : Studi Kasus PT.XYZ. *Jurnal Sains dan Seni ITS*, 6(2):2–6.
- Bulsara, J., Patel, P., Soni, A., and Acharya, S. (2021). A review: Brief insight into Polycystic Ovarian syndrome. *Endocrine and Metabolic Science*, 3(November 2020):100085.
- Chang, R. J. and Dumesic, D. A. (2019). *Polycystic Ovary Syndrome and Hyperandrogenic States*. Number 2. Elsevier Inc., eighth edi edition.
- Chen, R. C., Dewi, C., Huang, S. W., and Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1).
- Chen, W. and Pang, Y. (2021). Metabolic Syndrome and PCOS: Pathogenesis and the Role of Metabolites. *Metabolites*, 11(12).
- Cunha, A. and Póvoa, A. M. (2021). Infertility management in women with polycystic ovary syndrome: a review. *Porto Biomedical Journal*, 6(1):e116.

- El Hayek, S., Bitar, L., Hamdar, L. H., Mirza, F. G., and Daoud, G. (2016). Poly Cystic Ovarian Syndrome: An updated overview. *Frontiers in Physiology*, 7(APR):1–15.
- Fritz, M. A. and Speroff, L. (2011). Chronic anovulation and the polycystic ovary syndrome. In *Clinical Gynecologic Endocrinology and Infertility, 8th Edition*, pages 151–187. Lippincott Williams and Wilkins.
- Gokulnath, C. B. and Shantharajah, S. P. (2019). An optimized feature selection based on genetic approach and support vector machine for heart disease. *Cluster Computing*, 22:14777–14787.
- Gunawan, D., Riana, D., Ardiansyah, D., Akbar, F., and Alfarizi, S. (2020). Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Dengan Algoritma Genetika Pada Analisis Sentimen Calon Gubernur Jabar 2018-2023. *Jurnal Teknik Komputer AMIK BSI*, 6(2):174–180.
- Harafani, H. and Maulana, A. (2019). Penerapan Algoritma Genetika pada Support Vector Machine Sebagai Pengoptimasi Parameter untuk Memprediksi Kesuburan. *Jurnal Teknik Informatika STMIK Antar Bangsa*, V(1):51–59.
- Hestiantoro, A. and Pamungkas, D. T. (2020). Assessment of the Quality of Internet-Based Health Information in the Indonesian Language about Polycystic Ovarian Syndrome. *Indonesian Journal of Obstetrics and Gynecology*, 8(4):222–227.
- Iriany, R., Widodo, A. W., and Mahmudy, W. F. (2017). Pemanfaatan Algoritma Genetika Untuk Optimasi 0 / 1 Multi-Dimensional Knapsack Problem Dalam Pendistribusian Produk (Studi Kasus UD . TOSA). *Jurnal Pengembangan*

Teknologi Informasi dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya, 1(4):257–269.

Jabeen, A., Yamini, V., Rahman Amberina, A., Dinesh Eshwar, M., Vadakedath, S., Begum, G. S., and Kandi, V. (2022). Polycystic Ovarian Syndrome: Prevalence, Predisposing Factors, and Awareness Among Adolescent and Young Girls of South India. *Cureus*, 14(8):1–14.

Ji, H. R., Woo, H. L., Park, Y. J., Hwang, D. S., Lee, J. M., Lee, C. H., Jang, J. B., and Park, K. S. (2018). Characteristics of heart rate variability in women with polycystic ovary syndrome: A retrospective cross-sectional study. *Medicine (United States)*, 97(38):0–3.

Koncz, P. and Paralic, J. (2011). An approach to feature selection for sentiment analysis. *INES 2011 - 15th International Conference on Intelligent Engineering Systems, Proceedings*, pages 357–362.

Koneru, A. and Priyanka (2019). Polycystic Ovary Syndrome (PCOS) and Sexual Dysfunctions. *Journal of Psychosexual Health*, 1(2):154–158.

Krisnandi, K. and Agung, H. (2017). Implementasi Algoritma Genetika Untuk Memprediksi Waktu Dan Biaya Pengerjaan Proyek Konstruksi. *Jurnal Ilmiah FIFO*, 9(2):90.

Kuntjoro, D. A., Setiawan, B. D., and Perdana, R. S. (2018). Algoritme Genetika Untuk Optimasi K-Means Clustering Dalam Pengelompokan Data Tsunami. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(10):3865–3872.

Miazgowski, T., Martopullo, I., Widecka, J., Miazgowski, B., and Brodowska,

- A. (2021). National and regional trends in the prevalence of polycystic ovary syndrome since 1990 within Europe: The modeled estimates from the Global Burden of Disease Study 2016. *Archives of Medical Science*, 17(2):343–351.
- Moraes, R., Valiati, J. F., and Gavião Neto, W. P. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, 40(2):621–633.
- Mustari, A. D., Rostini, T., Indrati, A. R., Bayuaji, H., and Rachmayati, S. (2018). Korelasi Jumlah Folikel Antral dengan Kadar 25(OH)D Serum pada Penderita Sindrom Ovarium Polikistik. *Majalah Kedokteran Bandung*, 50(4):4–9.
- Nabil, D., Benali, R., and Bereksi Reguig, F. (2020). Epileptic seizure recognition using EEG wavelet decomposition based on nonlinear and statistical features with support vector machine classification. *Biomedizinische Technik*, 65(2):133–148.
- Nanda, M. A., Seminar, K. B., Nandika, D., and Maddu, A. (2018). A comparison study of kernel functions in the support vector machine and its application for termite detection. *Information (Switzerland)*, 9(1).
- Norman, R. J. and Teede, H. J. (2018). A new evidence-based guideline for assessment and management of polycystic ovary syndrome. *Medical Journal of Australia*, 209(7):299–300.
- Pahlevi, R., Fredlina, K. Q., and Utami, N. W. (2021). Penerapan Algoritma ID3 Dan SVM Pada Klasifikasi Penyakit Diabetes Melitus Tipe 2. *Prosiding Seminar Nasional Aplikasi Sains dan Teknologi (SNAST) 2021*, 2:64–75.
- Pane, S. F., Maulana Awangga, R., Rahmadani, E. V., and Permana, S.

- (2019). Implementasi Algoritma Genetika Untuk Optimalisasi Pelayanan Kependudukan. *Jurnal Tekno Insentif*, 13(2):36–43.
- Prasetio, R. T., Rismayadi, A. A., and Anshori, I. F. (2018). Implementasi Algoritma Genetika pada k-nearest neighbours untuk Klasifikasi Kerusakan Tulang Belakang. *Jurnal Informatika*, 5(2):186–194.
- Puspitasari, A. M., Ratnawati, D. E., and Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *J-Ptiik*, 2(2):802–810.
- Puspitasari, D. I. (2017). Penerapan Data Mining Menggunakan Perbandingan Algoritma Greedy Dengan Algoritma Genetika Pada Prediksi Rentet Waktu Harga Crude Palm Oil. *Elinvo (Electronics, Informatics, and Vocational Education)*, 2(1):21–26.
- Resmiati, R. and Arifin, T. (2021). Klasifikasi Pasien Kanker Payudara Menggunakan Metode Support Vector Machine dengan Backward Elimination. *Sistemasi*, 10(2):381.
- Rodrigues, A. L., de Santana, M. A., Azevedo, W. W., Bezerra, R. S., Barbosa, V. A., de Lima, R. C., and dos Santos, W. P. (2019). Identification of mammary lesions in thermographic images: feature selection study using genetic algorithms and particle swarm optimization. *Research on Biomedical Engineering*, 35(3-4):213–222.
- Saikin, Fadli, S., and Ashari, M. (2021). Optimization of Support Vector Machine Method Using Feature Selection to Improve Classification Results. *JISA(Jurnal Informatika dan Sains)*, 4(1):22–27.

- Santoso, B., Cholissodin, I., and Setiawan, B. D. (2017). Optimasi K-Means untuk Clustering Kinerja Akademik Dosen Menggunakan Algoritme Genetika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12):1652–1659.
- Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., and Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:6308–6325.
- Subha, Nayana, Radhakrishnan, R., and Sumalatha (2022). Computerized Diagnosis of Polycystic Ovary Syndrome Using Machine Learning and Swarm Intelligence Techniques. *Research Square*, pages 1–29.
- Sugeha, I. H., Inkiriwang, R. L., and Pratasih, P. A. (2019). Optimasi Penjadwalan Menggunakan Metode Algoritma Genetika Pada Proyek Rehabilitasi Puskesmas Minanga. *Jurnal Sipil Statik*, 7(12):1669–1680.
- Suparman, E. and Suparman, E. (2017). Amenorea Sekunder: Tinjauan dan Diagnosis. *Jurnal Biomedik (Jbm)*, 9(3).
- Taufiq, M. N., Dewi, C., and Mahmudy, W. F. (2017). Optimasi Komposisi Pakan Untuk Penggemukan Sapi Potong Menggunakan Algoritma Genetika. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(7):571–582.
- Teede, H., Misso, M., Costello, M., Dokras, A., Laven, J., Moran, L., Piltonen, T., and Norman, R. (2018). *International evidence-based guideline for the assessment and management of polycystic ovary syndrome 2018*.
- Tempola, F., Muhammad, M., and Khairan, A. (2018). Perbandingan Klasifikasi

- Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 5(5):577.
- Valero-Carreras, D., Alcaraz, J., and Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers and Operations Research*, 152(December 2022):106131.
- Vidya Bharathi, R., Swetha, S., Neerajaa, J., Varsha Madhavica, J., Janani, D. M., Rekha, S. N., Ramya, S., and Usha, B. (2017). An epidemiological survey: Effect of predisposing factors for PCOS in Indian urban and rural population. *Middle East Fertility Society Journal*, 22(4):313–316.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F., and Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1):134.
- Widayani, W. and Harliana, H. (2021). Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains dan Informatika*, 7(1):20–27.
- Wolf, W. M., Wattick, R. A., Kinkade, O. N., and Olfert, M. D. (2018). Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. *International Journal of Environmental Research and Public Health*, 15(11):1–13.