

Penerapan Kecerdasan Buatan Untuk Menentukan Gejala Utama Polycystic Ovary Syndrom (PCOS) Berdasarkan Data Klinis Pasien

Baiq Andriska Candra Permana^{1*}, Zulkipli², Muhammad Wasil³, Harianto⁴

^{1,2,3,4} Program Studi Informatika, Universitas Hamzanwadi

*andriska.cp@hamzanwadi.ac.id

Abstrak

Polycystic Ovary Syndrome (PCOS) adalah gangguan hormonal yang sering terjadi pada wanita usia reproduksi dan dapat menyebabkan gangguan menstruasi, infertilitas, serta masalah metabolisme. Diagnosis PCOS masih menjadi tantangan karena variasi gejala yang luas dan perbedaan respons pasien terhadap pengobatan. Oleh karena itu, penelitian ini bertujuan untuk menerapkan kecerdasan buatan (AI) dalam mengidentifikasi gejala utama yang berkontribusi terhadap PCOS berdasarkan data klinis pasien. Penelitian ini menggunakan pendekatan machine learning, dengan algoritma Support Vector Machine (SVM) sebagai metode utama dalam klasifikasi pasien dengan dan tanpa PCOS. Data yang digunakan bersumber dari rekam medis pasien, yang mencakup parameter data klinis pasien yang diperoleh dari website kaggle dengan jumlah sampel data sebanyak 541 data pasien. Tahapan penelitian meliputi pengumpulan dan preprocessing data, pemilihan fitur utama menggunakan teknik feature selection, pelatihan model dengan algoritma SVM. Model AI yang dikembangkan menghasilkan 10 fitur utama yang berpengaruh pada diagnosa dengan nilai akurasi 90.74% yang menunjukkan bahwa model memiliki kemampuan baik dalam melakukan klasifikasi penderita PCOS maupun tidak PCOS. Selain itu matriks menunjukkan adanya keseimbangan antara nilai matriks untuk presisi 87.5% , recall 82.35% dan F1 score 84.85%. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam bidang medis, khususnya dalam mendukung diagnosa dini yang lebih cepat dan akurat serta personalisasi pengobatan PCOS.

Kata kunci : Kecerdasan Buatan, PCOS, Pembelajaran Mesin, SVM, Klasifikasi.

Abstract

Polycystic Ovary Syndrome (PCOS) is a common hormonal disorder among women of reproductive age, often leading to menstrual irregularities, infertility, and metabolic issues. Diagnosing PCOS remains challenging due to the wide range of symptoms and varying patient responses to treatment. Therefore, this study aims to apply artificial intelligence (AI) to identify key symptoms contributing to PCOS based on patients' clinical data. This study employs a machine learning approach, with the Support Vector Machine (SVM) algorithm as the primary method for classifying patients with and without PCOS. The data used was sourced from patient medical records, which included clinical data parameters obtained from the Kaggle website, with a total of 541 patient data samples. The research stages include data collection and preprocessing, selection of main features using feature selection technique, model training with SVM algorithm. The AI model developed produces 10 main features that affect the diagnosis with an accuracy value of 90.74% which shows that the model has the ability to classify PCOS and non-PCOS sufferers. In addition, the matrix shows a balance between the matrix values for precision 87.5%, recall 82.35% and F1 score 84.85%. The results of this study are expected to contribute to the medical field, especially in supporting faster and more accurate early diagnosis and personalization of PCOS treatment.

Keywords : Artificial Intelligence, PCOS, Machine Learning, SVM, Classification.

1. Pendahuluan

Polycystic Ovary Syndrome (PCOS) adalah gangguan hormonal yang sering dialami oleh

wanita selama berada pada usia subur dan menjadi salah satu penyebab utama wanita mengalami kesulitan hamil [1]. Terdapat berbagai

tanda atau gejala yang dialami wanita pada kondisi ini, seperti siklus ovulasi yang tidak teratur, peningkatan kadar hormon androgen, serta keberadaan kista kecil di ovarium yang dapat dideteksi melalui pemeriksaan USG. Selain itu, PCOS juga sering dikaitkan dengan resistensi insulin, obesitas, serta peningkatan risiko penyakit metabolik, seperti diabetes tipe 2 dan tekanan darah tinggi. Karena gejala PCOS sangat bervariasi pada setiap individu, proses diagnosis bisa menjadi cukup menantang, terutama karena tidak semua penderita menunjukkan gejala yang sama [2],[3].

Kemajuan teknologi di bidang kecerdasan buatan (*Artificial Intelligence/AI*) telah membawa perubahan besar dalam dunia medis, terutama dalam membantu tenaga kesehatan mendiagnosis berbagai penyakit secara lebih cepat dan akurat [4],[5]. AI telah digunakan dalam berbagai aspek kedokteran, mulai dari analisis pencitraan medis, identifikasi pola penyakit, hingga pengolahan data klinis untuk mendukung pengambilan keputusan medis. Salah satu cabang AI yang banyak dimanfaatkan dalam dunia medis adalah Machine Learning (ML), yang memungkinkan sistem komputer untuk mempelajari pola dari data pasien dan membuat prediksi berdasarkan informasi tersebut[6]. Apabila dibandingkan dengan metode konvensional, penerapan AI dalam dunia medis menawarkan kecepatan, akurasi, dan efisiensi yang lebih

tinggi[7],[8].

Support Vector Machine (SVM) adalah satu dari sekian banyak algoritma *machine learning* yang banyak digunakan dalam diagnosis berbagai penyakit, termasuk kanker, diabetes, penyakit jantung, dan sindrom ovarium polikistik [9],[10]. Algoritma SVM merupakan salah satu teknik yang sering digunakan dalam klasifikasi medis khususnya pada diagnose PCOS karena memiliki kemampuan dalam mengidentifikasi pola kompleks, akan tetapi belum banyak yang secara spesifik menerapkan Support Vector Machine (SVM) dalam identifikasi gejala utama yang paling berpengaruh terhadap diagnose pasien. Berdasarkan kondisi tersebut maka perlu adanya model yang dapat membantu dokter atau pihak medis dalam memahami faktor paling dominan yang mempengaruhi PCOS, bukan hanya melakukan klasifikasi pasien sebagai "PCOS" atau "tidak PCOS". Oleh karena itu, penelitian ini bertujuan untuk menerapkan metode kecerdasan buatan, khususnya algoritma Support Vector Machine (SVM), dalam mengidentifikasi gejala utama yang berperan dalam diagnosis PCOS berdasarkan data klinis pasien. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan medis berbasis data, guna meningkatkan akurasi dan efisiensi diagnosis PCOS secara dini.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dari penelitian ini, penulis mengangkat beberapa penelitian sebagai referensi dalam memperkaya bahan kajian. Berikut merupakan penelitian beberapa jurnal terkait dengan penelitian yang dilakukan penulis, yaitu;

- Penelitian tahun 2020 oleh B.Andriska CP dan M.Djamaluddin dalam jurnal Infotek : Jurnal Informatika dan Teknologi yang berjudul “Penerapan Python Dalam *Data Mining* Untuk Prediksi Kanker Paru” menjelaskan bahwa penggunaan data mining dengan algoritma C4.5 telah mampu melakukan klasifikasi terhadap kanker paru dengan tingkat akurasi sebesar 86%. Hal ini membuktikan bahwa metode yang ditemukan dapat dijadikan sebagai acuan dalam pengembangan system kecerdasan buatan^[11]
- Penelitian tahun 2024 2024 oleh Rahmat H, Yandiko S, Teguh S dkk. Dalam jurnal Teknologi Informasi dan Rekayasa Komputer yang berjudul “Implementasi *Machine Learning* Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector machine” yang membahas tentang teknik prediksi penyakit jantung menggunakan algoritma SVM yang berhasil membuat system library streamlit yang dapat dengan mudah diakses melalui situs web. Dari

penelitian ini didapatkan kemampuan SVM dalam memprediksi penyakit jantung memiliki tingkat akurasi 85% ^[12].

- Penelitian tahun 2023 oleh Nahdira dan Kusnawi dalam jurnal Jutif : Jurnal Teknik Informatika yang berjudul “PCOS Disease Classification Using Feature Selection RFECV and EDA with KNN Algorithm method” menjelaskan bahwa hasil dari seleksi fitur EDA menghasilkan 10 atribut penting dan RFECV menghasilkan 7 atribut penting, dan algoritma KNN memiliki tingkat akurasi 93% untuk melakukan klasifikasi PCOS ^[13].
- Penelitian tahun 2025 oleh Ridwan CH, Mahesa R dkk dalam jurnal Jati : Jurnal Mahasiswa Teknik Informatika yang berjudul “Sinergi AI dan *Machine Learning* untuk Prediksi Multikeluhan Pada Diagnosis Penyakit Kepala” menjelaskan bahwa algoritma seperti Deep Neural Networks (DNN) dan Multilayer Perceptron (MLP) berhasil meningkatkan akurasi diagnosis hingga 99,6%. AI juga membuktikan kemampuannya dalam memproses data medis yang kompleks, menghasilkan prediksi yang lebih cepat dan akurat. Namun, penelitian ini menemukan sejumlah tantangan, termasuk variabilitas data medis keterbatasan teknik klasifikasi, dan perlunya validasi model pada populasi yang lebih luas^[14]

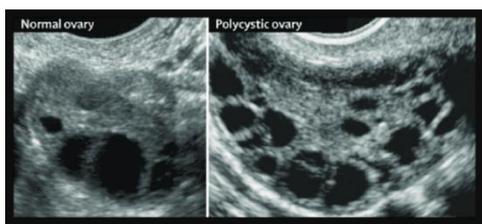
- Penelitian tahun 2025 oleh Tisa PD dan Nurul AS dalam jurnal Infotekmesin yang berjudul "Analisis Kinerja *Ensemble Learning* dan Algoritma Tunggal dalam Klasifikasi Sindrom Ovarium Polikistik Menggunakan *Random Forest, Logistic Regression* dan *XGBoost*" menjelaskan bahwa penelitian ini menggunakan metode *ensemble learning* dengan teknik *voting classifier* menggabungkan beberapa model klasifikasi lainnya. Hasil yang diperoleh metode ini memiliki performa yang lebih baik di banding KNN yaitu dengan tingkat akurasi 95% [15]

2.2. Landasan Teori

1. Polycystic Ovary Syndrome (PCOS).

Polycystic Ovary Syndrome (PCOS) merupakan gangguan endokrin kompleks yang memengaruhi 4–20% wanita usia reproduktif.

Diagnosis PCOS umumnya didasarkan pada kriteria Rotterdam yang di tandai oleh terjadinya gangguan ovulasi, peningkatan kadar hormon *androgen* (hiperandrogenisme), dan ovarium yang membesar dengan banyak kista kecil di dalamnya yang terlihat melalui ultrasonografi [16].



Gambar 1. Ovarium Normal dan Ovarium dengan PCOS [17]

PCOS ditandai oleh ketidak seimbangan hormonal, termasuk peningkatan hormon luteinizing (LH) relatif terhadap follicle-stimulating hormone (FSH), yang menyebabkan peningkatan produksi *androgen* oleh ovarium.

Hiperandrogenisme ini berkontribusi pada gejala seperti *hirsutisme*, jerawat, dan *alopecia*. Selain itu, resistensi insulin sering ditemukan pada penderita PCOS, yang memperburuk kondisi metabolik dan meningkatkan risiko diabetes tipe 2. Resistensi insulin memainkan peran penting dalam memperburuk *hiperandrogenisme* dan memperparah gejala PCOS. Kondisi ini menyebabkan tubuh memproduksi lebih banyak insulin, yang kemudian merangsang ovarium untuk menghasilkan lebih banyak hormon androgen [1,18].

2. Kecerdasan Buatan dalam Dunia Medis

Teknologi kecerdasan buatan yang di kenal dengan AI merupakan teknologi baru dalam dunia komputer yang dimanfaatkan untuk penelitian, mengembangkan teori, metode, teknik serta simulasi, penyuluhan serta perluasan kecerdasan manusia [19]. Beberapa tahun belakangan teknologi ini menjadi perbincangan karena di khawatirkan akan menggantikan layanan kesehatan di masa depan. Tentunya teknologi AI tidak akan mampu menggantikan peran manusia secara keseluruhan namun dapat dijadikan sebagai alat bantu diagnosa maupun pelayanan dalam kesehatan tentunya dengan kontrol

manusia. Beberapa bentuk pemanfaatan nyata AI dalam bidang kesehatan seperti potensi AI yang baik dalam mendiagnosa penyakit lambung, usus, maupun kanker lambung. Selain itu AI bahkan di manfaatkan pada terapi, tindakan pembedahan bahkan hingga tahap penyembuhan sehingga perawatan bedah menjadi lebih minim invasif, gambar menjadi lebih jelas.

3. Pembelajaran Mesin (*Machine learning*)

Pembelajaran mesin merupakan suatu tools yang dapat merubah data menjadi informasi yang diperlukan untuk menganalisis data. *Machine learning* banyak dimanfaatkan dalam dunia medis seperti untuk diagnosa awal penyakit dengan menggunakan algoritma *machine learning*. Salah satu tipe pembelajaran dalam *machine learning* adalah *supervised learning* dimana setiap input memiliki keluaran yang sudah di ketahui. Salah satu model *supervised learning* yang populer adalah Support Vector Machine (SVM) [12].

4. Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang dapat melakukan prediksi berdasarkan pola pada proses training. Klasifikasi yang di lakukan pada algoritma ini adalah dengan memisahkan kelas menggunakan garis pembatas (*hyperplane*) antara opini positif dengan opini negatif. Garis pembatas dikatakan baik ketika jarak terbesar memiliki jarak yang besar ke titik data pelatihan terdekat dari tiap kelas.

3. Metode Penelitian

3.1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan melakukan pengujian pada data klinis pasien yang bertujuan untuk menghasilkan suatu model klasifikasi yang dapat digunakan untuk mendeteksi kemungkinan seseorang menderita PCOS berdasarkan data klinis pasien tersebut. Metode kecerdasan muatan yang digunakan adalah Support Vector Machine (SVM) yang dikenal memiliki efektifitas yang baik dalam melakukan klasifikasi pada data yang berdimensi tinggi. Bahasa pemrograman yang digunakan adalah python pada seluruh tahapan. *Pandas* dan *NumPy* sebagai manipulasi dan analisis sementara implementasi kode menggunakan *IDE Visual Studio Code*.

3.2. Data Penelitian

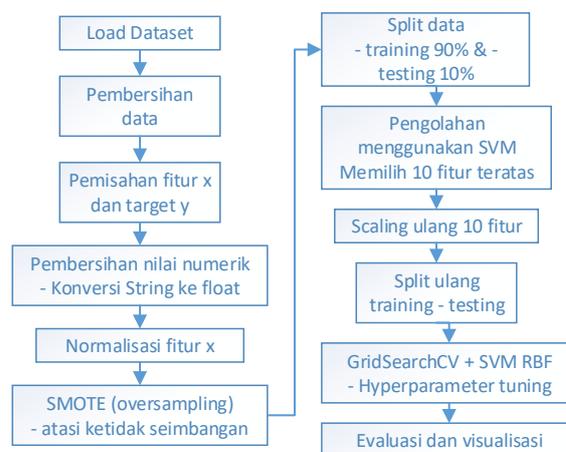
Penelitian ini menggunakan sumber data yang berasal dari situs kaggle berupa dataset klinis pasien yang terdiri dari berbagai parameter. Penelitian ini menggunakan 541 entri data pasien dan 41 fitur atribut klinis. Berbagai variable yang digunakan pada penelitian ini meliputi:

- *Variable independent* :

Tabel 1. *Variable independen*

Age	PulseRate,RR(breaths/min)
Weight	Cycle length(days)
Haight	BloodGroup
Hb(g/dl)	Avg F size (R) (mm)
Cycle(R/I)	Endometrium (mm).
BMI	Follicle No. (R)

Pregnant(Y/N)	Follicle No. (L)
No. of abortions	Marriage Status (Yrs)
I beta-HCG(mIU/mL)	FSH(mIU/mL)
II beta-HCG(mIU/mL)	LH(mIU/mL)
FSH/LH,Hip(inch)	Waist(inch)
Waist:Hip Ratio	PRL(ng/mL)
TSH (mIU/L)	Vit D3 (ng/mL)
AMH(ng/mL)	PRG(ng/mL)
RBS(mg/dl)	Weight gain(Y/N)
Skin darkening (Y/N)	hair growth(Y/N)
Hair loss(Y/N)	Pimples(Y/N)
Fast food (Y/N)	Reg.Exercise(Y/N)
BP_Systolic (mmHg)	BP_Diastolic (mmHg)
Avg. F size (L) (mm)	

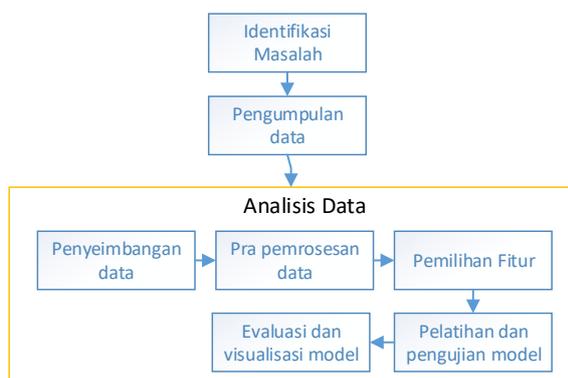


Gambar 3. Tahap Analisis Data

- *Variable dependent* (target) : Hasil diagnosa PCOS digunakan sebagai label dimana 1 untuk menyatakan penderita dan 0 bukan penderita.

3.3. Tahapan Penelitian

Alur tahapan penelitian sebagaimana ditunjukkan pada gambar berikut :



Gambar 2. Tahapan Penelitian

Tahapan yang lebih rinci dari proses analisis data pada penelitian ini sebagaimana ditunjukkan pada gambar berikut :

Dataset yang sudah ada di bersihkan dan pisahkan antara *variable* atau *fitur* (x) dan target (y). Konversi *String* ke *float* pada data digunakan karena SVM data bekerja dalam bentuk numerik. Normalisasi pada *fitur input* (x) dilakukan untuk menghindari dominasi dari fitur bersekala besar karena model akan fokus pada fitur bersekala besar walaupun memiliki kontribusi yang kecil. Misalnya saja dalam penelitian ini pada *fitur* TSH (mIU/L) yang memiliki satuan puluhan sementara *Waist:Hip Ratio* nilai desimal kecil seperti (0.83). Teknik penyeimbangan data menggunakan teknik SMOTE (Synthetic Minority Over Sampling Technique) untuk menyeimbangkan data yang terdapat pada kelas mayoritas dan minoritas sehingga performa model menjadi lebih seimbang dan adil. Ketika data jumlah pasien penderita PCOS (1) jauh lebih banyak dari data jumlah pasien yang tidak PCOS (0) atau sebaliknya dan tidak di tangani, maka kemungkinan akurasi akan tinggi akan tetapi *F1 score* dan *recall* untuk data yang minoritas kemungkinan akan menjadi

sangat rendah. Oleh sebab itu diperlukan SMOTE untuk mengatasi ketidak seimbangan data untuk memastikan model lebih sensitif terhadap fitur (x) dan tidak mengabaikan data minoritas. Interpolasi SMOTE dinyatakan dengan :

$$X_{\text{baru}} = X_i + \delta (X_{z_i} - X_i) \quad (1)$$

X_i = data minoritas (asli)

X_{z_i} = tetangga terdekat ke z dari X_i

$\delta \in [0,1]$ = bilangan acak dari distribusi uniform.

Data yang telah di seimbangkan dengan teknik SMOTE selanjutnya di *split* 90% *training* dan 10% *testing*. Data diolah dengan menggunakan model kecerdasan buatan yaitu model SVM dengan kernel RBF dengan tujuan merubah data pada dimensi yang lebih tinggi sehingga lebih mudah untuk di pisahkan. SVM dengan kernel RBF di nyatakan dengan :

$$\max_{\alpha} \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j K(x_i x_j) \quad (2)$$

dengan syarat:

$$0 \leq a_i \leq c, \sum_{i=1}^n a_i y_i = 0$$

Keterangan :

a_i = bobot dari SVM

C = parameter regulasi

$a_i \in \{-1, 1\}$ = label kelas

$K(x_i x_j)$ = kernel RBF

Hyperparameter tuning menggunakan *GridSearchCV* sebagai alat bantu dalam melakukan kombinasi silang (*cross validation*) untuk mengatur parameter c dan gamma. Adapun hyperparameter yang digunakan pada penelitian

Tabel 2. *Hyperparameter*

Parameter	Nilai	Ket
C	0.01, 0.1, 1, 10, 100	Cara kerja parameter menghindari misklasifikasi pada tiap sample
Gamma	1, 0.1, 0.01, 0.001, 0.0001	Menentukan pengaruh dari satu sample data laltih.
Kernel	RBF	Memetakan data ke dimensi lebih tinggi sehingga mudah di pisahkan oleh SVM

Hal ini dilakukan untuk memastikan model yang digunakan merupakan model terbaik berdasarkan performa terbaik dari hasil pelaltihan yang dilakukan pada data pasien. Tahap evaluasi akan menghasilkan luaran berupa :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

Keterangan :

TP= *True Positive*

TN= *True Negative*

FP= *False Positive*

FN = *False Negative*

$$F1 \text{ score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (6)$$

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

Berdasarkan penelitian yang dilakukan dengan menggunakan model Support Vector Machine (SVM) dengan karnael RBF (Radial Basic Function) yang dapat melakukan klasifikasi terhadap pasien yang menderita PCOS atau tidak

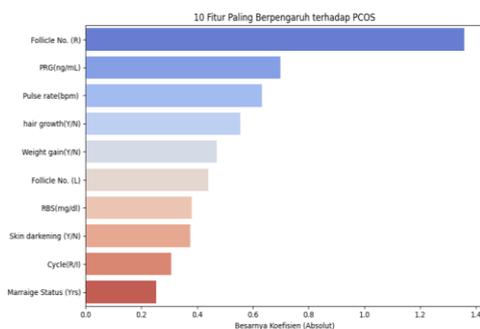
dilihat berdasarkan data klinis pasien, didapatkan model terbaik dengan SVC nilai $c= 100$ dan $\gamma = 0.001$ diperoleh model dengan hasil *Hyperparameter tuning* dengan 5 *cross validation* sebagai berikut :

Tabel 3. Matriks evaluasi

Matriks Evaluasi	Nilai
Akurasi	90.74%
Presisi	87.50%
Recall	82.35%
F1	84.85%

Hasil akurasi yang di dapatkan cukup tinggi yaitu 90.74% yang menunjukkan bahwa model yang digunakan memiliki kemampuan yang baik dalam melakukan pengklasifikasian data dan memprediksi pasien penderita PCOS dan tidak PCOS. Selain itu matriks menunjukkan adanya keseimbangan antara nilai matriks untuk presisi, *recall* dan *F1 score*.

Berikut adalah visualisasi fitur dominan penyebab PCOS yang di dapatkan dari pengolahan data klinis pasien :



Gambar 4. Fitur dominan penyebab PCOS

Tabel 4. *Fitur* paling berpengaruh

Follicle No.(R)= 1.3588	Jumlah folikel ovarium kanan
PRG (ng/ml)=0.6985	Kadar progesteron

Pulse rate (bpm)=0.6319	Detak Jantung
Hair Growth (Y/N)=0.5553	Pertumbuhan rambut berlebih
Weight gain (Y/N)=0.4703	Kenaikan berat badan
Follicle No. (L)=0.4404	Jumlah folikel ovarium kiri
RBS (mg/dL)=0.3800	Gula darah
Skin darkening (Y/N)=0.3762	Penggelapan kulit
Cycle (R/I)=0.3067	Pola siklus haid
Marriage Status (Yrs)=0.2530	Lama pernikahan

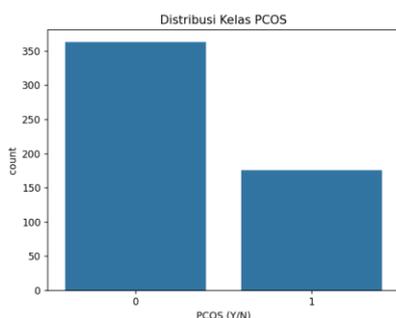
Berdasarkan visualisasi yang diperoleh pada penelitian ini menunjukkan adanya 10 faktor dari berbagai faktor yang mempengaruhi diagnosa PCOS. *Fitur* (x) yang dianggap paling berpengaruh diantaranya *follicle*, kadar *progesteron*, detak jantung, gejala fisik seperti pertumbuhan rambut yang tidak wajar serta kenaikan berat badan. Dapat dikatakan indikator hormon yang mengatur sistem reproduksi serta indikator metabolik yang mengatur fungsi metabolisme seperti pemrosesan gula darah, lemak dan berat badan memiliki pengaruh yang besar terhadap diagnosa PCOS.

4.2. Pembahasan

Hasil penelitian menunjukkan penggunaan SVM dengan kernel RBF dengan penggunaan 90% data *training* dan 10% data *testing* dapat dikatakan efektif dalam mengklasifikasikan PCOS berdasarkan data gejala klinis pada pasien. Besarnya nilai akurasi yang dihasilkan yaitu 90.74% menunjukkan bahwa model mampu melakukan klasifikasi pada data dengan tingkat akurasi yang tinggi. Nilai *F1 score* 84.85% menunjukkan tingkat keseimbangan dan

kemampuan yang baik pada model dalam mendeteksi PCOS. Presisi 87.50% menunjukkan bahwa model mampu memprediksi hampir 88% penderita positif PCOS dengan benar, dan dari seluruh pasien yang benar menderita PCOS model berhasil mendeteksi 82.35% diantaranya yang di tujukan oleh nilai *recall*.

Faktor utama paling berpengaruh sebagai diagnosa PCOS adalah jumlah folikel pada ovarium kanan. Semakin tinggi jumlah folikel ovarium kanan maka semakin besar kemungkinan seseorang terdiagnosa PCOS. Pada hasil terlihat yaitu jumlah koefisien folikel pada ovarium kanan sebesar 1.3588 dan kadar progesteron atau PRG sebesar 0.6985. Hal ini menunjukkan bahwa kontribusi dari jumlah folikel pada ovarium kanan memiliki kontribusi yang lebih besar dibandingkan dengan kadar progesteron maupun *fitur* (x) yang lainnya. Semakin besar koefisien nya maka *fitur* akan dianggap memiliki pengaruh yang semakin besar. Penggunaan karnel RBF pada model SVM dianggap memiliki kesesuaian untuk digunakan pada dataset ini karena data klinis pasien PCOS kompleks dan tidak linier.



Gambar 5. Distribusi kelas PCOS

Penggunaan teknik SMOTE untuk melakukan penyeimbang data pada dataset sangat penting untuk dapat menghasilkan performa yang baik, karena tanpa SMOTE maka terdapat ketidak seimbangan antara jumlah data pasien PCOS dan tidak PCOS sehingga dapat menyebabkan model cenderung menjadi bias pada data mayoritas.

Penelitian ini menggunakan karnel RBF dengan kombinasi c dan *gamma* 5 x 5 sehingga menghasilkan 25 kombinasi dimana masing masing kombinasi di uji dengan 5 *cross validasi* sehingga terdapat total 125 model pengujian.

Jika dibandingkan dengan penelitian sejenis yang telah dilakukan sebelumnya, penelitian ini menggunakan model klasifikasi yang berbeda serta persentase training testing digunakan pada penelitian ini 90% : 10 % untuk jumlah data yang sama. Terdapat kesamaan yang menunjukkan bahwa faktor paling dominan sebagai penyebab terjadinya PCOS adalah jumlah folikel pada ovarium kanan. Nilai.

5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan menggunakan algoritma SVM terdapat 10 *fitur* paling berpengaruh sebagai penyebab PCOS. Jumlah folikel ovarium kanan, pertumbuhan rambut yang berlebih serta kenaikan berat badan menjadi 3 faktor dominan diagnosa PCOS. SVM memiliki performa yang

baik dalam melakukan klasifikasi yaitu 86.11%. Namun *recall* belum optimal yaitu 68.97% yang artinya masih ada kasus PCOS yang belum terdeteksi dengan baik. Model ini dapat digunakan untuk membantu medis untuk melakukan diagnosa awal PCOS khususnya untuk diagnosa klinis, tapi tidak untuk menggantikan peran medis sepenuhnya. Perlu kiranya untuk menambah data ataupun perbaikan model untuk mendapatkan akurasi yang lebih baik pada kasus PCOS. Untuk penelitian selanjutnya di harapkan untuk dapat mengeksplorasi algoritma pembelajaran mesin lain seperti *Random Forest* atau *Neural Network* guna meningkatkan akurasi dan generalisasi model. Selain itu, penggunaan dataset yang lebih besar dan beragam, termasuk data dari institusi medis yang berbeda jug dapat meningkatkan validitas hasil. Penambahan fitur non-klinis seperti pola hidup dan riwayat keluarga juga dapat dipertimbangkan untuk membangun model yang lebih komprehensif..

6. Daftar Pustaka

- [1] J. P. Husada And A. Akbar, "Peran Olahraga Dalam Mengelola Gejala Pcos," *Jurnal Pandu Husada*, Vol. 6, No. 2, 2025, [Online]. Available: <https://jurnal.umsu.ac.id/index.php/jph>
- [2] Dakhaz Mustafa Abdullah And Adnan Mohsin Abdulazeez, "Machine Learning Applications Based On SVM Classification : A Review," *Qubahan Academic Journal*, Vol. 3, No. 4, Pp. 206–218, Nov. 2023, Doi: 10.48161/Issn.2709-8206.
- [3] D. Triyanto, "Support Vector Machine Untuk Klasifikasi Penyakit Diabetes Mellitus," *Media Teknologi Dan Informatika*, Vol. 1, Pp. 2024–147, 2024, [Online]. Available: [Http://jurnal.bsi.ac.id/index.php/mti](http://jurnal.bsi.ac.id/index.php/mti)
- [4] E. Y. Kurniawati, S. Hadisaputro, And A. Suwandono, "Profil Klinis Wanita Dengan Sindrom Ovarium Polikistik," *Media Ilmu Kesehatan*, Vol. 11, No. 2, Mar. 2023, Doi: 10.30989/Mik.V11i2.762.
- [5] W. Q. Salsabila, K. Adyani, And F. Realita, "Literatur Review: Faktor Resiko Sindrom Ovarium Polikistik Pada Remaja," *Journal Of Health (Joh)*, Vol. 11, No. 02, Pp. 164–174, Jun. 2024, Doi: 10.30590/Joh.V11n2.832.
- [6] B. A. Candra Permana And I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree Dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Infotek : Jurnal Informatika Dan Teknologi*, Vol. 4, No. 1, Pp. 63–69, Jan. 2021, Doi: 10.29408/Jit.V4i1.2994.
- [7] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan machine learning dalam prediksi tingkat kasus penyakit di Indonesia," *Journal of Information System Management (JOISM)*, vol. 5, no. 1, 2023.
- [8] M. A. Thariq, M. Dimas, M. Baskara, R. A. Chaniago, D. Christin, and I. Ernawati, "Systematic literature review: Analisis penerapan kecerdasan buatan dalam bidang kesehatan," in *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, Jakarta, Indonesia, Apr. 25, 2024. e-ISSN: 2962-6129..
- [9] H. Z. Elnaz Putri and H. Fahmi, "Implementasi Metode Support Vector Machine pada Klasifikasi Diagnosis Penyakit Hipertensi," *Jurnal Riset Mahasiswa Matematika*, vol. 3, no. 5, pp. 241–250, Jun. 2024, doi: 10.18860/jrmm.v3i5.27312.
- [10] A. Putranto, N. L. Azizah, I. Ratna, I. Astutik, F. Sains, and D. Teknologi, "Sistem

- Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit,” Apr. 2023. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>
- [11] B. A. C. Permana and M. Djamaluddin, “Penerapan Python Dalam Data Mining Untuk Prediksi Kanker Paru,” *Infotek : Jurnal Informatika dan Teknologi*, vol. 6, no. 2, pp. 470–477, Jul. 2023, doi: 10.29408/jit.v6i2.17816.
- [12] R. Hidayat, Y. S. Sy, T. Sujana, M. Husnah, H. T. Saputra, and F. Okmayura, “Implementasi Machine Learning Untuk Prediksi Penyakit Jantung Menggunakan Algoritma Support Vector Machine,” *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 5, no. 2, pp. 161–168, Sep. 2024, doi: 10.37148/bios.v5i2.152.
- [13] N. T. Pitaloka, “PCOS disease classification using feature selection RFECV and EDA with KNN algorithm method,” *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 4, pp. 693–701, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.693.
- [14] H. R. Cahya *et al.*, “Sinergi AI dan machine learning untuk prediksi multikeluhan pada diagnosis penyakit kepala: Systematic literature review,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, Feb. 2025.
- [15] T. Permatasari Djaka and A. S. Winarsih, “Analisis Kinerja Ensemble Learning dan Algoritma Tunggal dalam Klasifikasi Sindrom Ovarium Polikistik Menggunakan Random Forest, Logistic Regression, dan XGBoost,” vol. 16, no. 01, 2024, doi: 10.35970/infotekmesin.v16i1.2504.
- [16] A. Ghafari, M. Maftoohi, M. E. Samarin, S. Barani, M. Banimohammad, and R. Samie, “The last update on polycystic ovary syndrome(PCOS), diagnosis criteria, and novel treatment,” Mar. 01, 2025, *Elsevier B.V.* doi: 10.1016/j.endmts.2025.100228.
- [17] A. Agrawal, R. Ambad, R. Lahoti, P. Muley, and P. Pande, “Role of artificial intelligence in PCOS detection,” Apr. 01, 2022, *Wolters Kluwer Medknow Publications*. doi: 10.4103/jdmimsu.jdmimsu_278_22.
- [18] N. Komang Widiastuti, K. Dwi, A. Mas Putrayana, And I. G. Widhiantara, “Resistensi Insulin Dan Kaitannya Dengan Hiperandrogenisme Pada Penderita Pcos,” 2021.
- [19] J. T. Atmojo *Et Al.*, “Artificial Intelligence Dalam Praktik Kesehatan.” [Online]. Available: [Http://Journal.Stikeskendal.Ac.Id/Index.Php/Pskm](http://Journal.Stikeskendal.Ac.Id/Index.Php/Pskm)