Deteksi Faktor Risiko Polycystic Ovarian Syndrome (PCOS) pada Wanita Menggunakan Algoritma Support Vector Machine

Syahratul Muthi'ah M Masiming
Program Studi Informatika PJJ
Fakultas Informatika
Telkom University
Bandung, Indonesia
mutiaahm@student.telkomuniversity.ac.id

Fitriyani
Program Studi Informatika PJJ
Fakultas Informatika
Telkom University
Bandung, Indonesia
fitriyani@telkomuniversity.ac.id

Lazuardy Syahrul Darfiansa
Program Studi Informatika PJJ
Fakultas Informatika
Telkom University
Bandung, Indonesia
lazuardysyahrul@telkomuniversity.ac.id

***Abstrak* — Polycystic Ovarian Syndrome (PCOS) merupakan gangguan hormonal yang umum terjadi pada perempuan usia produktif dan sering kali tidak terdiagnosis secara dini karena keterbatasan waktu dan biaya pemeriksaan laboratorium. Penelitian ini bertujuan untuk mendeteksi faktor risiko PCOS serta mengklasifikasikan pasien PCOS dan Non-PCOS menggunakan algoritma Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Dataset yang digunakan berasal dari sumber terbuka Kaggle dan terdiri dari 1000 data pasien dengan beberapa atribut klinis dan fisik. Tahapan penelitian meliputi preprocessing data, penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE), optimasi hyperparameter menggunakan GridSearchCV, serta evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan F1-Score. Hasil penelitian menunjukkan bahwa model SVM dengan kernel RBF mampu mengklasifikasikan pasien PCOS dan Non-PCOS dengan tingkat akurasi sebesar 97%. Selain itu, penelitian ini mengidentifikasi tiga faktor risiko yang paling berpengaruh terhadap PCOS, yaitu ketidakteraturan siklus menstruasi, Body Mass Index (BMI), dan kadar hormon testosteron. Dengan demikian, model yang diusulkan berpotensi menjadi sistem pendukung keputusan medis untuk deteksi dini PCOS secara lebih cepat dan efisien.**

Kata kunci: pcos, support vector machine, kernel rbf, deteksi faktor risiko, klasifikasi pcos

I. PENDAHULUAN

Kesehatan memegang peranan penting dalam menunjang kehidupan, khususnya bagi perempuan. Bagi perempuan, mereka memiliki kerentanan terhadap gangguan kesehatan terutama pada organ reproduksi, seperti ovarium. Gangguan pada ovarium dapat menibulkan masalah kesehatan serius, salah satunya adalah *Polycystic Ovarian Syndrome* (PCOS). PCOS dilkasifikasikan sebagai masalah hormonal yang sering ditemukan pada perempuan dan banyak dialami oleh wanita usia produktif atau subur [1]. PCOS memengaruhi sistem reproduksi wanita yang ditandai dengan beberapa gejala, seperti menstruasi yang tidak teratur, kelebihan hormon androgen sehingga mengalami pertumbuhan rambut berlebih pada wajah atau tubuh serta kesulitan untuh hamil [2]. Berdasarkan data WHO (World Health Organization), revalensi PCOS pada perempuan usia reproduktif berkisar antara 6-13%. Namun demikian, sekitar 70% kasus POS belum terdeteksi [2]. Deteksi awal pada faktor risiko PCOS sangat diperlukan agar dapat menghindari gangguan kesehatan berjangka panjang, seperti diabetes, hipertensi dan infertilitas. Saat ini, untuk mendeteksi dan mendiagnosis PCOS diperlukan tes laboratorium dan pemeriksaan medis lainnya dan membutuhkan alokasi waktu dan anggaran. Oleh karena itu, diperlukan pendekatan berbasis AI agar dapat meningkatkan keakuratan dalam mendeteksi gangguan PCOS lebih awal.

Berbagai penelitian telah dilakukan dengan memanfaatkan pendekatan berbasis AI untuk mendeteksi faktor risiko PCOS dan mengklasifikasikan pasien PCOS dan Non-PCOS. Penelitian oleh B. Yamini et. al mengungkapkan bahwa model *Support Vector Machine* (SVM) adalah metode kecerdasan buatan yang sering digunakan dalam deteksi risiko PCOS. SVM termasuk teknik *machine learning* yang banyak diterapkan pada klasifikasi biner, seperti pemisahan data PCOS dan Non-PCOS. Metode tersebut menentukan *hyperplane* optimal untuk mem*filter* data untuk dua kelas. [3]. Adapun penelitian yang dilakukan oleh Ejay Nsugbe menunjukkan bahwa metode SVM menyediakan sistem pendukung keputusan diagnosis karena mampu mengidentifikasi kondisi pasien PCOS menggunakan trik kernel yang digunakan dalam menangani data yang tumpang tindih [4]. Disisi lain, penelitian yang dilakukan oleh Zainul Arifin et. al juga mengatakan bahwa model SVM menggunakan kernel RBF efektif dalam klasifikasi sel kanker jinak dan ganas. Meskipun yang diklasifikasikan adalah sel kanker, namun berdasarkan penelitian tersebut membuktikan bahwa model SVM dapat memisahkan dua kelas data dengan tingkat akurasi yang tinggi dengan menggunakan kernel RBF [5]. Berdasarkan kajian yang dilakukan dari penelitian sebelumnya, masih terdapat kekurangan pada penelitian tersebut, seperti penggunaan dataset dengan informasi dan atribut data yang kurang menyebabkan hasil pengevaluasian model cenderung kurang maksimal.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model SVM dengan kernel RBF untuk

mendeteksi dan mengklasifikasikan pasien dengan risiko PCOS dan Non-PCOS serta mendeteksi faktor-faktor yang paling berisiko dalam menyebabkan PCOS pada wanita. Penelitian ini juga menerapkan beberapa teknik, seperti teknik SMOTE untuk mengatasi ketidakseimbangan distribusi pada data, teknik GridSearch untuk pengoptimalan *hyperparameter* serta mengevaluasi performa model dengan menggunakan empat metrik evaluasi, yaitu akurasi, presisi, *recall*, dan *F1-Score*. Dari penelitian, diharapkan dapat menjadi cara alternatif untuk mendeteksi pasien dengan risiko PCOS dan Non-PCOS serta faktor yang paling berpengaruh yang dapat menyebabkan risiko PCOS agar dapat menghembat biaya pemeriksaan, terutama didaerah yang masih kurang dalam fasilitas medis.

II. KAJIAN TEORI

A. Polycystic Ovarian Syndrome (PCOS)

Polycystic Ovarian Syndrome (PCOS) diklasifikasikan sebaai masalah hormonal yang menyerang sistem reproduksi perempuan dan menyerang diusia produktif. PCOS ditandai dengan gangguan siklus mentruasi yang tidak teratur, ketidakseimbangan hormon, dan kemungkinan adanya pembentukan kista pada ovarium [6]. Hingga saat ini, etiologi PCOS belum sepenuhnya dipahami, tetapi, ada beberapa penyebab dari PCOS. Diantara penyebab adalah ketidakseimbangan hormon, faktor genetik, gaya hidup yang tidak sehat, pola makan yang tidak teratur dan memiliki riwayat diabetes dan obesitas.

B. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode *machine learning* untuk tugas regresi dan klasifikasi. SVM bekerja dnegan menentukan *hyperplane* sebagai batas pemisah antar kelas biner pada data. Model ini bertujuan untuk memaksimalkan margin antar kelas, di mana margin yang lebih besar menandakan performa model yang baik dalam melakukan generalisasi pada data baru [7].

C. Kernel RBF

Kernel RBF merupakan salah satu fungsi yang digunakan sebagai pendekatan model SVM yang berperan memetakan data dari ruang berdimensi rendah ke ruang berdimensi tinggi [5]. Kernel RBF diigunakan untuk menangani data non-linear dengan persamaan sebagai berikut:

$$K(x,y) = \exp\left(-\gamma \sum_{j=1}^{p}(x_{i,j} - y_{i,j})^2\right) \quad (1)$$

K : Nilai kernel
x,y : Dua vektor yang mewakili dua data
$x_{i,j}$, $y_{i,j}$: Elemen ke-j dari vektor x dan y. i merupakan data ke-i.
γ : Parameter yang mengatur lebar dari fungsi kernel.
exp : Fungsi Gaussian

Parameter γ pada persamaan kernel diatas berperan dalam menentukan keputusan dan mengendalikan varian model [8]. Apabila parameter γ bernilai kecil, maka pengaruh terhadap titik data akan lebih besar dan model akan lebih halus. Sebaliknya, apabila parameter γ bernilai besar maka model akan lebih kompleks dan sensitif pada data latih [5].

D. SMOTE

Synthetic Minority Over-sampling Technique atau teknik SMOTE merupakan strategi yang dimanfaatkan guna menangani ketidakseimbangan pada kelas data yang digunakan [9]. Melalui teknik ini, data akan ditambahkan menggunakan data sintetis ke dalam kelas data yang minoritas agar data yang akan diolah oleh model menjadi seimbang. Dengan menggunakan teknik SMOTE ini juga dapat memberikan hasil evaluasi model yang tidak bias dan kemungkinan lebih dominan ke kelas data mayoritas.

E. GridSearch

Teknik *GridSearch* digunakan untuk mengoptimalkan *hyperparameter* dengan melibatkan rentang nilai yang mungkin untuk setiap *hyperparameter* dan membuat grid dari kemungkinan kombinasi *hyperparameter* serta menguji semua kombinasi validasi silang agar dapat mencegah terjadinya *overfitting* ketika mengevaluasi kinerja model pada data yang tidak digunakan dalam pelatihan model [5].

F. Confussion Matrix

Confussion Matrix merupakan matriks evaluasi kinerja model klasifikasi berdasarkan data uji. Dalam *confussion matrix* terdapat empat bagian penting yang bertujuan guna mengevaluasi hasil prediksi dari model [10]. Empat bagian penting tersebut dapat divisualisasikan sebagai berikut:

TABEL 1
Tabel Confussion Matrix

TP	FP
FN	TN

TP: model memprediksi positif yang sebenarnya positif, yang artinya prediksi benar
FP: model mempreksi positif yang sebenarnya negatif, yang artinya model salah menilai data menjadi positif
FN: model memprediksi negatif yang sebenarnya positif, yang artinya model salah menilai data menjadi positif
TN: model memprediksi negatif yang sebenarnya negatif, yang artinya prediksi model bernilai negatif.

G. Akurasi, Presisi, *Recall*, dan *F1-Score*

Akurasi merupakan metrik evaluasi yang digunakan untuk mengukur tingkat ketepatan model dalam menghasilkan prediksi yang benar. Metrik ini menunjukkan kemampuan model dalam mengklasifikasikan data secara tepat, namun tidak dapat dijadikan satu-satunya acuan dalam mengevaluasi kinerja model secara keseluruhan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Presisi merupakan metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam menghasilkan prediksi yang benar pada kelas positif dari keseluruhan data. Dalam konteks klasifikasi, presisi menunjukkan sejauh mana model mampu mengidentifikasi data positif

tanpa melakukan kesalahan klasifikasi terhadap data yang seharusnya bukan termasuk kelas positif.

$$Precision = \frac{TP}{TP+F} \qquad (3)$$

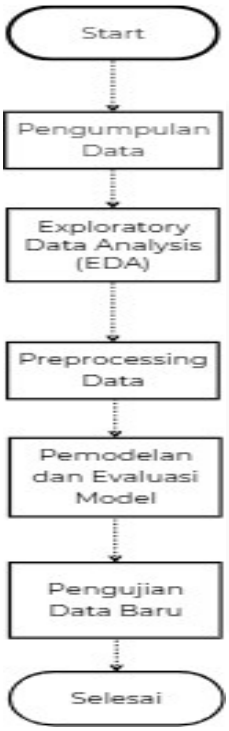
Recall merupakan metrik evaluasi yang mengindikasikan kemampuan model dalam mengenali kelas data secara benar. Nilai *recall* akan menunjukkan seberapa jauh model mampu mengklasifikasikan data dengan benar.

$$recall = \frac{TP}{TP+FN} \qquad (4)$$

F1-Score merupakan metrik evaluasi untuk menggambarkan keseimbangan antara nilai presisi dan *recall*. Nilai *F1-Score* dapat menunjukkan bagaimana model dalam menggabungkan kemampuan presisi dan *recall* untuk memahami efektifitas model dalam mengklasifikasikan data dengan benar [10].

$$F1\ Score = 2 \times \frac{Recall \times Precission}{Recall + Precission} \qquad (5)$$

III. METODE



GAMBAR 1
Flowchart Penelitian

A. Pengumpulan Data

Dataset pada penelitian ini menggunakan dataset PCOS yang diambil dari sumber Kaggle. Dataset ini berisi informasi sebanyak 1000 baris dengan 5 atribut data faktor risiko PCOS. Dataset ini digunakan dengan tujuan fokus utama penelitian, yaitu untuk mengetahui kinerja model SVM dalam mengklasifikasi dan mengidentifikasi faktor risiko yang paling berpengaruh terhadap PCOS.

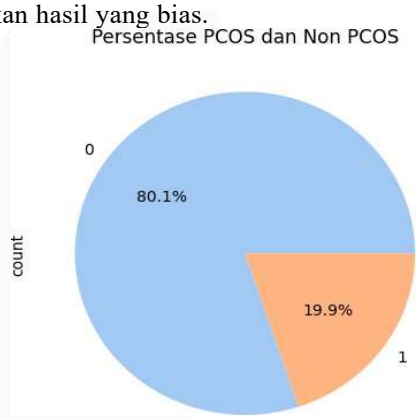
Age	BMI	Menstrual_Irregularity	Testosterone_Level(ng/dL)	Antral_Follicle_Count	PCOS_Diagnosis
24	34.7	1	25.2	20	0
37	26.4	0	57.1	25	0
32	23.6	0	92.7	28	0
28	28.8	0	63.1	26	0
25	22.1	1	59.8	8	0
...	...	...	...	...	...
34	18.4	1	95.7	23	0
45	28.9	1	28.5	7	0
37	28.3	0	32.4	28	0
41	27.3	0	95.6	9	0
22	21.9	1	78.9	7	0

1000 rows × 5 columns

GAMBAR 2
Tabel Data Yang Digunakan

B. Exploratory Data

Exploratory Data Analysis (EDA) adalah proses dalam memahami karakteristik dan mengidentifikasi data sebelum dilakukan pemodelan. Pada penelitian ini, EDA dilakukan dengan mengvisualisasikan data pasien yang berisiko PCOS dan Non-PCOS serta umur pasien. Kemudian dilakukan pemeriksaan terhadap data untuk memastikan apakah terdapat atribut yang memiliki *missing* value dan data yang terduplikasi. Selain itu, dilakukan juga pemeriksaan terhadap data apakah data mengalami ketidakseimbangan atau tidak. Dari pemeriksaan tersebut, diketahui persentase data PCOS lebih dominan 80% dibandingkan dengan data Non-PCOS sehingga diperlukan penanganan terhadap data yang tidak seimbang agar model yang dibangun tidak memberikan hasil yang bias.



GAMBAR 3
Persentase Data PCOS dan Non-PCOS

C. Preprocessing Data

Preprocessing data dilakukan untuk mempersiapkan data agar mampu menghasilkan analisis yang lebih optimal dan akurat. Pada tahapan ini diawali dengan pembuatan variabel X dan Y. Variabel X merupakan atribut data "PCOS_Diagnosis" dan variabel Y merupakan label dari variabel X yang direpresentasikan dengan angka 0 untuk data Non-PCOS dan angka 1 untuk data PCOS. Kemudian, dataset dibagi ke dalam data pelatihan dan data pengujian dengan pembagian data sebesar 20% untuk data uji dan 80% untuk data latih. Kemudian, dilakukan standarisasi data dengan memisahkan variabel X dan Y agar dapat memastikan semua atribut memiliki skala yang sama. Kemudian, menentukan nilai mean dan standar deviasi dari data latih variabel X dan menggunakannya untuk menormalisasi data latih variabel X serta data uji variabel X sehingga hasil penghitungan tersebut siap digunakan untuk pemodelan.

Karena data yang digunakan tidak seimbang, maka dilakukan penanganan terhadap ketidakseimbangan data yang diatasi melalui teknik SMOTE. Penggunaan teknik ini hanya diterapkan hanya pada data latih dan bukan pada data uji agar dapat memastikan hasil nilai akurasi, presisi, *recall*, dan *F1-Score* menggambarkan kemampuan model terhadap data baru. Tahap terakhir pada proses preprocessing setelah melakukan penanganan data yang tidak seimbang adalah dengan melakukan pengoptimalan model SVM melalui *Hyperparameter Tuning*. Proses ini menggunakan teknik GridSearchCV dengan skema *5-fold Cross Validation* untuk pencarian sistematis dari kombinasi nilai dua parameter grid utama, yaitu parameter C dan parameter Gamma yang kemudian performa kombinasi kedua parameter tersebut akan dievaluasi dan memperoleh kombinasi *hyperparameter* terbaik

D. Pemodelan dan Evaluasi Model

Pemodelan merupakan proses inti dari penelitian ini. Model SVM pada penelitian ini dibangun dengan menggunakan fungsi kernel radial. Agar mendapatkan performa model yang akurat, digunakan metode *GridSearchCV* dan validasi silang. Setelah melakukan pemodelan, model dievaluasi dengan menggunakan akurasi, presisi, *recall*, dan *F1- Score* dan menggunakan grafik *Confusion Matrix*. Evaluasi model ini dilakukan dengan tujuan untuk mengukur bagaimana model dapat mengklasifikasikan pasien PCOS dan Non-PCOS secara akurat.

E. Pengujian Model pada Data Baru

Setelah membangun dan melakukan evaluasi model SVM, model akan diuji dengan menggunakan data baru. Data baru yang digunakan merujuk pada data asli namun dengan nilai yang ditentukan sendiri. Pengujian ini bertujuan untuk menunjukkan bahwa model SVM tidak hanya akurat dalam mendiagnosis PCOS pada data latih tetapi juga pada data baru.

IV. HASIL DAN PEMBAHASAN

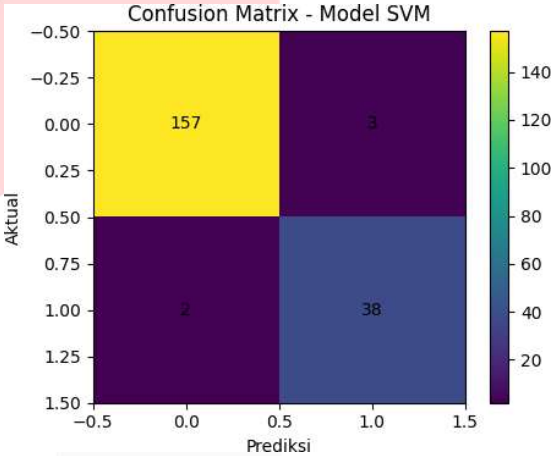
Pengolahan dataset dilakukan menggunakan Google Colab dan bahasa pemrograman Python melalui tahapan pengumpulan data, EDA, preprocessing data, penanganan terhadap ketidakseimbangan data, pemodelan dan evaluasi model, serta perngujian terhadap data baru.

A. Hasil Kinerja Model SVM

TABEL 2
Tabel *Classification Report*

	Akurasi	Presisi	Recall	F1-Score
Non PCOS (0)	97%	99%	98%	98%
PCOS (1)	97%	93%	95%	94%

Setelah dilakukan pengolahan data pada dataset yang digunakan, berdasarkan hasil pengolahan dataset pada Tabel 1, diketahui bahwa kinerja model SVM menunjukkan kinerja yang sangat baik karena dapat mengklasifikasikan data uji sebesar 97% dengan benar. Untuk kelas Non-PCOS, dihasilkan nilai presisi 99% yang berarti dari seluruh data yang diprediksi sebagai kelas PCOS benar sebagai kelas PCOS. Selain itu, terdapat nilai *recall* untuk kelas Non-PCOS sebesar 98%. Dari nilai tersebut model berhasil mendeteksi data PCOS dengan benar. Disisi lain, nilai *F1-Score* menunjukkan keseimbanan yang baik antara nilai presisi dan *recall* dengan nilai sebesar 98%. Adapun untuk kelas PCOS, terdapat nilai presisi sebesar 93% yang diprediksi benar sebagai PCOS dari keseluruhan data. Nilai *recall* sebesar 95% menunjukkan bahwa pasien berhasil terdeteksi sebagai PCOS dan hanya sedikit data yang terlewat untuk terdeteksi. Performa model SVM juga cukup baik dalam mengklasifikasikan kelas PCOS yang dapat dilihat dari nilai *F1-Score* sebesar 94%. terdeteksi. Performa model SVM juga cukup baik dalam mengklasifikasikan kelas PCOS yang dapat dilihat dari nilai *F1-Score* sebesar 94%.



GAMBAR 4
Grafik Confusion Matrix

Berdasarkan Gambar 4, diketahui hasil pengujian terhadap kinerja model SVM dalam klasifikasi dan mendeteksi PCOS pada wanita. Dengan menggunakan grafik *Confusion Matrix*, untuk kelas PCOS diketahui bahwa model SVM dapat mengklasifikasikan nilai *True Positve* sebesar 38. Artinya, terdapat 38 pasien yang diprediksi dan terdeteksi benar sebagai pasien PCOS. Kemudian, untuk *False Positive* menghasilkan nilai 3 yang berarti terdapat 3 data Non-PCOS yang salah diprediksi oleh model sebagai PCOS. Untuk nilai *False Negative* menghasilkan 2 data PCOS yang gagal terdeteksi. Adapun untuk kelas Non-PCOS, menghasilkan nilai *True Negative* sebesar 157. Artinya, terdapat 157 data Non-PCOS yang berhasil dikenali oleh model sebagai pasien Non-PCOS.

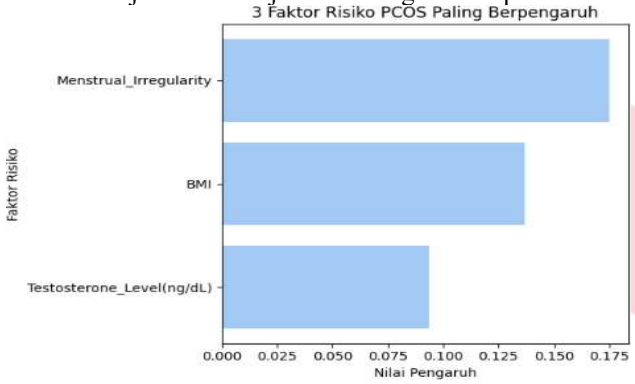
B. Faktor Risiko PCOS

TABEL 3
Tabel Faktor Paling Berpengaruh Terhadap Risiko PCOS

Faktor	Pengaruh
Menstrual Irregularity	0.175000
BMI	0.136667
Testosterone Level (ng/dL)	0.093333

Berdasarkan Tabel 2, diperoleh tiga faktor utama yang paling berpengaruh terhadap risiko PCOS pada wanita, yaitu Menstrual Irregularity, Body Mass Index (BMI), dan Testosterone Level, yang diidentifikasi menggunakan metode *Permutation Feature Importance*. Metode ini

mengukur penurunan performa model ketika nilai suatu atribut diacak, sehingga semakin besar importance score maka semakin besar pengaruh atribut tersebut terhadap keputusan model. Hasil menunjukkan bahwa ketidakteraturan siklus menstruasi memiliki pengaruh terbesar (0.175000), diikuti oleh BMI (0.136667) yang berkaitan dengan obesitas dan ketidakseimbangan hormon, serta kadar hormon testosteron (0.093333) yang berhubungan dengan gejala hiperandrogenisme pada PCOS. Meskipun demikian, hasil analisis ini menunjukkan tingkat pengaruh atribut terhadap model klasifikasi dan tidak dapat dijadikan sebagai kesimpulan medis, sehingga tetap diperlukan validasi klinis dan kajian lebih lanjut oleh tenaga medis profesional.



GAMBAR 5
Diagram Faktor Risiko PCOS Paling Berpengaruh

V. KESIMPULAN

Penelitian ini berhasil mengembangkan model *Support Vector Machine* berbasis kernel RBF untuk mendeteksi faktor yang paling berpengaruh menyebabkan risiko PCOS dan mengklasifikasikan kelas PCOS dan Non-PCOS dengan tingkat akurasi yang tinggi. Berdasarkan hasil evaluasi menyatakan bahwa SVM menunjukkan kinerja yang baik dengan tingkat akurasi sebesar 97% dalam pengklasifikasian pasien PCOS dan Non-PCOS secara akurat setelah melakukan penanganan terhadap data yang tidak seimbang. Pada kelas PCOS, model menunjukkan nilai presisi, *recall*, dan *F1-Score* yang tinggi dan berhasil memprediksikan PCOS dengan benar meskipun masih ada sejumlah kecil kasus PCOS yang belum terdeteksi sepenuhnya oleh model. Sedangkan untuk kelas Non-PCOS, model juga memberikan hasil nilai presisi, *recall*, dan *F1-Score* yang sangat tinggi dan mengindikasikan bahwa model dapat mendeteksi pasien Non-PCOS dengan baik. Sehingga secara keseluruhan, model SVM dapat membaca data pasien Non-PCOS dengan baik meskipun model masih ada kesalahan dalam mendeteksi seluruh pasien PCOS secara optimal. Disisi lain, penelitian ini juga dapat mengidentifikasi faktor-faktor yang paling berpengaruh menyebabkan risiko PCOS berdasarkan tingkat pengaruh fitur data yang digunakan. Berdasarkan hasil analisis yang dilakukan, faktor *Menstrual Irregularity*, BMI, dan *Testosterone Level (ng/dL)* yang menjadi faktor paling berpengaruh terhadap risiko PCOS. Namun, perlu diketahui bahwa faktor-faktor tersebut merupakan hasil deteksi dari model dan tidak bisa dijadikan sebagai kesimpulan medis sehingga perlu kajian lebih lanjut oleh tenaga ahli kesehatan profesional. Meski begitu, penelitian ini menunjukkan bahwa model SVM mampu mengklasifikasikan pasien PCOS dan Non-PCOS serta mendeteksi faktor-faktor yang berisiko PCOS sehingga model dapat digunakan sebagai alat pendukung dalam pendeteksian dini PCOS pada wanita.

REFERENSI

[1] Kodipalli, Ashwini., and Devi, Susheela, “Prediction of PCOS and Mental Health Using Fuzzy Inference and SVM,” Frontiers in Publich Health., Vol. 9, November, 2021

[2] (2025, 7 February), “Polycystic Ovary Syndrome”, WHO, [Online]. Available: https://www.who.int/news-room/factsheets/detail/polycystic-ovary-syndrome [Accessed: 5 November 2025]

[3] Yamin, B., Kaneti, Venkata Ramana., P, Prema., C, Ambhika., Nalini, M., and R, Siva Subramanian, “Machine Learning-Driven PCOS Prediction for Early Detection and Tailored Interventions,” SSRG International JournalI of Electrical and Electronics Engineering, vol. 10(9) pp 61-75. September 2023

[4] Nsugbe, Ejay, “An artificial intelligence-based decision support system for early diagnosis of polycystic ovaries syndrome,” Healthcare Analytics, vol.3 pp 2772-4425. Mar. 2023

[5] Arifin, Zainul., Rahman, Dimas Fachri., Rintyarna, Bagus Setya., and Daryanto, “Penerapan Algoritma Support Vector Machine Berbasis Kernel Radial Function dalam Klasifikasi Sel Kanker,” BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer. vol. 4 no.2 pp. 100-106. Sept. 2023

[6] (2024, 14 Oktober). “Memahami PCOS: Gejala, Penyebab, dan Perawatan yang Efektif”, MorulaIVF, [Online],Available: https://www.morulaivf.co.id/id/blog/pcosadalah/#:~:text=Memahami%20PCOS:%20Gejala%2C%20Penyebab%2C%20dan%20Perawatan%20yang%20Efektif&text=Artikel%20ini%20telah%20direview%20secara,%2C%20diabetes%2C%20dan%20%5B%E2%80%A6%5D&text=Ardianto%20Airlangga%2C%20SpOG,Polycystic%20Ovary%20Syndrome%20(PCOS)%20adalah%20gangguan%20hormonal%20yang%20umum%20terjadi,%2C%20diabetes%2C%20dan%20penyakit%20jantung , [Accessed: 5 November 2025]

[7] (2025, 13 November), “Support Vector Machine (SVM) Algorithm,” Geeksforgeeks, [Online], Available: https://www.geeksforgeeks.org/machine-learning/support-vector-machine-algorithm/ , [Accessed: 8 November 2025]

[8] Tiwari, Shamik., Kane, Lalit., Koundal, Deepika., Jain, Anurag., Alhudhaif, Adi., Polat, Kemal., Zaguia, Atef., Alenezi, Fayadh., and Althubiti, Sara A, “SPOSDS: A smart Polycystic Ovary Syndrome diagnostic system using machine learning,” Expert Systems With Applications. vol. 203, May, 2022

[9] Priani, Wina., Sulistianingsih, Evy., and Imro’ah, Nurfitri, “Implementasi Teknik *Oversampling SMOTE-NC* Pada Algoritma *K- Nearest Neighbor,*” Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster). vol. 14 no. 3. pp 272-281, 2025

[10] Rina, (2023, 11 Juni), “ Memahami Confussion Matrix: Accuracy, Precision, Recall, Specificity, dan F1-Score untuk Evaluasi Model Klasifikasi.” [Online], Available: https://esairina.medium.com/memahami-confusion-matrix-accuracy-precision-recall-specificity-

dan-f1-score-610d4f0db7cf , [Accessed: 8 November 2025]
[11] Akanbi, Kemi., Adepoju, Odunayo Gabriel., and Nti, Kofi Isaac. "Developing A System for Automatic Prediction of Polycystic Ovary Syndrome Using Machine Learning," In 2024 The 7th International Conference on Machine Learning and Machine Intelligence (MLMI) (MLMI 2024). pp 20-26. Aug, 2024
[12] Mridul, Aunik Hasan., Ahsan, Nowreen., Alam, Syeda Sadia., Afrose, Sonia., Sultana, Zakia., and kafi, Md. Tanvir Mahmud, "Polycystic Ovary Syndrome (PCOS) Disease Prediction Using Traditional Machine Learning and Deep Learning Algorithms," International Journal of Computer Information Systems and Industrial Management Applications, vol. 16 pp 322-346. July, 2024
[13] Silambarasan, E., Nirmala, G., and Mishara, Ishani, "Polycystic ovary syndrome detection using optimized and DenseNet," International Journal of Information Technology. vol. 17 no. 2 pp 1039-1047, Mar. 2025
[14] Thakre, Vaidehi., Vedpathak, Shreyas., Thakre, Kalpana., and Sonawani, Shilpa, "PCOcare: Detection and Prediction using Machine Learning Algorithms," Bioscience Biotechnology Research Communications, vol. 13 no. 14 pp 240-244, Dec. 2020
[15] Marshall, M.D. Ph.D., John C., and Dunaif, M.D, Andrea, "Should all woman with PCOS be treated for insulin resistance?," Fertility and Sterility, vol. 97 no.1 pp. 18-22, 2012
[17] Roe, MD, Andrea Hsu., and Dokras, MD, PhD, Anuja' "The Diagnosis of Polycystic Ovary Syndrome in Adolescents," Reviews in Obstetrics & Gynecology, vol. 4 no.2 pp. 45-51, 2011
[18] Ade Aulia. (2023, 23 Agustus). *Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Diabetes | Machine Learning Project 3*. [Video]. YouTube. https://youtu.be/kRTrwNWGsiA?si=BAPTRFNhwCYeHJXu