

ANALISIS POLA GEJALA PCOS MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Cantika Aulia¹⁾, M Tsabita Robani²⁾, Muhammad Fakhri Nadrota Acta³⁾, Khalid Al Mas Ud⁴⁾, Marco Saputra⁵⁾, Ken Ditha Tania⁶⁾, Winda Kurnia Sari⁷⁾

¹⁻⁷ Fakultas Ilmu Komputer, Sistem Informasi, Universitas Sriwijaya, Indralaya, Indonesia
Jl. Raya Palembang - Prabumulih Km. 32, Indralaya, Ogan Ilir, Sumatera Selatan, Indonesia
Email: ¹09031282227124@student.unsri.ac.id, ²09031382227140@student.unsri.ac.id,
³0903182227048@student.unsri.ac.id, ⁴09031182227027@student.unsri.ac.id,
⁵0903182227030@student.unsri.ac.id, ⁶kenya.tania@gmail.com, ⁷windakurniasari@gmail.com

Abstract

This research implements the K-Means clustering algorithm to group patients with Polycystic Ovary Syndrome (PCOS) based on symptom severity. PCOS is a complex hormonal disorder characterized by a wide range of symptoms, making diagnosis challenging. Through K-Means clustering, three distinct groups were identified: low similarity, medium similarity, and high similarity,. Each cluster represents specific symptom combinations, including irregular menstrual cycles, excessive hair growth, and mood swings. This method offers valuable insights into the distribution patterns of PCOS symptoms and provides a data-driven basis to support more informed medical decision-making.

Keywords: PCOS, Clustering, K-Means, Diagnosis, Data Mining

Abstrak

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan pasien PCOS berdasarkan intensitas gejala. PCOS adalah gangguan hormonal yang sulit didiagnosa karena gejalanya beragam. Dengan menggunakan algoritma K-Means, hasil *clustering* menunjukkan tiga kategori utama: ringan, sedang, dan berat. Setiap klaster mencerminkan kombinasi gejala seperti siklus menstruasi tidak teratur, pertumbuhan rambut berlebih, dan perubahan suasana hati. Pendekatan ini efektif dalam membantu pemahaman pola distribusi gejala PCOS serta mendukung pengambilan keputusan medis yang lebih tepat.

Kata kunci: PCOS, Clustering, K-Means, Diagnosis, Data Mining

1. PENDAHULUAN

Pemanfaatan algoritma dalam bidang medis terus berkembang, terutama untuk mendiagnosis penyakit berdasarkan analisis data. Salah satu penyakit yang kompleks dan sulit didiagnosis adalah Polycystic Ovary Syndrome (PCOS), gangguan hormonal yang mempengaruhi sekitar 13% wanita usia subur di dunia (Kurniawati et al., 2023). PCOS berdampak signifikan pada sistem reproduksi dan metabolisme perempuan, sehingga memengaruhi kualitas hidup penderitanya. Gejala umum yang muncul pada remaja meliputi jerawat (acne), siklus menstruasi tidak teratur (oligomenorea), dan perubahan warna kulit seperti acanthosis nigricans (Sari et al., 2023).

PCOS ditandai oleh ketidakseimbangan hormon, gangguan ovulasi, dan peningkatan hormon androgen yang menyebabkan munculnya kista kecil pada ovarium (Tanjung & Achmad Fauzi, 2023). Selain menjadi penyebab utama infertilitas, PCOS juga berkaitan dengan kondisi seperti kelebihan berat badan, stres, dan depresi. Gejala fisik seperti hirsutisme, jerawat, serta gangguan berat badan berdampak pada kepercayaan diri dan kesehatan mental penderita. Sayangnya, sekitar 70% wanita yang mengalaminya belum mendapatkan diagnosis resmi (WHO, 2025) (WHO, 2025), menunjukkan perlunya pendekatan deteksi dini yang lebih efektif. Tercatat remaja putri yang terkena PCOS sering kali memiliki tingkat pengetahuan yang kurang mengenai kondisi tersebut (Oktavia et al., 2024). Sementara itu, berdasarkan kriteria National Institutes of Health, (2024), PCOS mempengaruhi sekitar 4-8% wanita di usia subur dan ditandai dengan gejala seperti oligomenore, hiperandrogenisme, serta ovarium polikistik (Karadayı Atas, 2024). Oleh karena itu, diperlukan pendekatan untuk mendeteksi

penyakit ini dengan menerapkan algoritma clustering yang mampu mengelompokkan data secara lebih terstruktur.

Penelitian terkini berfokus pada penggunaan berbagai algoritma dalam Machine Learning untuk mendiagnosis PCOS menggunakan data terstruktur atau tidak terstruktur (Arora et al., 2024). Salah satu pendekatan yang dapat digunakan adalah algoritma clustering, yang mampu mengelompokkan data berdasarkan karakteristik tertentu. Clustering memungkinkan identifikasi pola dalam dataset medis, termasuk analisis hormon sebagai indikator utama penyakit (Kholil et al., 2024). Dalam konteks diagnosis PCOS, teknik ini membantu memetakan karakteristik pasien berdasarkan variasi gejala dan kadar hormon, sehingga mendukung proses pengambilan keputusan medis yang lebih akurat.

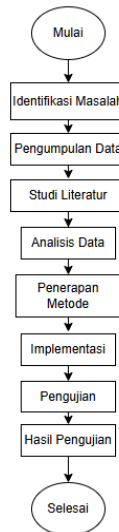
Metode K-Means merupakan algoritma clustering yang sering digunakan dalam segmentasi data medis, termasuk dalam identifikasi folikel ovarium melalui citra ultrasonografi (Sugianto et al., 2020). Penggunaan metode ini memungkinkan analisis data dalam skala besar secara efisien. Penelitian terdahulu juga menunjukkan efektivitas kombinasi K-Means dengan algoritma lain. Misalnya, Sheikdavood & Bala (2023) menggabungkan Adaptive K-Means dengan Reptile Search Algorithm (RSA), serta CNN dan DNN untuk meningkatkan akurasi diagnosis. N. N. Faris & Miften, (2022) menggabungkan K-Means untuk seleksi fitur dan LS-SVM untuk klasifikasi, mencapai akurasi hingga 99%.

Pendekatan berbasis genomik juga menunjukkan potensi besar dalam klasifikasi PCOS. Studi oleh Dapas & Dunaif (2022) mengidentifikasi sub tipe PCOS berdasarkan profil genetik dan fenotipik yang berbeda. Metode K-Means dapat diterapkan untuk mengelompokkan pasien berdasarkan kesamaan genetik, memungkinkan personalisasi diagnosis dan pengobatan. Di sisi lain, pendekatan berbasis citra menggunakan CNN, seperti yang dilakukan oleh (Nurdiawan et al., 2024), juga menunjukkan hasil akurat dalam mendeteksi PCOS, dengan akurasi mencapai 89,47%. Pendekatan lain yang dapat digunakan dalam klasifikasi ini adalah metode K-Means Clustering, yang dapat mengelompokkan data citra ovarium berdasarkan kesamaan fitur sehingga membantu proses segmentasi awal sebelum diterapkan pada model deep learning. Dengan kombinasi metode K-Means dan CNN, akurasi deteksi PCOS dapat ditingkatkan melalui pemisahan lebih optimal antara data normal dan yang terkena PCOS (Putri et al., 2022).

Penelitian ini bertujuan untuk menganalisis faktor risiko utama PCOS, termasuk gaya hidup dan berat badan berlebih yang berkontribusi terhadap peningkatan kasus pada remaja. Selain itu, penelitian ini mengeksplorasi penerapan algoritma K-Means Clustering dalam menganalisis dataset PCOS guna mengidentifikasi pola-pola gejala yang tidak terlihat melalui metode konvensional. Diharapkan hasil dari penelitian ini dapat meningkatkan akurasi diagnosis dan mendukung pengambilan keputusan medis yang lebih efektif melalui pemanfaatan teknologi analisis data.

2. METODE PENELITIAN

Tahapan penelitian merupakan langkah-langkah dari sebuah proses yang akan dilaksanakan sesuai dengan penelitian. Dengan tahapan penelitian maka sebuah alur proses akan tersusun secara baik. Sebuah alur proses pada tahapan penelitian dimulai dari identifikasi kemunculan masalah yang ada sampai dengan mendapatkan hasil dari pengujian. Pada gambar 1. dapat dilihat visualisasi berupa gambar alur penelitian, dimana pada gambar tersebut tersusun secara jelas setiap tahapan yang dijalani pada penelitian.



Gambar 1. Tahapan Penelitian

Adapun pendefinisian terkait dengan setiap tahapan pada penelitian dapat dilihat berikut:

2.1. PROSES IDENTIFIKASI MASALAH

Identifikasi masalah merupakan tahap awal yang krusial dalam suatu penelitian karena menentukan arah dan fokus kajian yang akan dilakukan. Menurut (Nasution, 2021), identifikasi masalah berfungsi untuk menjelaskan kepada pembaca mengapa suatu kajian penting dilakukan, sekaligus memperjelas alasan pemilihan topik oleh peneliti. Dalam konteks ini, Polycystic Ovary Syndrome (PCOS) menjadi isu yang signifikan karena tidak hanya berdampak pada kesuburan, tetapi juga berhubungan dengan komplikasi metabolik seperti resistensi insulin, obesitas, dan peningkatan risiko diabetes tipe 2 (Elmannai et al., 2023). Dengan mempertimbangkan tingginya prevalensi dan dampak jangka panjangnya, diperlukan pendekatan yang tepat untuk mendeteksi dan memahami pola gejala PCOS secara dini.

2.2. PROSES PENGUMPULAN DATA

Pada tahap ini dilakukan pengumpulan data. Data yang akan digunakan dalam proses penelitian berupa data sekunder mengambil dataset dari kaggle (Emanuel et al., 2024). Dataset yang digunakan untuk penelitian ini menyediakan data terkait PCOS berupa variabel gejala dan kadar hormon pasien (Koli et al., 2023). Penulis melakukan pemahaman terhadap objek yang ingin diteliti. Penulis membaca berbagai sumber referensi seperti, jurnal penelitian sebelumnya, buku - buku, dan website yang berkaitan dengan penelitian. Setelah dilakukannya pengumpulan data maka dilakukan analisa data yang telah didapatkan untuk menguji dan memvalidasi apakah data tersebut bisa digunakan dalam identifikasi pola PCOS pada pasien.

2.3. PENERAPAN METODE ALGORITMA K-MEANS

Pada tahap ini, dilakukan proses dekomposisi masalah dengan menerapkan Algoritma K-Means dalam memproses data untuk menanggapi perumusan masalah. K- Means Clustering merupakan suatu metode dalam analisis data dalam penggunaan pengelompokkan data ke dalam beberapa cluster atau kelompok berdasarkan persamaan karakteristik tertentu (Mantik & Hidayati, 2022). Algoritma K-Means Clustering juga bisa dikatakan sebagai metode pengelompokan data yang bekerja dengan mengelompokkan titik-titik data ke dalam beberapa klaster berdasarkan kedekatan jarak matematis, umumnya menggunakan metrik *Euclidean*. Proses ini bertujuan untuk meminimalkan jarak total antara setiap titik data dengan pusat klaster yang telah ditentukan yang diimplementasikan melalui aplikasi *Rapidminer* guna mendapatkan hasil dari algoritma K-Means . Titik data yang memiliki jarak paling dekat dengan centroid akan dikelompokkan ke dalam klaster yang sama. Parameter k dalam algoritma ini menunjukkan jumlah cluster yang akan dibentuk, di mana nilai k yang lebih tinggi menghasilkan klaster yang lebih kecil dengan tingkat granularitas yang lebih tinggi, sedangkan nilai

k yang lebih rendah menghasilkan kluster yang lebih luas dengan detail yang lebih sedikit (Wibowo & Mulyastuti, 2022).

Langkah- langkah untuk melakukan K-Means Clustering sebagai berikut : Proses klusterisasi dilakukan untuk mendapatkan hasil dari algoritma K -Means yang diimplementasikan melalui aplikasi Rapidminer.

1. Tentukan jumlah cluster k. Pilih k centroid secara acak dari data points sebagai inisialisasi awal.
2. Untuk setiap data point, hitung jarak ke setiap centroid dan tetapkan data point tersebut ke cluster dengan centroid terdekat. Jarak dihitung menggunakan jarak Euclidean:

$$d(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$$

Keterangan

$d(x, y)$ = Jarak data ke x ke pusat kluster y

y_i = titik pusat ke $-y$ pada observasi ke $-i$

x_i = data x pada observasi ke $-i$

n = banyaknya observasi

3. Menghitung titik pusat cluster yang baru dengan menghitung rata - rata jarak data dengan titik pusat cluster. Jika pusat cluster tidak berubah maka proses cluster sudah selesai tetapi jika pusat cluster masih berubah maka ulangi langkah menghitung jarak pada langkah ke dua sampai pusat cluster tidak berubah lagi.

2.4. IMPLEMENTASI

Pada tahap ini, implementasi merupakan suatu proses penerapan dari metode yang digunakan. Penerapan metodologi yang sudah dirancang untuk mendapatkan suatu hasil dari tujuan penelitian itu sendiri. Proses implementasi dilalui dengan prosedur yang sistematis dan terdokumentasi guna memastikan hasil yang diperoleh memiliki tingkat validitas yang tinggi.

2.5. PENGUJIAN DAN HASIL PENGUJIAN

Pada tahap ini dilakukan proses pengujian yang hasilnya akan dianalisa untuk memastikan dari hasil data tersebut sesuai dengan tujuan dari penelitian ini. Setelah dilakukannya pengujian, maka didapatkan hasil pengujian tersebut untuk dijadikan penyajian informasi yang valid dan kredibel melalui bukti bukti gambar, grafik atau data yang sudah didapatkan dan dimiliki dari penulisan.

3. HASIL DAN PEMBAHASAN

3.1. ANALISIS MASALAH

Polycystic Ovary Syndrome (PCOS) merupakan kondisi yang bervariasi dan memengaruhi sekitar 4%–21% wanita usia reproduksi. Kondisi ini bervariasi dan dapat menyebabkan ketidakseimbangan hormon, gangguan ovulasi, dan peningkatan risiko penyakit metabolik. PCOS memiliki gejala yang berbeda pada setiap orang, jadi sulit untuk mendiagnosisnya. Hingga kini, belum ada fenotipe spesifik yang dapat digunakan untuk mengelompokkan pasien berdasarkan karakteristik reproduktif dan metaboliknya (Mohamed et al., 2025). Penelitian lebih lanjut diperlukan untuk memahami berbagai manifestasi klinis PCOS dan metode pengobatan yang lebih baik.

3.2. HASIL PENGUMPULAN DATA

Data yang digunakan dalam penelitian ini mencakup karakteristik pasien PCOS berdasarkan usia, indeks massa tubuh (BMI), dan berbagai gejala. Data ini dikumpulkan untuk mencari pola umum pada pasien PCOS dan membantu dalam analisis kelompok. Memahami bagaimana karakteristik pasien tersebar dapat memberikan pemahaman yang lebih baik tentang variasi kondisi PCOS serta variabel yang mempengaruhi intensitas gejala. Hasil pengolahan data disajikan dalam tabel 1 dan tabel 2 berikut:

Tabel 1. Karakteristik rata-rata pasien PCOS

Gejala	Rata-rata	Satuan	Jumlah Pasien	% Pasien
Age	18 – 21	years	180	38,7

BMI	19 – 22	kg/m ²	117	25
Menstrual Cycle Frequency Period Duration	1 ±	day	390	83
	4 ±	day	144	31

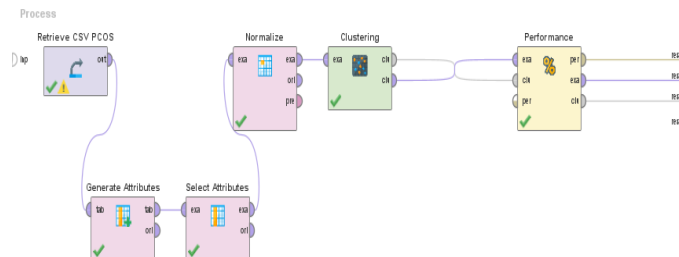
Tabel 2. Diagnosis pasien PCOS

Gejala	0 (No)	1 (Yes)	% Pasien
Regular Period	100	365	78,4
Weight Gain	232	233	50,
Excessive Hair Growth	342	123	26,4
Skin Darkening	308	157	33,7
Hair Loss	177	288	62
Acne	273	192	41
Mood Swing	111	354	76
PCOS Diagnosed	363	102	22

Hasil ini mengonfirmasi bahwa PCOS memiliki manifestasi gejala yang beragam, dan pendekatan berbasis data diperlukan untuk memahami pola karakteristik pasien serta memberikan intervensi yang lebih tepat.

3.3. IMPLEMENTASI ALGORITMA K-MEANS

Tujuan algoritma K-Means adalah untuk menempatkan pasien PCOS dalam kelompok berdasarkan gejalanya. Dalam studi N. Faris et al., (2025), algoritma K-Means juga digunakan sebagai bagian dari model seleksi fitur, di mana ia berperan penting dalam proses identifikasi atribut yang paling releva. Untuk memastikan pengolahan data yang optimal, proses ini dilakukan menggunakan *RapidMiner* yang hasilnya digunakan untuk memeriksa pola gejala PCOS dan variasi tingkat keparahannya. Ini mencakup pemrosesan data awal, normalisasi, pemilihan atribut, penerapan *clustering*, evaluasi dan implementasi. Proses implementasi K-Means dalam *RapidMiner*, dari pemrosesan data hingga evaluasi model clustering, digambarkan dalam Gambar 2.



Gambar 2. Implementasi K-Means pada RapidMiner

Tabel 3. Evaluasi Clustering menggunakan DBI

Cluster	DBI Value
2	1981
3	1922
4	1869
5	1792
6	1679
7	1776
8	1666

Berdasarkan Tabel 3.3. evaluasi clustering menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa nilai DBI menurun seiring bertambahnya jumlah cluster. Nilai DBI untuk cluster 6 dalam penelitian ini (1679), yang lebih rendah daripada nilai DBI untuk cluster lain, menunjukkan bahwa pemisahan antar cluster lebih baik dengan tingkat kesamaan yang tinggi. Oleh karena itu,

penggunaan 6 Cluster dianggap sebagai pilihan yang sesuai untuk mendapatkan hasil clustering yang lebih optimal dan representatif terhadap data yang digunakan.

3.3.1. ANALISIS HASIL CLUSTERING

Berdasarkan hasil clustering menggunakan algoritma K-Means, diperoleh tiga kluster utama yang mengelompokkan pasien berdasarkan tingkat intensitas gejala PCOS yang dapat dilihat pada Tabel 4 berikut:

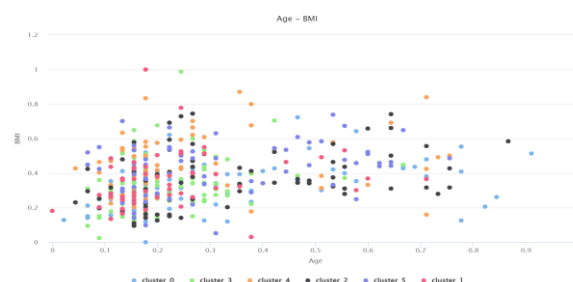
Tabel 4. Hasil Clustering

Gejala	C0	C1	C2	C3	C 4	C5
	84	54	73	106	71	77
Age	0.321	0.232	0.334	0.210	0.264	0.309
BMI	0.319	0.361	0.366	0.321	0.451	0.407
Menstrual Cycle Frequency	0.048	0.232	0.334	0.210	0.264	0.026
Period Duration	0.293	0.325	0.289	0.313	0.324	0.301
Regular Period	0.893	0.963	0.904	0.925	0	0.961
Weight Gain	0	0.778	0.178	0.415	0.803	1
Excessive Hair Growth	0.024	0.963	0.096	0.066	0.606	0.156
Skin Darkening	0.143	0.741	0.068	0.302	0.662	0.273
Hair Loss	0.476	0.741	0.342	0.774	0.761	0.610
Acne	0	0.704	0.137	1	0.535	0
Mood Swing	1	0.926	0	0.887	0.901	0.805
PCOS Diagnosis	0.071	0.315	0.068	0.075	0.873	0.052
Level Severity	High	Medium	Low	Medium	High	Low

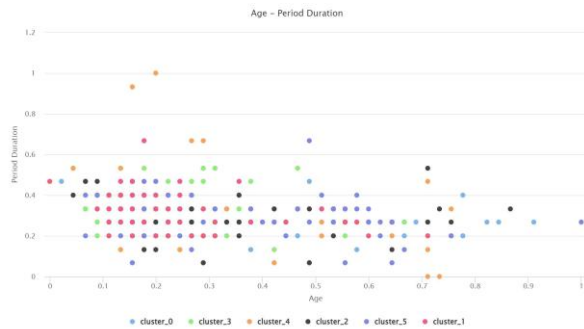
Hasil analisis kluster menunjukkan bahwa pasien dengan PCOS dapat diklasifikasikan ke dalam tiga kategori berdasarkan kemiripan pola gejala, yaitu *High Similarity* (C0 dan C4), *Medium Similarity* (C1 dan C3), serta *Low Similarity* (C2 dan C5). Kluster dengan tingkat kemiripan tinggi dicirikan oleh tingkat *mood swing* yang signifikan, siklus menstruasi yang relatif teratur, serta tingkat kerontokan rambut yang mencolok, dengan C4 juga menunjukkan tingkat diagnosis PCOS dan peningkatan berat badan yang tinggi, yang mengindikasikan kecenderungan terhadap sub tipe metabolik. Kluster dengan kemiripan sedang didominasi oleh manifestasi hiperandrogenisme, seperti hirsutisme, hiperpigmentasi kulit, serta gangguan siklus menstruasi. Sementara itu, kluster dengan kemiripan rendah menunjukkan gejala yang minimal, di mana C2 hampir tidak menampilkan indikasi klinis signifikan, sedangkan C5 menunjukkan *mood swing* yang cukup tinggi dengan gejala fisik lain yang relatif ringan. Temuan ini menekankan pentingnya pendekatan segmentatif dalam mengidentifikasi heterogenitas gejala PCOS, yang dapat mendukung proses diagnosis dan penanganan yang lebih tepat dan personal.

3.3.2. VISUALISASI DATA

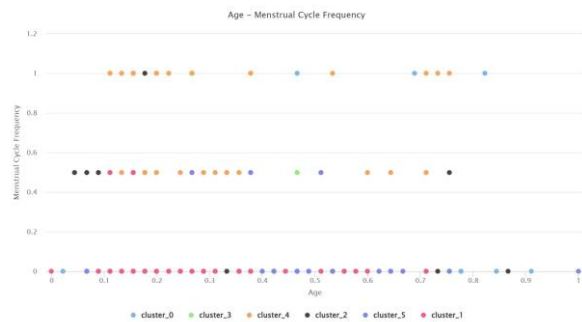
Untuk memahami pola variasi gejala PCOS pada pasien, dilakukan visualisasi data yang menggambarkan hubungan antara usia dengan beberapa parameter klinis yang acak. Visualisasi data menunjukkan hubungan antara usia pasien dengan durasi menstruasi, frekuensi siklus menstruasi, dan BMI.



Gambar 3. Scatter Age and BMI



Gambar 4. Scatter Age and Period Duration



Gambar 5. Scatter Age and Menstrual Cycle Frequency

PEMBAHASAN

Analisis kluster menggunakan algoritma K-Means berhasil mengelompokkan pasien PCOS ke dalam tiga kategori utama berdasarkan kemiripan pola gejala, yakni High, Medium, dan Low Similarity. Kluster *High Similarity* (C0 dan C4) menunjukkan gejala mencolok, seperti *mood swing*, siklus menstruasi teratur, dan kerontokan rambut. Khususnya, C4 juga memperlihatkan prevalensi diagnosis PCOS dan peningkatan berat badan yang tinggi, yang dapat dikaitkan dengan fenotipe metabolik. Temuan ini selaras dengan studi Hoeger et al., (2021), yang menyatakan bahwa pengelompokan fenotipik tanpa supervisi dapat memisahkan antara fenotipe reproduktif dan metabolik berdasarkan temuan genetik terkini.

Kluster *Medium Similarity* (C1 dan C3) didominasi oleh manifestasi hiperandrogenisme, seperti hirsutisme, hiperpigmentasi, dan gangguan menstruasi, yang mengindikasikan fenotipe reproduktif. Sementara itu, kluster *Low Similarity* (C2 dan C5) menunjukkan gejala minimal; C2 hampir asimtomatik, sedangkan C5 didominasi *mood swing* ringan. Hal ini mengindikasikan adanya subkelompok yang belum sepenuhnya memenuhi kriteria diagnostik PCOS, namun tetap berada dalam spektrum risiko.

Secara keseluruhan, pendekatan klustering berbasis data terbukti efektif dalam mengidentifikasi sub tipe klinis PCOS, mendukung strategi diagnosis dan intervensi yang lebih personal. Dengan demikian, pendekatan klustering berbasis gejala klinis menjadi alat bantu yang relevan dan efisien.

4. SIMPULAN

Penelitian ini menerapkan algoritma K-Means untuk mengelompokkan pasien PCOS berdasarkan intensitas gejalanya. Hasil clustering menunjukkan bahwa pasien dapat dikategorikan ke dalam tiga tingkat keparahan: ringan, sedang, dan berat. Cluster dengan keparahan tinggi ditandai oleh dominasi faktor metabolik seperti kenaikan berat badan dan resistensi insulin, sedangkan cluster dengan keparahan sedang menunjukkan gejala hiperandrogenisme seperti pertumbuhan rambut berlebih dan kerontokan rambut. Cluster dengan keparahan rendah memiliki gejala minimal dan lebih didominasi gangguan emosional seperti perubahan suasana hati.

Metode K-Means terbukti efektif dalam mengelompokkan pasien PCOS berdasarkan karakteristik pola gejala yang dialami pasien, membantu dalam pemahaman pola distribusi penyakit, serta mendukung pendekatan diagnosis yang lebih spesifik. Pendekatan ini berpotensi mendukung diagnosis yang lebih terpersonalisasi, pemantauan kondisi yang lebih akurat, serta intervensi terapeutik yang selaras dengan profil klinis pasien. Dalam jangka panjang, pendekatan ini dapat meningkatkan efisiensi layanan kesehatan, mempercepat deteksi komplikasi, dan meminimalkan pendekatan empiris dalam terapi. Implementasi metode ini membuka peluang pengembangan sistem pendukung keputusan klinis berbasis kecerdasan buatan guna memperkuat praktik kedokteran berbasis bukti.

Pengembangan selanjutnya disarankan mencakup komparasi dengan algoritma lain seperti DBSCAN atau Hierarchical Clustering, penambahan variabel klinis yang lebih kompleks, serta

perluasan cakupan dataset. Integrasi pendekatan ini dengan teknologi AI diperkirakan dapat meningkatkan akurasi diagnosis serta efektivitas tata laksana PCOS secara menyeluruh.

5. DAFTAR PUSTAKA

- Arora, S., Vedpal, & Chauhan, N. (2024). Polycystic Ovary Syndrome (PCOS) diagnostic methods in machine learning: a systematic literature review. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-024-19707-6>
- Dapas, M., & Dunaif, A. (2022). Deconstructing a Syndrome: Genomic Insights Into PCOS Causal Mechanisms and Classification. In *Endocrine Reviews* (Vol. 43, Issue 6, pp. 927–965). Endocrine Society. <https://doi.org/10.1210/endrev/bnac001>
- Elmannai, H., El-Rashidy, N., Mashal, I., Alohal, M. A., Farag, S., El-Sappagh, S., & Saleh, H. (2023). Polycystic Ovary Syndrome Detection Machine Learning Model Based on Optimized Feature Selection and Explainable Artificial Intelligence. *Diagnostics*, 13(8). <https://doi.org/10.3390/diagnostics13081506>
- Emanuel, R. H. K., Docherty, P. D., Lunt, H., Murray, R., & Campbell, R. E. (2024). Clustering polycystic ovary syndrome laboratory results extracted from a large internet forum with machine learning. *Intelligence-Based Medicine*, 9, 100135. <https://doi.org/10.1016/j.ibmed.2024.100135>
- Faris, N. N., & Miften, F. S. (2022). An intelligence model for detection of PCOS based on k-means coupled with LS-SVM. *Concurrency and Computation: Practice and Experience*, 34(21). <https://doi.org/10.1002/cpe.7139>
- Faris, N., Sahi, A., Diykh, M., Abdulla, S., & Siuly, S. (2025). Enhanced Polycystic Ovary Syndrome diagnosis model leveraging a K-means based genetic algorithm and ensemble approach. *Intelligence-Based Medicine*, 11. <https://doi.org/10.1016/j.ibmed.2025.100253>
- Hoeger, K. M., Dokras, A., & Piltonen, T. (2021). Update on PCOS: Consequences, Challenges, and Guiding Treatment. In *Journal of Clinical Endocrinology and Metabolism* (Vol. 106, Issue 3, pp. E1071–E1083). Endocrine Society. <https://doi.org/10.1210/clinem/dgaa839>
- Karadayı Atas, P. (2024). Exploring the Molecular Interaction of PCOS and Endometrial Carcinoma through Novel Hyperparameter-Optimized Ensemble Clustering Approaches. *Mathematics*, 12(2), 295. <https://doi.org/10.3390/math12020295>
- Kholil, K. A., Rahaningsih, N., & Danar Dana, R. (2024). Penerapan Data Mining Untuk Clustering Penyakit Diare Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3124–3131. <https://doi.org/10.36040/jati.v8i3.9616>
- Koli, S., Tayde, J., & Khan, F. (2023). *PCOS 2023 Dataset*. Kaggle. <https://www.kaggle.com/datasets/sahilkoli04/pcos2023/suggestions?status=pending>
- Kurniawati, E. Y., Hutabarat, N. C., & Noviasari, E. (2023). Status Gizi dan Gaya Hidup Wanita dengan Sindrom Ovarium Polikistik (PCOS) di Yogyakarta. *JURNAL KESEHATAN PERINTIS (Perintis's Health Journal)*, 10(1), 74–82. <https://doi.org/10.33653/jkp.v10i1.971>
- Mantik, J., & Hidayati, N. (2022). Implementation of data mining to determine stock inventory at kenza grocery stores using the k-means clustering method. In *Jurnal Mantik* (Vol. 6, Issue 3). Online.
- Mohamed, A. H., Albasheer, O., Ghoniem, M. A., Abdalghani, N., Ayish, F., Abdelwahab, S. I., Abdelmageed, M. M., Hakami, A. M. S., Khormi, A. H., Altraifi, A. A., Medani, I., Chourasia, U., Ali, S. A., Abdelmola, A., & Ahmed, A. E. (2025). Impact of lifestyle interventions on reproductive and psychological outcomes in women with polycystic ovary syndrome: A systematic review. *Medicine*, 104(3), e41178. <https://doi.org/10.1097/MD.00000000000041178>
- Nasution, A. R. S. (2021). Identifikasi Permasalahan Penelitian. *ALACRITY: Journal of Education*, 13–19. <https://doi.org/10.52121/alacrity.v1i2.21>
- National Institutes of Health, (NIH). (2024, August 21). *Polycystic Ovary Syndrome (PCOS)*. Eunice Kennedy Shriver National Institute of Child Health and Human Development. <https://www.nichd.nih.gov/health/topics/pcos>
- Nurdiawan, O., Susana, H., & Faqih, A. (2024). Deep Learning For Polycystic Ovarian Syndrome Classification Using Convolutional Neural Network. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 9(2), 218–226. <https://doi.org/10.33480/jitk.v9i2.4575>
- Novalia, V., Tania, K. D., Meiriza, A., & Wedhasmara, A. (2024, September). Knowledge discovery of application review using word embedding's comparison with CNN-LSTM model on <https://doi.org/10.35145/joisie.v9i1.4772>

- sentiment analysis. In 2024 International Conference on Electrical Engineering and Computer Science (ICECOS) (pp. 234–238). IEEE. <https://doi.org/10.1109/icecos63900.2024.10791113>
- Oktavia, R., Lanita, U., Siregar, S. A., Reskiaddin, L. O., & Perdana, S. M. (2024). Efektivitas Edukasi Kesehatan Tentang Polycystic Ovary Syndrome (PCOS) Melalui Media Sosial Terhadap Pengetahuan dan Sikap Remaja Putri di Madrasah Aliyah Laboratorium Jambi. *Jurnal Kesmas Jambi*, 8(2), 81–93. <https://doi.org/10.22437/jkmj.v8i2.32755>
- Putri, S. N., Santoso, B., & Budiono, B. (2022). Sedentary Lifestyle and Overweight in Relation to the Risk of Polycystic Ovary Syndrome in Senior High School Students in Surabaya. *JUXTA: Jurnal Ilmiah Mahasiswa Kedokteran Universitas Airlangga*, 13(2), 57–61. <https://doi.org/10.20473/juxta.V13I22022.57-61>
- Sari, D. A., Kurniawati, E. Y., & Ashari, M. A. (2023). SKRINING DAN DETERMINAN KEJADIAN SINDROM OVARIUM POLIKISTIK (SOPK) PADA REMAJA. *Jurnal Ilmu Kebidanan*, 9(2), 102–106. <https://doi.org/10.48092/jik.v9i2.211>
- Sheikdavood, K., & Bala, M. P. (2023). Polycystic Ovary Cyst Segmentation Using Adaptive K-means with Reptile Search Algorithm. *Information Technology and Control*, 52, 85–99.
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44. <https://doi.org/10.47292/joint.v2i2.30>
- Tanjung, N. N. M., & Achmad Fauzi. (2023). Hubungan Antara Kejadian Polycystic Ovarium Syndrome Dengan Akne Pada Wajah Di Nu Beauty Medical Aesthetics Jonggol TAHUN 2022. *Jurnal Ilmiah Keperawatan (Scientific Journal of Nursing)*, 9(3), 74–83. <https://doi.org/10.33023/jikep.v9i3.1592>
- WHO. (2025, February 7). *Polycystic ovary syndrome*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/polycystic-ovary-syndrome>
- Wibowo, A. S., & Mulyastuti, I. D. (2022). Penerapan Algoritma K-Means Clustering Pada Jumlah Fasilitas Kesehatan Menurut Pemerintah Provinsi DKI Jakarta. *Tekinfo: Jurnal Bidang Teknik Industri Dan Teknik Informatika*, 23(2), 116–122. <https://doi.org/10.37817/tekinfo.v23i2.2603>