

Peningkatan Identifikasi PCOS dengan KELM melalui Seleksi Fitur LDA dan Deteksi *Outlier* LOF

Panreshma Rizkha Ambadar¹, Dian C Rini Novitasari^{2,*}, Abdulloh Hamid³

Program Studi Matematika, Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia

²diancrini@uinsa.ac.id

*penulis korespondensi

INFO ARTIKEL

ABSTRAK

Sejarah Artikel

Diterima: 7 Maret 2025

Direvisi: 22 April 2025

Diterbitkan: 30 April 2025

Kata Kunci

KELM

Klasifikasi

LDA

LOF

PCOS

Polycystic Ovary Syndrome (PCOS) merupakan kelainan yang terjadi pada organ reproduksi perempuan. Kelainan ini mempunyai dampak implikasi klinis yang beragam dan serius, diantaranya pada bagian reproduksi, metabolisme, hingga gangguan psikologis. Identifikasi yang tepat sangat penting untuk meningkatkan penanganan. Penelitian ini bertujuan untuk menguji efektivitas metode *Kernel Extreme Learning Machine* (KELM) dalam mengidentifikasi PCOS setelah penghapusan *outlier* dengan *Local Outlier Factor* (LOF) dan seleksi fitur menggunakan *Linear Discriminant Analysis* (LDA). Dalam penelitian ini, metode KELM mengidentifikasi kelainan PCOS dengan klasifikasi berdasarkan data rekam medis pasien. Penelitian ini juga melibatkan pengolahan data dengan LOF untuk menangani data *outlier* dan seleksi fitur terbaik menggunakan LDA guna meningkatkan akurasi identifikasi kelainan PCOS. Berbagai uji coba dilakukan, untuk mengoptimalkan hasil identifikasi kelainan PCOS. Hasil penelitian menunjukkan bahwa ketiga kombinasi dari metode LOF, LDA, dan KELM memperoleh nilai akurasi sebesar 100% dengan eliminasi 10% data *outlier* dan 10 fitur utama. Hal ini yang menunjukkan kombinasi ketiga metode ini mampu meningkatkan kualitas deteksi dan identifikasi kelainan PCOS.

PENDAHULUAN

Kesehatan perempuan adalah konsep yang merujuk pada kondisi jasmani dari fungsi biologis seorang perempuan. Menurut *World Health Organization* (WHO), kesehatan reproduksi adalah keadaan fisik dan mental serta sosial yang utuh, bebas dari penyakit atau kecacatan yang berhubungan dengan sistem reproduksi, fungsi, serta prosesnya [1]. Banyak kasus yang dapat ditemukan mengenai perempuan yang mengalami banyak kesulitan karena fungsi reproduksinya. Sehingga paham mengenai tanda-tanda pada kelainan reproduksi dapat membantu kita untuk lebih berhati-hati dan menjaga kesehatan tubuh khususnya organ ovarium sebagai salah satu organ pada tubuh yang berperan penting dalam reproduksi perempuan [2]. Salah satu gangguan pada reproduksi perempuan yaitu kelainan *Polycystic Ovary Syndrome* (PCOS). PCOS merupakan kelainan pada organ reproduksi perempuan, yang mengarah pada kelainan reproduksi, metabolisme, dan psikologis [3]. Penderita PCOS dialami oleh sekitar 10-15% perempuan usia produktif. Berdasarkan *World Health Organization* (WHO), terdapat kurang lebih 3.4% populasi perempuan dari seluruh dunia atau sekitar 116 juta jiwa mengalami kelainan PCOS [4]. Menurut *Institute for Health Metrics and Evaluation* (IHME), di Indonesia angka PCOS mencapai 15-25% dari total jumlah penduduk Indonesia [5]. Banyaknya kasus pada kelainan PCOS, maka diperlukan pemahaman mengenai gejala dan penyebab yang berhubungan dengan kelainan PCOS.

Terdapat penelitian mengenai identifikasi data rekam medis PCOS menggunakan SVM dengan seleksi fitur GA (GA-SVM) mendapatkan hasil rata-rata akurasi, sensitivitas, dan spesifisitas masing-masing sebesar 92.14%, 84.31%, dan 95.95% [6]. Selain itu, penelitian tentang identifikasi gangguan PCOS menggunakan KNN dengan seleksi fitur RFECV dan EDA mendapatkan nilai akurasi, presisi, recall dan f1-score masing-masing sebesar 93%, 100%, 82%, 90% [7]. Pada perbandingan metode KNN dan KELM dalam mengidentifikasi tumor otak mendapatkan nilai akurasi sebesar 93,31% untuk KELM sedangkan 80,40% untuk KNN. Hal ini menunjukkan bahwa metode KELM signifikan lebih baik daripada metode KNN [8]. Selanjutnya, terdapat penelitian yang dilakukan Sheng-wei Fei dan Ying-zhe Liu [9] yang menunjukkan metode KELM memiliki kemampuan mengidentifikasi lebih baik daripada KNN, dengan akurasi untuk KNN 94,375% dan akurasi KELM 100%.

Pada penelitian mengenai PCOS perlu adanya *pra-processing data* berupa deteksi *outlier* atau data pencilan pada data sebelum dilakukan tahapan lebih lanjut. *Local Outlier Factor* (LOF) merupakan salah satu metode untuk mendeteksi *outlier* dengan mempertimbangkan struktur kerapatan lokal dari dataset secara keseluruhan. Deteksi *outlier* ini digunakan untuk menghapus data *outlier* dari data yang akan diolah [10]. Terdapat penelitian mengenai perbandingan algoritma *Deep Q-Network* (DQN) dengan LOF pada deteksi *outlier* yang menghasilkan bahwa metode LOF dapat melakukan deteksi *outlier* lebih akurat daripada DQN, dengan algoritma LOF memberikan hasil terbaik dengan nilai F1-Score tertinggi sebesar 0.36 pada skenario ke-5 [11].

Tahapan *pra-processing data* berikutnya berupa seleksi fitur yang diharapkan dapat memperbaiki tingkat ketepatan prediksi [12]. Beragamnya penanda klinis pada kelainan PCOS, diperlukan pemilihan fitur penanda klinis yang paling berpengaruh untuk mendeteksi kelainan PCOS. Salah satu algoritma yang dapat digunakan yaitu *Linear Discriminant Analysis* (LDA) yang memiliki tujuan untuk memaksimalkan jarak antar kelas dan meminimalkan jarak antara data serupa [13]. Penelitian yang menggunakan LDA dalam proses prediksi kenaikan jabatan, dan diperoleh Nilai *F1 Score training* sebesar 0,9345 sedangkan nilai *F1 Score testing* sebesar 0,9765 [14]. Berikutnya, penelitian tentang analisis efek reduksi dimensi pada klasifikasi KNN, menunjukkan perbandingan metode LDA dan PCA dengan masing-masing akurasi 96.15% dan 81.15% pada ISOLET dataset [15].

Penjelasan sebelumnya mengenai berbagai implikasi klinis yang ditimbulkan oleh kelainan PCOS menegaskan pentingnya penanganan segera bagi penderita kelainan ini. Penelitian-penelitian terdahulu telah menunjukkan efektivitas penggunaan metode identifikasi dengan klasifikasi yang dipadukan dengan seleksi fitur. Oleh karena itu, penerapan metode KELM yang dilengkapi dengan seleksi fitur LDA dengan didukung deteksi data *outlier* menggunakan LOF ini, diharapkan dapat memberikan kontribusi signifikan bagi tenaga medis, khususnya dalam upaya deteksi lebih awal terhadap pasien yang mengalami kelainan PCOS, maupun yang tidak terdiagnosis, sehingga dapat diambil langkah-langkah pengobatan yang lebih tepat dan terarah.

METODE

Local Outlier Factor (LOF)

Local Outlier Factor (LOF) merupakan metode deteksi data pencilan (*outlier*) dengan konsep membandingkan seberapa kepadatan pada data pencilan dengan sampel di sekitarnya [10]. Data yang memiliki sedikit sampel disekitarnya dan terisolasi dari lingkungan lokalnya akan disebut data pencilan. Dan sebaliknya data tersebut dikatakan normal, apabila termasuk ke dalam suatu kluster dan memiliki banyak tetangga. Metode LOF efektif dalam mendeteksi data berdimensi tinggi [11].

Langkah-langkah pada perhitungan LOF yaitu sebagai berikut [16].:

1. Menghitung jarak k pada setiap titik p di kumpulan data

$$d(p, o) = \sqrt{\sum_{i=1}^n (p_i - o_i)^2} \quad (1)$$

dimana $d(p, o)$ adalah jarak Euclidean antara titik p dan o . Lalu, n adalah pada jumlah banyak fitur.

2. Menentukan tetangga terdekat terhadap p

$$N_{k-distance(p)}(p) = \{q \in D \setminus \{p\} \mid d(p, q) \leq k - distance(p)\} \quad (2)$$

dimana $N_{k-distance(p)}$ merujuk pada himpunan semua titik q dalam dataset D , kecuali titik p . Jarak $k - distance(p)$ adalah jarak maksimum tetangga terdekat yang sudah ditentukan sebelumnya.

3. Menghitung *Reachability Distance* (RD) ke setiap tetangga (o) untuk setiap titik p

$$RD_k(p, o) = \max [k - distance(o), d(p, o)] \quad (3)$$

dimana $RD_k(p, o)$ merujuk pada nilai maksimum antara $k - distance(o)$ dan jarak antara p dan o .

4. Menghitung *Local Reachability Density*

$$LRD_k(p) = \left(\frac{\sum_{o \in N_k(p)} RD_k(p, o)}{|N_k(p)|} \right)^{-1} \quad (4)$$

dimana $LRD_k(p)$ dihitung dengan menjumlahkan nilai $RD_k(p, o)$ untuk semua titik o dalam himpunan tetangga k -distance $N_k(p)$ kemudian dibagi dengan jumlah titik dalam himpunan tersebut.

5. Menghitung *Local Outlier Factor*

$$LOF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{LRD_k(o)}{LRD_k(p)}}{|N_k(p)|} \quad (5)$$

dimana $LOF_k(p)$ dihitung dengan membandingkan LRD titik p dengan LRD dari tetangga-tetangganya dalam himpunan $N_k(p)$ yang merupakan titik-titik dalam jarak k -distance dari titik p . Jika titik p memiliki LRD yang jauh lebih rendah daripada tetangganya, maka nilai LOF makin besar yang menunjukkan bahwa p adalah outlier.

6. Mengidentifikasi *outlier* berdasarkan nilai LOF. Nilai hasil LOF yang jauh lebih tinggi daripada satu dianggap *outlier* [17].

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) adalah teknik seleksi fitur yang banyak digunakan. Teknik ini terutama menghilangkan fitur yang berlebihan dan bergantung dari dataset. Pada dasarnya, teknik ini terdiri dari tiga tahap utama. Pada tahap pertama, itu menghitung perbedaan antara rata-rata yang berbeda kelas. Perbedaan ini dikenal sebagai varians antar kelas. Pada tahap kedua, ia menghitung perbedaan antara rata-rata dan nilai sampel dari setiap kelas. Perbedaan ini dikenal sebagai varians dalam kelas. Pada tahap ketiga, itu memilih fitur yang memiliki varians antar kelas yang lebih besar yang lebih besar dan varians dalam kelas yang lebih kecil [18]. Pada LDA, terdapat perluasan metode dengan menggunakan Fisher yang dikenal dengan Fisher's discriminant (FLDA) yang memiliki

kegunaan yang hampir sama dengan LDA. Namun, pada FLDA ini tidak perlu melakukan asumsi multivariat normal, tetapi hanya menggunakan asumsi bahwa kovarian matriks adalah sama atau homogen. Lebih lanjut, langkah-langkah dalam melakukan seleksi fitur data menggunakan LDA sebagai berikut:

1. Menghitung rata-rata setiap kelas label. Dapat dijabarkan dengan proses menghitung rata-rata dari tiap-tiap kelas (\bar{x}_i) dan rata-rata kelas keseluruhan dari seluruh data (\bar{x}) [19].

$$\bar{x}_i = \frac{1}{r} \sum_{j=1}^r p_{ij} \quad i = 1, 2, \dots, z \quad (6)$$

$$\bar{x} = \frac{\sum_{i=1}^g r_i \bar{x}_i}{\sum_{i=1}^g r_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} p_{ij}}{\sum_{i=1}^g r_i} \quad (7)$$

dimana \bar{x} adalah rata-rata keseluruhan dengan menjumlahkan rata-rata setiap kelas \bar{x}_i yang masing-masing dikalikan dengan jumlah elemen r_i dalam kelompok lalu dibagi dengan total jumlah elemen dari semua kelompok. Pada \bar{x}_i , untuk setiap kelas i , nilai \bar{x}_i diperoleh dengan menjumlahkan semua elemen p_j (dari 1 hingga r), kemudian membaginya dengan r yang menunjukkan rata-rata dalam kelompok tersebut.

2. Menghitung matriks sebaran dalam kelas (S_W) dan matriks antarkelas (S_B)

$$S_W = \sum_{i=1}^g \sum_{j=1}^{r_i} (p_{ij} - \bar{x}_i)^T (p_{ij} - \bar{x}_i) \quad (8)$$

$$S_B = \sum_{i=1}^g r_i (\bar{x}_i - \bar{x})^T (\bar{x}_i - \bar{x}) \quad (9)$$

dimana S_W mengukur sebaran data dalam masing-masing kelas, dengan menghitung selisih antara setiap data p_{ij} dan rata-rata kelas \bar{x}_i kemudian kalikan dengan hasil *transpose*-nya. Sedangkan S_B mengukur variansi antar kelas dalam data, dengan menghitung antara rata-rata kelas \bar{x}_i dan rata-rata keseluruhan \bar{x} .

3. Menghitung matriks kovarian atau dapat disebut juga dengan kriteria Fisher untuk menemukan fitur terbaik.

$$S = (S_W)^{-1} S_B \quad (10)$$

dimana S_W adalah matriks sebaran dalam kelas dan S_B adalah matriks sebaran antar kelas. Pembagian antara kedua matriks tersebut menggunakan nilai diagonalnya saja untuk menunjukkan nilai variansi dari matriks S_W . Hasil pembagian berupa nilai kriteria fisher yang menunjukkan ranking pada fitur-fitur yang ada.

Kernel Extreme Learning Machine (KELM)

Kernel Extreme Learning Machine (KELM) adalah metode perkembangan dari *Extreme Learning Machine* (ELM). ELM adalah jenis jaringan saraf *feedforward* dengan satu lapisan tersembunyi, yang menjadikan pembelajaran ELM lebih cepat daripada jaringan saraf tradisional [20]. Perbedaan antara ELM dan KELM tersebut terletak pada fungsi kernel di *hidden layer*. Penambahan fungsi kernel pada metode KELM, merupakan adaptasi dari metode *Support Vector Machine* (SVM). SVM menggunakan kernel untuk meningkatkan dimensi data sehingga data dapat dipisahkan secara optimal [21]. Terdapat dua tahapan pada KELM, yaitu proses data *training* dan data *testing*. Proses training bertujuan untuk

memperoleh *output* yang berupa nilai keluaran bobot [22]. Berikut ini tahapan *training* pada KELM:

1. Menyiapkan matriks dengan x_i adalah input data training dan t_i adalah label data *training*
2. Inisialisasi fungsi kernel dengan K. Fungsi kernel yang dapat digunakan yaitu:

- a. Fungsi Linear

$$K(x_i, x_j) = x_i x_j^T \quad (11)$$

- b. Fungsi Polinomial

$$K(x_i, x_j) = (x_i x_j^T + 1)^d \quad (12)$$

- c. Fungsi RBF (*Radial Basis Function*)

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right) \quad (13)$$

Dilanjutkan membentuk matriks Ω_{KELM}

3. Setelah didapatkan matriks Ω_{KELM} , dilanjutkan menghitung nilai bobot keluaran (β) dengan Y yaitu kelas dari data training dengan persamaan

Tahapan selanjutnya yaitu proses testing data. Proses testing bertujuan untuk mengevaluasi proses training di tahap sebelumnya pada implementasi metode KELM [22]. Berikut ini tahapan *testing* pada KELM:

1. Menghitung matriks omega (Ω), dengan menggunakan jenis fungsi kernel seperti langkah 2 pada proses training, akan tetapi pada tahap testing menggunakan data testing, dengan N merupakan banyaknya data training dan M merupakan banyaknya data testing, sehingga persamaan kernel berubah menjadi:
2. Setelah didapatkan matriks omega (Ω), maka selanjutnya yaitu menghitung label kelas (Y_k) dengan menggunakan nilai (β) yang diperoleh dari tahap *training*.

Confusion Matrix

Confusion Matrix merupakan teknik yang digunakan untuk menyajikan hasil akurasi dari model yang telah dibuat [23]. Perhitungan ini berguna untuk mengukur kinerja dari suatu sistem klasifikasi [24]. *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja model identifikasi kelainan PCOS pada serangkaian data uji yang nilai sebenarnya diketahui sebagaimana Tabel 1 [25].

Tabel 1. *Confusion matrix*

Kelas Aktual	Kelas Prediksi	
	Positif PCOS	Negatif PCOS
Positif PCOS	TP	FP
Negatif PCOS	FN	TN

dimana TP (*True Positive*) adalah jumlah dokumen dari kelas positif yang benar diidentifikasi sebagai kelas positif. TN (*True Negative*) adalah jumlah dokumen dari kelas negatif yang benar diidentifikasi sebagai kelas negative. FP (*False Positive*) adalah jumlah dokumen dari kelas negatif yang salah diidentifikasi sebagai kelas positif. FN (*False Negative*) adalah jumlah dokumen dari kelas positif yang salah diidentifikasi sebagai kelas negatif

Confusion Matrix digunakan untuk menghitung nilai validitas model berupa evaluasi akurasi, sensitivitas dan spesifisitas. Akurasi adalah tingkat keberhasilan dari sebuah sistem identifikasi kelainan PCOS. Sensitivitas adalah nilai yang terdiagnosis penyakit dalam hasil

identifikasi kelainan PCOS, sedangkan spesifisitas adalah nilai yang penderita yang tidak terdiagnosis dalam hasil, berikut perhitungan akurasi, sensitivitas, dan spesifisitas [26]:

$$akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

$$sensitivitas = \frac{TP}{TP + FN} \quad (15)$$

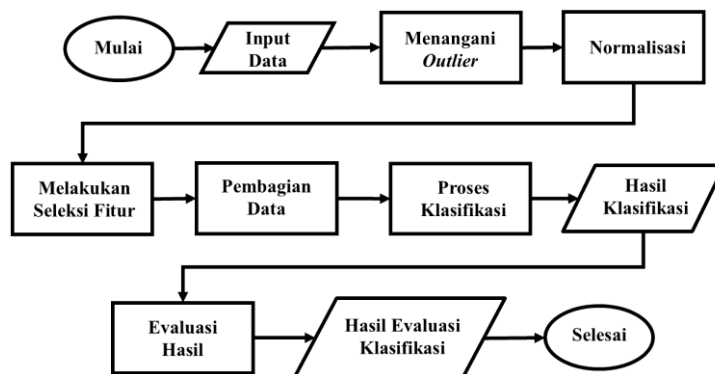
$$spesifisitas = \frac{TN}{TN + FP} \quad (16)$$

Penelitian ini menggunakan data penanda klinis dan biokimia dalam mendiagnosis PCOS yang bersumber dari Kaggle [27]. Total data yang digunakan terdiri atas 541 dengan dua kelas yaitu sebanyak 177 pasien positif mengidap PCOS dan sebanyak 364 pasien negatif mengidap PCOS. Terdapat 39 fitur penanda klinis dan fisik penderita PCOS dengan 1 fitur target. Tabel 2 menunjukkan sampel data kelainan PCOS yang akan digunakan pada penelitian ini.

Tabel 2. Data Sampel Kelainan PCOS

No	x_1	x_2	x_3	...	x_{37}	x_{38}	x_{39}	y
1	28	44.6	152.0	...	0	1	0	0
2	36	65.0	161.5	...	0	0	0	0
3	33	68.8	165.0	...	1	1	0	1
4	37	65.0	148.0	...	0	0	0	0
...
541	23	82.0	165.0	...	1	1	0	1

Penjelasan fitur yang digunakan yaitu: Usia pasien saat survei dilakukan (x_1), Berat badan pasien saat survei dilakukan (x_2), Tinggi badan pasien saat survei dilakukan (x_3), Indeks Massa Tubuh pasien (x_4), Denyut nadi pasien (x_5), Tingkat pernapasan pasien (x_6), Hemoglobin pasien (x_7), Panjang siklus menstruasi (x_8), Lama pernikahan pasien (x_9), Jumlah aborsi pada pasien (x_{10}), Kadar hormon perangsang folikel (x_{11}), Kadar hormon luteinisasi (x_{12}), Rasio FSH dan LH (x_{13}), Ukuran pinggul pasien (x_{14}), Ukuran pinggang pasien (x_{15}), Rasio Pinggang dan Pinggul (x_{16}), Kadar hormon perangsang tiroid. (x_{17}), Kadar hormon anti-mullerian (x_{18}), Kadar prolaktin pasien (x_{19}), Kadar vitamin D3 (x_{20}), Kadar progesteron pasien (x_{21}), Kadar gula darah pasien (x_{22}), Tekanan darah sistolik pasien (x_{23}), Tekanan darah diastolik pasien (x_{24}), Jumlah folikel ovarium kiri (x_{25}), Jumlah folikel ovarium kanan (x_{26}), Rata-rata ukuran folikel kiri (x_{27}), Rata-rata ukuran folikel kanan (x_{28}), Ketebalan endometrium pasien (x_{29}), Golongan darah pasien (x_{30}), Siklus menstruasi pasien (x_{31}), Apakah pasien sedang hamil (x_{32}), Apakah pasien mengalami penambahan berat badan (x_{33}), Apakah pasien mengalami pertumbuhan rambut berlebih (x_{34}), Apakah pasien mengalami penggelapan kulit (x_{35}), Apakah pasien mengalami rambut rontok (x_{36}), Apakah pasien memiliki jerawat (x_{37}), Apakah pasien sering mengonsumsi makanan cepat saji (x_{38}), Apakah pasien rutin berolahraga (x_{39}), Kondisi PCOS pada pasien (y). Berikutnya Gambar 1 merupakan diagram alir yang digunakan pada penelitian ini.



Gambar 1. Diagram Alir Penelitian

Berdasarkan Gambar 1, Tahapan yang dilakukan pada penelitian ini dimulai dengan tahapan awal yaitu menginput data yang diperoleh dari web Kaggle [27]. Data yang telah terunduh berjumlah 541 data dengan 39 fitur penanda klinis dan fisik penderita PCOS, terdiri atas dua kelas yaitu positif PCOS dan negatif PCOS. Tahapan selanjutnya yaitu deteksi *outlier* atau data pencilan. Data pencilan ini berpengaruh dalam hasil pengolahan data. Penanganan data *outlier* dapat menggunakan metode LOF. Selanjutnya, fitur pada data yang beragam dapat dilakukan seleksi fitur untuk memilah fitur yang paling berpengaruh ke dalam pengolahan data. Seleksi data dapat menggunakan metode LDA. Tahapan sebelum melakukan identifikasi kelainan PCOS, adalah melakukan normalisasi data yang bertujuan untuk melakukan standarisasi terhadap nilai baku atau nilai mentah dari data. Metode yang digunakan yaitu *Z-Score Normalization*. Pada identifikasi kelainan PCOS, dilakukan pembagian data sama besar untuk data *training* dan data *testing* menggunakan metode *K-Fold Cross Validation* [28]. Nilai *k* pada *K-Fold Cross Validation* yang digunakan adalah 10 yang berarti terdapat 10 bagian, dengan tiap satu bagian pasti akan menjadi data *training* dan data *testing*. Selanjutnya, data *training* dan data *testing* tersebut diolah ke dalam proses identifikasi kelainan PCOS. Pada klasifikasi KELM menggunakan uji coba 3 kernel, yaitu kernel linear, polinomial, dan RBF. Berikut ini proses data training dan proses data testing. Selanjutnya dilakukan evaluasi berdasarkan hasil uji coba yang telah dilakukan menggunakan *confusion matrix* untuk mendapatkan nilai akurasi, sensitivitas, dan spesifisitas dalam penelitian ini.

HASIL DAN PEMBAHASAN

Berdasarkan Gambar 1, tahap awal yaitu menangani data *outlier*. Pada Tabel 1, menunjukkan data awal kelainan PCOS yang terdiri dari 541 data. Berdasarkan hasil perhitungan yang telah dilakukan dengan metode LOF menggunakan persamaan (1)-(5), terdapat data *outlier* sebesar 54 data atau 10% dari data awal. Tahap selanjutnya yaitu data *outlier* tersebut akan dihapus dari data awal PCOS, sehingga total dataset menjadi 487 data yang ditunjukkan pada Tabel 3.

Tabel 3. Data Inlier Kelainan PCOS

Data ke-	x_1	x_2	x_3	x_4	...	x_{39}
1	28	44.6	152	19.3	...	0
2	36	65	161.5	24.9	...	0
3	33	68.8	165	25.2	...	0

4	37	65	148	29.6	...	0
...
487	23	82	165	30.1	...	0

Pada Tabel 3 menunjukkan data *inlier* PCOS yaitu data yang bebas dari data *outlier*. Data bebas *outlier* terdiri dari 487 data yang akan dilanjutkan ke proses berikutnya berupa seleksi fitur. Fitur awal data kelainan PCOS yaitu sebanyak 39 fitur x berupa penanda fisik dan klinis. Berdasarkan hasil perhitungan dengan metode LDA menggunakan persamaan (6)-(10), terdapat 10 fitur dengan urutan pertama teratas, dimana 10 fitur ini akan digunakan dalam proses selanjutnya. Berikut adalah nama-nama fitur terpilih: *Follicle No. (R)*, *Follicle No. (L)*, *Skin darkening (Y/N)*, *Hair growth (Y/N)*, *Weight gain (Y/N)*, *Fast food (Y/N)*, *Cycle (R/I)*, *Pimples (Y/N)*, *AMH (ng/mL)*, dan *Weight (Kg)*, masing-masing fitur ke- $x_{27}, x_{26}, x_{35}, x_{34}, x_{33}, x_{38}, x_{31}, x_{37}, x_{19}$, dan x_2 . Selanjutnya melakukan normalisasi menggunakan *Z-Score Normalization* dengan cara menggunakan nilai rata-rata dan standar deviasi pada data, sehingga data memiliki rentang nilai yang sama. Hasil normalisasi ditunjukkan pada Tabel 4.

Tabel 4. Hasil Normalisasi

Data ke-	x_{27}	x_{26}	x_{35}	...	x_2
1	-0.8203	-0.7290	-0.6543	...	-1.3870
2	-0.3606	-0.7290	-0.6543	...	0.5456
3	1.9380	1.6378	-0.6543	...	0.9055
4	-1.0502	-0.9657	-0.6543	...	0.5456
...
487	0.7887	0.6911	1.5283	...	2.1561

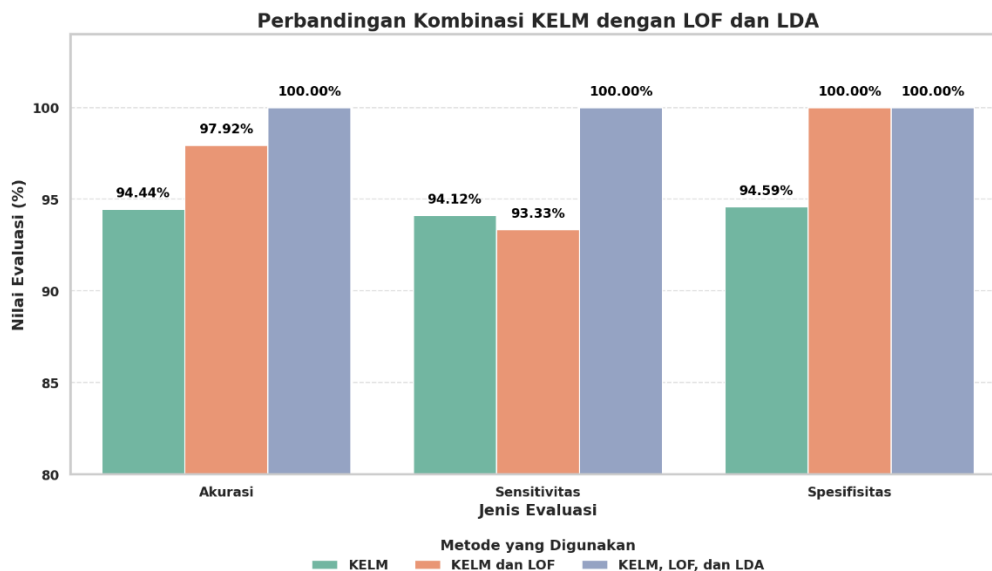
Pada Tabel 3 menunjukkan hasil normalisasi dari perhitungan *Z-Score*. Data terdiri dari 10 fitur pilihan dan data yang sudah terbebas *outlier*, dimana data ini digunakan pada proses berikutnya yaitu identifikasi kelainan PCOS menggunakan KELM. Proses identifikasi menggunakan KELM dengan pembagian data *k-fold* sebesar 10. Data dibagi menjadi 10 bagian yang mana tiap bagian tersebut akan mendapati giliran menjadi data latih (*training*) maupun data uji (*testing*). Pada identifikasi kelainan PCOS dengan KELM saja, uji coba menggunakan tiga kernel, yaitu linear, polinomial, dan RBF masing-masing ditunjukkan di persamaan (11), (12), dan (13), didapatkan hasil masing-masing sebesar 94.44%, 72.22%, dan 78.18%. Hal ini menunjukkan kernel linear sebagai kernel dengan nilai akurasi tertinggi daripada ketiga kernel yang lain. Sehingga, dilakukan perhitungan uji coba KELM menggunakan kernel linear dengan gabungan metode LOF dan LDA mendapatkan hasil yang dirangkum pada Tabel 5.

Tabel 5. Hasil Identifikasi Kelainan PCOS KELM dengan LOF dan LDA

Metode	Akurasi	Sensitivitas	Spesifisitas
KELM	94.44%	94.12%	94.59%
KELM dan LOF	97.92%	93.33%	100%
KELM, LOF, dan LDA	100%	100%	100%

Berdasarkan Tabel 5 menunjukkan hasil evaluasi performa berbagai kombinasi metode berbasis kernel linear dengan nilai validitas berupa akurasi, sensitivitas, dan spesifisitas. Metode KELM dasar memberikan hasil yang baik dengan akurasi sebesar 94.44%, sensitivitas 94.12%, dan spesifisitas 94.59%. Hasil ini meningkat seiring dengan

penambahan metode LOF dan LDA. Kombinasi dari ketiga metode KELM, LOF, dan LDA memberikan hasil terbaik dengan akurasi, sensitivitas, dan spesifisitas masing-masing sebesar 100%. Hal ini menunjukkan bahwa penggunaan ketiga teknik ini secara bersamaan mampu mengidentifikasi kelainan PCOS secara optimal, baik dalam mengenali kelas positif maupun negatif. Berikutnya, disajikan visualisasi digram batang pada Gambar 2.



Gambar 2. Perbandingan Kombinasi KELM dengan LOF dan LDA

Metode KELM saja sudah cukup baik dalam mengidentifikasi kelainan PCOS, namun adanya kombinasi dengan LOF dan LDA memberikan peningkatan kinerja yang signifikan. Hal ini menunjukkan bahwa kombinasi dari beberapa metode mampu meningkatkan kemampuan model untuk menangani proses identifikasi kelainan PCOS dengan lebih baik. Selanjutnya dilakukan evaluasi berdasarkan hasil uji coba yang telah dilakukan menggunakan *Confusion Matrix* sebagaimana Tabel 6 untuk mendapatkan nilai akurasi, sensitivitas, dan spesifisitas dalam penelitian ini.

Tabel 6. *Confusion Matrix*

Kelas Aktual	Kelas Prediksi	
	Positif PCOS	Negatif PCOS
Positif PCOS	33	0
Negatif PCOS	0	15

Hasil perhitungan menunjukkan bahwa model identifikasi kelainan PCOS memiliki kinerja yang sempurna dengan nilai akurasi, sensitivitas, dan spesifisitas masing-masing sebesar 100%. Nilai akurasi yang tinggi ini menunjukkan bahwa model mampu mengidentifikasi semua data dengan benar, baik untuk kelas positif maupun negatif. Sensitivitas sebesar 100% menunjukkan bahwa model mampu mengidentifikasi seluruh data positif dengan benar. Dengan kata lain, semua kasus positif *True Positive* (TP) berhasil diidentifikasi, tanpa ada satu pun kasus positif yang salah diidentifikasi sebagai negatif *False Negative* (FN). Ini menandakan kemampuan model yang sangat baik dalam menangani data dari kelas positif. Nilai spesifisitas sebesar 100% juga menunjukkan bahwa model juga mampu mendeteksi seluruh data negatif dengan benar. Tidak ada satu pun kasus negatif yang salah diidentifikasi sebagai positif *False Positive* (FP). Hal ini memperlihatkan bahwa model memiliki akurasi yang sangat baik dalam menangani data dari kelas negatif. Namun, untuk memastikan bahwa performa ini tidak hanya berlaku pada data

uji saat ini, penting untuk melakukan validasi lanjutan seperti *cross-validation* atau pengujian dengan dataset eksternal. Hal ini untuk menghindari *overfitting* dan memastikan bahwa model benar-benar generalisasi dengan baik terhadap data nyata di luar sampel pelatihan.

KESIMPULAN

Identifikasi kelainan PCOS menggunakan metode *Kernel Extreme Learning Machine* (KELM) mendapatkan nilai akurasi yang cukup tinggi yaitu sebesar 94.44%. Namun, kombinasi dari ketiga metode ini (LOF, LDA, dan KELM) mendapatkan peningkatan akurasi sebesar 5.36% hingga nilai akurasi dapat mencapai 100%. Hal ini menunjukkan bahwa integrasi metode LOF dan LDA mampu memperkuat proses pra-pemrosesan data dan seleksi fitur yang tepat, sehingga meningkatkan kemampuan KELM dalam melakukan identifikasi kelainan PCOS. Kombinasi ini dapat menjadi pendekatan yang sangat efektif untuk mendeteksi kelainan PCOS sehingga mendapatkan nilai akurasi lebih tinggi daripada hanya KELM saja.

REFERENSI

- [1] H. Nurlaeli, "Pentingnya Pendidikan Kesehatan Reproduksi dan Seksualitas pada Remaja Santri Putri Pondok Pesantren Watu Ringkel Darussalam-Karangpucung," *Pros. Semin. Nas. WIJAYAKUSUMA*, pp. 204–215, 2020, [Online]. Available: <https://ejournal.unugha.ac.id/index.php/jarlit/article/download/289/235>
- [2] M. Mintarsih and P. Pitrotussaadah, "Hak-Hak Reproduksi Perempuan dalam Islam," *J. Stud. Gen. dan Anak*, vol. 9, no. 01, p. 93, 2022, doi: 10.32678/jsga.v9i01.6060.
- [3] A. Yuliadha and R. H. Setyaningrum, "Psikoneuroimunologi Depresi pada Polycystic Ovary Syndrome (PCOS)," *Smart Med. J.*, vol. 5, no. 1, p. 38, 2022, doi: 10.13057/smj.v5i1.43238.
- [4] C. Rezki, "Literature Review : Coping Stress Pada Wanita dengan Polycystic Ovary Syndrome (PCOS)," *Psikobuletin*, vol. 5, no. 3, pp. 371–381, 2024, doi: <https://doi.org/10.24014/pib.v5i3.25897>.
- [5] M. Fasitasari, R. Aryani, and S. Samsudin, "Pendampingan bagi Pasien Polycystic Ovary Syndrome (PCOS) di RSI Sultan Agung Semarang dalam Modifikasi Lifestyle dan Pendekatan Aspek Psikospiritual," *J. ABDIMAS-KU J. Pengabd. Masy. Kedokt.*, vol. 3, no. 1, pp. 13–26, 2024, doi: <http://dx.doi.org/10.30659/abdimasku.3.1.13-26>.
- [6] F. Novianti and N. Ulinnuha, "Seleksi Fitur Algoritma Genetika dalam Klasifikasi Data Rekam Medis PCOS Menggunakan SVM," *J. Ilm. NERO*, vol. 9, no. 1, pp. 9–19, 2024.
- [7] N. T. Pitaloka and K. Kusnawi, "Pcos Disease Classification Using Feature Selection Rfecv and Eda With Knn Algorithm Method," *J. Tek. Inform.*, vol. 4, no. 4, pp. 693–701, 2023, doi: 10.52436/1.jutif.2023.4.4.831.
- [8] V. V. V. Reddy and P. S. U. Priyadarsini, "Brain tumor detection using enhanced kernel extreme learning with deep belief network and compare prediction accuracy with K-nearest neighbors," *AIP Conf. Proc.*, vol. 2822, no. 1, p. 20034, Nov. 2023, doi: 10.1063/5.0172877.
- [9] S. Fei and Y. Liu, "Fault diagnosis method of bearing utilizing GLCM and MBASA - based," *Sci. Rep.*, pp. 1–8, 2022, doi: 10.1038/s41598-022-19209-1.
- [10] Mutmainah and W. Yustanti, "Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen," *JINACS (Journal Informatics Comput. Sci. ISSN)*, vol. 06, no. 02, pp. 532–540, 2024.
- [11] A. A. Daniswara, B. Rahmat, and E. Y. Puspaningrum, "Perbandingan Algoritma Deep Q-Network dan Local Outlier Factor Untuk Deteksi Anomali Konsumsi Air Minum Pelanggan PUDAM Kabupaten Banyuwangi," *Mars J. Tek. Mesin, Ind. Elektro Dan Ilmu Komput.*, vol. 2, no. 4, pp. 144–156, 2024.
- [12] W. I. Sabilla, C. B. Vista, and D. S. Hormansyah, "Implementasi Multilayer Perceptron Untuk Memprediksi Harapan Hidup Pada Pasien Penyakit Kardiovaskular," *J. Sains Komput. Inform.*, vol. 6, no. 1, pp. 57–68, 2022.
- [13] A. Ariwikri, G. A. Pratama, H. Fathurahman, M. A. F. S. Fikri, and M. F. Alrizki, "Perbandingan Akurasi Metode Naive Bayes Classifier Dan Random Forest Menggunakan Reduksi Dimensi Linear Dicitiran Analysis (Lda) Untuk Diagnosis Penyakit Diabetes," *J. Rekayasa Elektro Sriwij.*, vol. 4, no. 1, pp. 24–31, 2023, doi: 10.36706/jres.v4i1.58.

- [14] S. R. P. Ariyanto and W. Yustanti, "Prediksi Kenaikan Jabatan Pranata Komputer pada Kementerian X dengan Menggunakan Model Algoritma Klasifikasi Linear Discriminant Analysis (LDA)," *JEISBI (Journal Emerg. Inf. Syst. Bus. Intell.)*, vol. 04, no. 03, pp. 40–49, 2023.
- [15] E. B. N. S. E. Taufiqurrahman, "Analysis of dimensional reduction effect on K-Nearest Neighbor classification method," *J. dan Penelit. Tek. Inform.*, vol. 6, no. 1, pp. 222–230, 2021.
- [16] O. Alghushairy, R. Alsini, T. Soule, and X. Ma, "A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams," *big data Cogn. Comput.*, pp. 1–24, 2021.
- [17] G. Airlangga, "Advanced Machine Learning Techniques for Seismic Anomaly Detection in Indonesia: a Comparative Study of Lof, Isolation Forest, and One-Class Svm," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 5, no. 1, pp. 49–61, 2024, doi: 10.46306/lb.v5i1.490.
- [18] S. Pande, A. Khamparia, and D. Gupta, "Feature selection and comparison of classification algorithms for wireless sensor networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 3, pp. 1977–1989, 2023, doi: 10.1007/s12652-021-03411-6.
- [19] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Pearson Education, Inc., 2007.
- [20] A. J and P. Suresh L, "A novel fast hybrid face recognition approach using convolutional Kernel extreme learning machine with HOG feature extractor," *Meas. Sensors*, vol. 30, p. 100907, 2023, doi: <https://doi.org/10.1016/j.measen.2023.100907>.
- [21] D. Candra, R. Novitasari, A. N. Ramadanti, and D. Z. Haq, "Enhancing Covid-19 Diagnosis : Grlm Texture Analysis And Kelm For Lung X-Ray Classification Abstrak," *Fountain Informatics J.*, vol. 9, no. 1, pp. 11–18, 2024, doi: <http://dx.doi.org/10.21111/fij.v9i1.12166>.
- [22] M. A. P. Nasution, I. Cholissodin, and I. Indriati, "Prediksi Price Earning Ratio Saham Menggunakan Algoritme Kernel Extreme Learning Machine (Studi Kasus: PT TELKOM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 4, no. 10, pp. 3455–3462, 2020, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/8006>
- [23] A. B. Pranata, A. R. Abdillah, and F. Irwiensyah, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 6, pp. 3091–3098, 2024, doi: 10.23960/jitet.v11i3s1.3348.
- [24] E. F. Rusydiyah *et al.*, "Skin Cancer Diagnosis System on Object Detection Using Various Cnn Yolov5 in Android Mobile," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 15, pp. 5976–5986, 2024.
- [25] K. Kenny, D. Arisandi, and T. Sutrisno, "Evaluasi Penilaian Kinerja Karyawan dengan Metode Naïve Bayes," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 8, no. 1, pp. 110–118, 2024, doi: 10.24912/computatio.v8i1.20287.
- [26] S. N. Aisah, D. Candra, R. Novitasari, and Y. Farida, "Perbandingan Metode Extreme Learning Machine (ELM) dan Kernel Extreme Learning Machine (KELM) Pada Klasifikasi Penyakit Cedera Panggul," *J. FOURIER*, vol. 12, no. 2, pp. 69–78, 2023, doi: 10.14421/fourier.2023.122.69-78.
- [27] D. P. N. V K, "pcos-data," *Kaggle*, 2022. <https://www.kaggle.com/datasets/prasadbobby/pcosdata> (accessed Aug. 19, 2024).
- [28] K. Auliyatuz Zahroh *et al.*, "Perbandingan Ekstraksi Fitur Untuk Klasifikasi COVID-19, MERS, dan SARS Menggunakan Algoritma Extreme Learning Machine," *J. Fourier*, vol. 13, no. 1, pp. 30–41, 2024, doi: 10.14421/fourier.2024.131.30-41.