

PERBANDINGAN TINGKAT AKURASI ANTARA SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN KERNEL POLINOMIAL & KERNEL RADIAL BASIS FUNCTION (RBF) UNTUK KLASIFIKASI PENYAKIT POLYCYSTIC OVARY SYNDROME (PCOS)

Winda Fortuna Br Situmorang
Universitas Sumatera Utara
 e-mail: windafortuna19@gmail.com

INFORMASI ARTIKEL

Submitted : 2025-6-30
Review : 2025-6-30
Accepted : 2025-6-30
Published : 2025-6-30

KATA KUNCI

PCOS, Klasifikasi, Support Vector Machine, Kernel Polynomial, Kernel Rbf, Regresi Logistik.

Keywords: *PCOS, Classification, Support Vector Machine, Polynomial Kernel, RBF Kernel, Logistic Regression.*

ABSTRAK

Polycystic Ovary Syndrome (PCOS) merupakan gangguan hormonal yang umum dialami wanita usia reproduksi. Diagnosis dini terhadap PCOS sangat penting untuk mencegah komplikasi jangka panjang. Dalam penelitian ini, penulis melakukan klasifikasi data PCOS menggunakan metode Support Vector Machine (SVM) dengan dua jenis kernel, yaitu Kernel Polinomial dan Kernel Radial Basis Function (RBF), serta melakukan perbandingan terhadap metode Regresi Logistik. Proses klasifikasi dengan menggunakan dataset yang telah diproses melalui tahap preprocessing dan pembagian data. Proses training dan testing dilakukan untuk mengevaluasi kinerja model. Model SVM dengan Kernel Polinomial menghasilkan akurasi sebesar 71%, sedangkan model SVM dengan Kernel RBF memberikan hasil akurasi sebesar 72%. Sebagai perbandingan, metode Regresi Logistik memperoleh akurasi tertinggi sebesar 74%. Hasil ini menunjukkan bahwa SVM dengan Kernel RBF mampu memberikan performa yang kompetitif dibandingkan metode lainnya, terutama dalam menangani data yang bersifat non-linier.

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is a hormonal disorder that is common in women of reproductive age. Early diagnosis of PCOS is very important to prevent long-term complications. In this study, the author classified PCOS data using the Support Vector Machine (SVM) method with two types of kernels, namely the Polynomial Kernel and the Radial Basis Function (RBF) Kernel, and compared it to the Logistic Regression method. The classification process uses a dataset that has been processed through the preprocessing and data division stages. The training and testing processes

are carried out to evaluate model performance. The SVM model with the Polynomial Kernel produces an accuracy of 71%, while the SVM model with the RBF Kernel produces an accuracy of 72%. In comparison, the Logistic Regression method obtained the highest accuracy of 74%. These results indicate that SVM with the RBF Kernel is able to provide competitive performance compared to other methods, especially in handling non-linear data.

PENDAHULUAN

Support Vector Machine (SVM) merupakan salah satu metode yang dapat digunakan untuk proses klasifikasi dengan tingkat akurasi yang tinggi. Tujuan dari klasifikasi adalah mengelompokkan data ke dalam kategori atau kelas berdasarkan fitur-fitur yang telah tersedia. Secara umum, metode SVM bekerja dengan cara menentukan batas pemisah antara dua kelas dengan jarak maksimal dari data yang paling dekat dengan batas tersebut. Untuk memperoleh batas maksimal ini, sistem harus membentuk hyperplane (garis pemisah) terbaik pada ruang berdimensi tertentu. Hyperplane ini diperoleh dengan mengukur margin dan mencari titik maksimal margin antara dua kelas. Margin adalah jarak antara hyperplane dan data terdekat dari masing-masing kelas, yang disebut sebagai support vector (Achmad Rizal et al., 2019).

Dalam proses klasifikasi, pemilihan metode yang tepat sangat memengaruhi tingkat akurasi dari sistem. Dua metode yang sering digunakan untuk klasifikasi adalah Support Vector Machine (SVM) dan regresi logistik. SVM digunakan untuk memisahkan data ke dalam dua kelas dengan mencari hyperplane terbaik yang memiliki margin terbesar antara kelas. Berbeda dengan SVM, regresi logistik bertujuan untuk memodelkan hubungan antara variabel input dan output yang dikategorikan dengan menghitung probabilitas kelas menggunakan fungsi sigmoid. Perbedaan utama antara kedua metode ini terletak pada pemisahan data; SVM fokus pada margin maksimum, sedangkan regresi logistik fokus pada estimasi probabilitas. Regresi logistik lebih cocok untuk data yang memiliki hubungan linier antar fitur dan kelas. Sebaliknya, SVM lebih efektif untuk data yang bersifat non-linier dengan menggunakan kernel seperti polinomial dan RBF, sehingga metode SVM sering dipilih untuk kasus klasifikasi.

Dalam penelitian ini, peneliti menggunakan kernel polinomial dan kernel Radial Basis Function (RBF) untuk meningkatkan akurasi dalam proses klasifikasi data. Kernel polinomial digunakan karena mampu mengenali data yang tidak linier, namun masih memiliki hubungan tertentu antar variabel. Sementara itu, kernel RBF digunakan karena mampu mengenali pola data yang sangat kompleks dan tidak memiliki hubungan yang jelas antar fitur. Pemilihan kedua kernel ini didasarkan pada kemampuannya dalam meningkatkan margin pemisah SVM, sehingga model dapat menghasilkan akurasi yang lebih optimal dalam klasifikasi.

Beberapa penelitian terdahulu juga menunjukkan keberhasilan penerapan metode SVM. Misalnya, penelitian oleh (Ghaniaviyanto Ramadhan & Khoirunnisa, 2021) menggunakan metode Support Vector Machine (SVM) dalam mendeteksi dan mengklasifikasikan penyakit malaria. Hasil penelitian tersebut menunjukkan bahwa model SVM memiliki akurasi tertinggi dibandingkan dengan model Naïve Bayes. Selain itu, penelitian oleh (Ayu Wijayanti et al., 2018) menerapkan Support Vector

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

Machine (SVM) untuk klasifikasi tingkat risiko pasien gagal ginjal, dengan hasil akurasi terbaik mencapai 83,998%, sedangkan akurasi tertinggi sebesar 98,33%.

Polycystic Ovary Syndrome (PCOS) adalah kondisi hormonal umum yang terjadi pada wanita usia reproduksi. Penyakit ini ditandai dengan ketidakseimbangan hormon yang mempengaruhi proses ovulasi, dan menyebabkan gangguan menstruasi yang tidak teratur, pertumbuhan kista, serta berbagai masalah metabolisme. Kondisi ini dapat dimulai sejak masa remaja, dan gejalanya bisa berubah seiring waktu. PCOS memengaruhi sekitar 8-13% wanita usia reproduksi, dan hingga 70% kasus tidak terdiagnosis. Gejala umum dari PCOS meliputi tanda-tanda kelebihan hormon androgen seperti rambut wajah atau tubuh yang berlebihan, rambut rontok di kepala, jerawat serta gejala periode menstruasi tidak teratur atau tidak ada (Organization Health, 2023).

Penelitian terkait klasifikasi PCOS juga telah dilakukan sebelumnya, seperti oleh (Budi Prayuga et al., 2017) yang menggunakan metode Principal Component Analysis (PCA) dan metode Naïve Bayes untuk membantu mendeteksi Tingkat kesuburan wanita yang mengalami PCOS.

Berdasarkan uraian tersebut, maka peneliti tertarik untuk melakukan penelitian klasifikasi penyakit Polycystic Ovary Syndrome (PCOS) menggunakan metode Support Vector Machine (SVM), karena metode ini dinilai mampu memberikan akurasi tertinggi serta efektif dalam menangani data dengan pola non-linier.

METODE

Jenis Penelitian

Jenis penelitian yang digunakan adalah penelitian terapan, artinya penelitian menggunakan metode yang sudah ada. Bantuan software yang digunakan adalah Microsoft Excel dan Phyton.

Sumber Data

Sumber data pada penelitian ini adalah data didapatkan dari beberapa Rumah Sakit India pada penyakit Polycystic Ovary Syndrome (PCOS). Data dapat diperoleh dari website resmi <https://www.kaggle.com/> yang dipublikasikan oleh Prasoon Kottarathil pada tahun 2020.

Populasi dan Sampel Penelitian

Populasi dalam penelitian merupakan Kumpulan data yang akan digunakan dalam suatu penelitian. Dalam penelitian ini, populasi yang digunakan yaitu dataset numerik yang berupa dataset pasien penyakit Polycystic Ovary Syndrome (PCOS). Sampel dalam penelitian merupakan bagian dari populasi yang diambil untuk dianalisis dalam suatu penelitian. Dataset dalam penelitian sebanyak 541 sampel, dengan dataset dibagi 70% training dataset dan 30% testing dataset.

HASIL PEMBAHASAN

Deskripsi Data

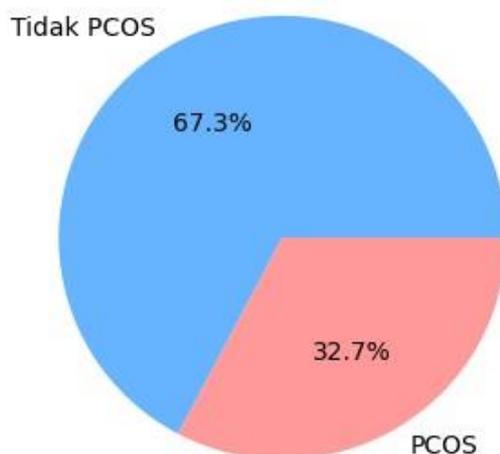
Penelitian ini menggunakan data pasien *Polycystic Ovary Syndrome* (PCOS) dengan jumlah sampel sebanyak 541 pasien. Data tersebut dapat diperoleh melalui *website Kaggle* dan dipublikasikan oleh Prasoon Kottarathil pada tahun 2020. Variabel dependen (Y) pada penelitian ini adalah pasien yang terkena penyakit *Polycystic Ovary Syndrome* ($Y = 1$), dan pasien yang tidak terkena penyakit *Polycystic Ovary Syndrome* ($Y = 0$). Pada Tabel 4.1 akan menunjukkan data deskriptif dari masing-masing variabel dependen (Y) dan variabel independen (X) yang digunakan pada penelitian ini.

Tabel 1 Variabel Dependen (Y) dan Variabel Independen (X)

	Variabel	Skala
Y	Status Penyakit	Nominal
X ₁	Usia	Kontinu
X ₂	Indeks Massa Tubuh	Kontinu
X ₃	Diabetes	Kontinu
X ₄	<i>Follicle-Stimulating Hormone</i>	Kontinu
X ₅	<i>Luteinizing Hormone</i>	Kontinu
X ₆	<i>Thyroid Stimulating Hormone</i>	Kontinu
X ₇	<i>Anti-Mullerian Hormone</i>	Kontinu
X ₈	Vitamin D3	Kontinu

Pemeriksaan Keseimbangan Data

Penelitian melakukan pemeriksaan keseimbangan data untuk mengetahui distribusi jumlah data pada masing-masing kelas dalam dataset. Pemeriksaan ini dilakukan dengan menghitung jumlah data untuk setiap label kelas, baik kelas 0 (Tidak PCOS) maupun kelas 1 (PCOS).



Gambar 1 *Pie Chart Polycystic Ovary Syndrome*

Berdasarkan Gambar 1 di atas, dapat disimpulkan bahwa dataset yang digunakan termasuk ke dalam kategori data tidak seimbang (*imbalanced*), karena persentase pasien penyakit PCOS yaitu 32,7% atau berjumlah 177 pasien. Sedangkan persentase pasien yang tidak terkena PCOS yaitu 67,3% atau berjumlah 364 pasien.

Metode Support Vector Machine

Metode analisis yang digunakan dalam penelitian adalah metode *Support Vector Machine* (SVM) dalam proses klasifikasi. Beberapa tahapan dalam memulai proses analisis yaitu melakukan pembagian data, melatih model, memprediksi model, hingga evaluasi performa berdasarkan akurasi masing-masing kernel, antara lain Kernel Polinomial dan Kernel *Radial Basis Function* (RBF) yang akan dibahas.

Peneliti membagi dataset menjadi dua bagian, yaitu *training* dataset sebanyak 70% dan *testing* dataset sebanyak 30% dari total keseluruhan data menggunakan *train_test_split* (kode dapat dilihat pada lampiran 3). Proses ini bertujuan untuk mempersiapkan data yang akan digunakan dalam melatih model dan menguji performa

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

model, maka didapat jumlah pada *training* dataset yaitu 378 sampel dan jumlah pada *testing* dataset yaitu 163 sampel dari seluruh total data sebanyak 541 sampel.

Estimasi Parameter SVM dengan Kernel Polinomial

Kernel Polinomial digunakan karena kemampuannya dalam menangani data non-linier dengan melakukan transformasi ke ruang fitur berdimensi lebih tinggi. Penelitian melakukan proses estimasi parameter untuk model Support Vector Machine (SVM) menggunakan kernel polinomial. Tujuan dari estimasi adalah untuk menentukan kombinasi parameter terbaik yang menghasilkan akurasi tertinggi pada training dataset.

Pada penelitian ini, nilai parameter C sebagai parameter regulasi dioptimalkan secara otomatis menggunakan metode Grid Search dengan validasi silang (Cross Validation). Tujuan dari optimasi ini adalah untuk memperoleh nilai C yang memberikan kinerja klasifikasi terbaik dalam memisahkan kelas data pada data Polycystic Ovary Syndrome (PCOS). GridSearchCV dilakukan terhadap beberapa kombinasi nilai C bersamaan dengan parameter kernel polinomial seperti degree (derajat polinomial), dan coef0 (konstanta penyesuaian dalam fungsi kernel).

Untuk mengatasi ketidakseimbangan kelas dalam dataset, penelitian ini menerapkan `class_weight='balanced'` pada model SVM (lihat Lampiran 4 untuk detail kode program). Selain itu, untuk memastikan bahwa model prediksi dapat memberikan klasifikasi yang seimbang antara setiap kelas, maka digunakan metrik f1-score (`scoring='f1'`) dalam proses pemilihan parameter terbaik.

Pendekatan ini melakukan seluruh kombinasi nilai parameter dalam proses pengujian. Kombinasi parameter yang diuji dalam proses ini meliputi:

- C= [0.1,1,10]
 - C=0.1 digunakan untuk menguji model dengan regularisasi tinggi.
 - C=1 digunakan sebagai nilai tengah umum yang banyak digunakan.
 - C=10 digunakan untuk menguji model dengan regularisasi rendah.
- degree = [2,3]
 - degree=2 dipilih untuk membentuk hubungan non-linier sederhana antar fitur.
 - degree=3 dipilih untuk melihat pengaruh peningkatan kompleksitas model terhadap performa. Degree lebih dari 3 tidak digunakan agar menghindari risiko overfitting.
- Coef0=[0,1]
 - Coef0=0 digunakan untuk melihat performa model tanpa penambahan bias.
 - Coef0=1 digunakan untuk menguji pengaruh penambahan bias terhadap pemisahan data non-linier.

Total kombinasi parameter yang diuji adalah $3 \times 2 \times 2 = 12$ model berbeda. Setiap kombinasi nilai parameter tersebut akan diuji menggunakan metode 5-fold cross-validation pada training dataset, yang artinya data latih akan dibagi menjadi lima bagian dan proses pelatihan serta pengujian dilakukan sebanyak lima kali secara bergiliran.

Setelah seluruh kombinasi diuji, GridSearchCV memilih kombinasi parameter dengan nilai rata-rata f1-score tertinggi pada proses cross-validation. Dalam penelitian ini, hasil f1-score dari setiap kombinasi parameter yang diperoleh adalah:

Tabel 2 Hasil f1-score (CV) dari Kombinasi Parameter Kernel Polinomial

Degree	C	Coef0	Rata-rata F1-score (CV)
2	0.1	1	0.499062
2	1	1	0.494696
2	10	1	0.494085
3	0.1	1	0.472063
3	1	1	0.446857

3	10	1	0.423522
3	10	0	0.325270
2	10	0	0.312393
3	1	0	0.277464
2	1	0	0.246036
3	0.1	0	0.127167
2	0.1	0	0.074685

Dari Tabel 4.2, Kombinasi parameter terbaik diperoleh pada nilai degree =2, C=0.1, dan coef0=1, dengan rata-rata f1-score sebesar 0.499(≈50%) karena kombinasi ini menunjukkan performa terbaik dalam memisahkan kelas pada training dataset yang tidak seimbang, artinya mampu membentuk hyperplane yang cukup fleksibel dan efektif dalam memodelkan data yang bersifat non-linier.

Pembentukan Model dan Hyperplane pada SVM Kernel Polinomial

Setelah dilakukan pelatihan dengan parameter terbaik, yaitu degree =2, C=0.1 dan coef0=1. Tahap selanjutnya adalah membentuk model fungsi keputusan f(x) untuk proses klasifikasi dan menentukan hyperplane SVM yang mencakup pemilihan fitur dan visualisasi hyperplane.

Penentuan Support Vector (x_i), Koefisien Dual (α_i), dan Bias (b)

Pada tahap ini, penelitian ini menentukan elemen-elemen penting yang digunakan dalam proses klasifikasi dengan SVM kernel polinomial, yaitu support vector (x_i), koefisien dual (α_i), dan bias (b). Support vector adalah sejumlah data latih yang berperan dalam menentukan posisi hyperplane. Dalam penelitian ini, support vector diperoleh secara otomatis dari proses training model SVM. Jumlah support vector dapat ditampilkan melalui atribut model support_vectors_. Berdasarkan hasil pelatihan, sistem menghasilkan total 317 support vector yang terdiri atas:

- 107 support vector dari kelas positif (PCOS).
- 210 support vector dari kelas negatif (Tidak PCOS).

Setiap support vector memiliki nilai koefisien dual yang digunakan dalam fungsi keputusan untuk menentukan kelas dari suatu data. Jika nilai support vector dengan koefisien dual yang bernilai nol secara otomatis dieliminasi selama proses training, karena tidak berpengaruh terhadap fungsi keputusan dan tidak dianggap sebagai support vector. Nilai koefisien dual dan label kelas dari setiap support vector dapat diperoleh melalui atribut model dual_coef_ dan support_labels.

Setelah support vector dan koefisien dual diperoleh, SVM menghitung nilai bias sebagai konstanta dalam persamaan hyperplane. Nilai bias ini diperoleh melalui atribut model intercept_ dari model SVM yang telah dilatih. Dalam penelitian ini, nilai bias yang diperoleh adalah:

$$b = -0.331588$$

Pembentukan Fungsi Keputusan f(x)

Setelah diperoleh nilai bias sebesar b = -0.331588, serta memperoleh support vector, koefisien dual, dan label kelas dari hasil pelatihan model, maka tahap selanjutnya yaitu menghitung nilai fungsi keputusan f(x). Perhitungan fungsi keputusan bertujuan untuk mengetahui jauhnya jarak data training dan testing terhadap hyperplane pada ruang fitur hasil transformasi kernel. Fungsi keputusan f(x) dalam metode SVM kernel dituliskan sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b$$

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

Pada penelitian ini, digunakan fungsi kernel polinomial dengan parameter degree = 2 dan coef0 = 1. Berdasarkan parameter tersebut, fungsi kernel Polinomial $K(x_i, x)$ diperoleh:

$$K(x, x_i) = (x, x_i + 1)^2$$

Dengan demikian, bentuk akhir dari fungsi keputusan $f(x)$ yang digunakan dalam model ini adalah sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i (x_i, x + 1)^2 - 0.331588$$

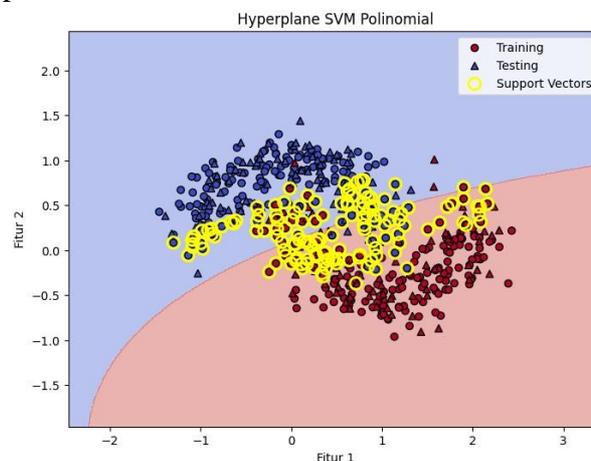
Pemilihan Fitur Terbaik

Untuk menampilkan visualisasi hyperplane dari model Support Vector Machine (SVM) dengan kernel polinomial, dilakukan terlebih dahulu pemilihan dua fitur utama dengan menggunakan metode SelectKBest. SelectKBest adalah metode seleksi fitur yang memilih sejumlah fitur terbaik berdasarkan skor terhadap variabel target. Tujuan dari pemilihan dua fitur ini untuk menyederhanakan dimensi data sehingga dapat divisualisasikan dalam ruang dua dimensi. Berdasarkan proses seleksi tersebut, diperoleh dua fitur terbaik yaitu:

- Fitur 1: Indeks Massa Tubuh
- Fitur 2: Anti-Mullerian Hormone (AMH)

Visualisasi Hyperplane SVM pada Kernel Polinomial

Setelah fitur terpilih, dilakukan pelatihan ulang model SVM dengan kernel polinomial menggunakan dua fitur tersebut, sehingga diperoleh visualisasi hyperplane seperti ditunjukkan pada Gambar 4.2:



Gambar 2 *Hyperplane SVM Polinomial*

Pada Gambar 2 tersebut memperlihatkan hasil visualisasi pemisahan kelas oleh model Support Vector Machine (SVM) dengan kernel polinomial menggunakan dua fitur utama yaitu Fitur 1 (Indeks Massa Tubuh) dan Fitur 2 (AMH). Model SVM membentuk data ke dalam ruang berdimensi lebih tinggi dengan menggunakan kernel polinomial berparameter terbaik yaitu degree = 2, Coef0 = 1, C=0.1. Warna latar biru pada grafik menunjukkan wilayah keputusan untuk kelas negatif (Tidak PCOS), sedangkan warna latar merah muda menunjukkan wilayah keputusan untuk kelas positif (PCOS). Titik-titik bulat berwarna merah menunjukkan data training, sedangkan simbol segitiga berwarna biru menunjukkan data testing.

Beberapa titik pada gambar dilingkari dengan warna kuning, yang menandakan support vector yaitu data-data yang paling dekat dengan hyperplane dan secara langsung mempengaruhi pembentukan margin pemisah. Hyperplane yang terbentuk terlihat dalam bentuk garis lengkung yang memisahkan dua wilayah klasifikasi secara non-linier. Bentuk lengkung ini mencerminkan kemampuan kernel polinomial dalam menyediakan pola data yang tidak dapat dipisahkan secara linier. Dengan demikian, visualisasi ini memberikan pemahaman menyeluruh tentang bagaimana model SVM dengan kernel polinomial memisah dua kelas berdasarkan fitur terpilih dan bagaimana kontribusi support vector mempengaruhi struktur model.

Estimasi Parameter dengan SVM Kernel RBF

Kernel Radial Basis Function (RBF) merupakan salah satu jenis kernel yang paling umum digunakan dalam algoritma Support Vector Machine (SVM), untuk mengatasi permasalahan klasifikasi yang bersifat non-linier. Estimasi parameter dalam SVM dengan kernel RBF melibatkan dua parameter utama yang diperlukan selama melatih model yaitu C (regularization parameter) dan γ (gamma). Kombinasi yang tepat dari kedua parameter sangat penting dalam mempengaruhi performa model dalam melakukan klasifikasi terhadap data yang tidak terpisah secara linier.

Sama seperti pada penggunaan kernel polinomial, penelitian ini menggunakan metode Grid Search Cross Validation (GridSearchCV) untuk melakukan estimasi parameter. Teknik ini mencoba berbagai kombinasi nilai parameter yang kemudian diuji untuk memperoleh parameter terbaik. Dalam proses ini, digunakan `class_weight='balanced'` untuk mengatasi masalah ketidakseimbangan kelas pada data, serta `scoring='f1'` sebagai proses pemilihan parameter terbaik.

Adapun kombinasi parameter yang diuji dalam proses Grid Search adalah sebagai berikut:

- $C = [0.01, 0.1, 1, 10, 100]$
 - $C=0.01$ atau 0.1 digunakan agar model toleran terhadap kesalahan pada margin yang lebih besar.
 - $C=10$ atau 100 digunakan untuk menghasilkan margin yang lebih sempit dan berusaha memisahkan data seakurat mungkin.
- $\gamma = [0.001, 0.01, 0.1, 1]$
 - $\gamma=0.001$ atau 0.01 digunakan untuk menghasilkan model yang lebih umum.
 - $\gamma=0.1$ atau 1 digunakan agar model lebih sensitif terhadap data di sekitarnya.

Total kombinasi yang dihasilkan adalah $5 \times 4 = 20$ kombinasi model yang berbeda. Untuk setiap kombinasi parameter, training dataset dibagi menjadi lima bagian yang (5-fold cross-validation) artinya bahwa setiap kombinasi, proses dilatih dan diuji sebanyak lima kali. Dari setiap fold diperoleh nilai $f1$ -score, kemudian dihitung rata-ratanya untuk setiap kombinasi parameter. Proses ini dilakukan secara otomatis oleh fungsi GridSearchCV. Setelah seluruh kombinasi parameter diuji, GridSearchCV akan memilih kombinasi parameter dengan rata-rata $f1$ -score tertinggi sebagai parameter terbaik. Hasil dari pengujian ini disajikan pada Tabel 4.3 berikut:

Tabel 3 Hasil $f1$ -score (CV) dari Kombinasi Parameter Kernel RBF

C	γ	Rata-rata F1-score (CV)
1	0.01	0.505024
10	0.001	0.503917
0.10	0.1	0.501063
1	0.1	0.485145
0.10	1	0.485010

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

100	0.001	0.479496
10	0.01	0.471304
100	0.01	0.469802
100	0.1	0.460338
0.10	0.01	0.448513
10	0.1	0.433818
100	1	0.428271
1	1	0.427803
10	1	0.414595
0.10	0.001	0.397030
0.01	0.001	0.297030
0.01	1	0.297030
0.01	0.1	0.297030
0.01	0.01	0.297030
1	0.001	0.136742

Dari Tabel 3 dapat dilihat bahwa kombinasi parameter terbaik diperoleh pada nilai $C=1$ dan $\gamma=0.01$, dengan nilai rata-rata f1-score sebesar 0.505 atau setara dengan 51%. Dengan kombinasi parameter ini, model SVM diharapkan mampu membentuk hyperplane yang cukup fleksibel untuk memisahkan kelas pada data non-linier.

Pembentukan Model dan Hyperplane pada SVM Kernel RBF

Pada tahap ini, penelitian membentuk model klasifikasi menggunakan metode SVM dengan fungsi keputusan $f(x)$ kernel RBF dan visualisasi hyperplane pada kernel RBF. Model dibentuk menggunakan parameter terbaik yang diperoleh dari proses pelatihan yaitu $C=1$ dan $\gamma=0.01$.

Penentuan Support Vector (x_i), Koefisien Dual (α_i), dan Bias (b)

Pada tahap ini, penelitian menentukan elemen-elemen penting dalam proses klasifikasi menggunakan model SVM dengan kernel Radial Basis Function (RBF) yaitu support vector (x_i), koefisien dual (α_i), dan bias (b). Support vector diperoleh menggunakan model `support_vectors_` dan jumlah support vector yang dihasilkan adalah 322 yang terdiri dari:

- 107 Support Vector dari kelas positif (PCOS).
- 215 Support Vector dari kelas negatif (Tidak PCOS).

Setiap support vector memiliki nilai koefisien dual dan label kelas yang digunakan dalam pembentukan fungsi keputusan $f(x)$, di mana nilainya diperoleh melalui atribut model `dual_coef_` dan `support_labels`.

Selain itu, model juga menghasilkan nilai bias sebagai konstanta pada fungsi keputusan $f(x)$ yang diperoleh melalui atribut model `intercept_`. Maka, nilai bias yang diperoleh yaitu:

$$b = 0.486978$$

Pembentukan Fungsi Keputusan $f(x)$

Setelah memperoleh support vector, koefisien dual, dan nilai bias dari proses pelatihan model SVM dengan kernel Radial Basis Function (RBF), maka penelitian ini membentuk fungsi keputusan $f(x)$ untuk menentukan kelas dari setiap data. fungsi keputusan $f(x)$ pada kernel RBF dihitung berdasarkan kontribusi seluruh support vector terhadap suatu titik data dependen (X) dengan memperhatikan nilai kernel RBF di antara support vector dan titik tersebut. Dengan fungsi kernel $f(x)$ dalam metode SVM kernel sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i K(x, x_i) + b$$

Model SVM kernel RBF menggunakan parameter $\gamma=0.01$, maka fungsi kernel dengan parameter sebagai berikut:

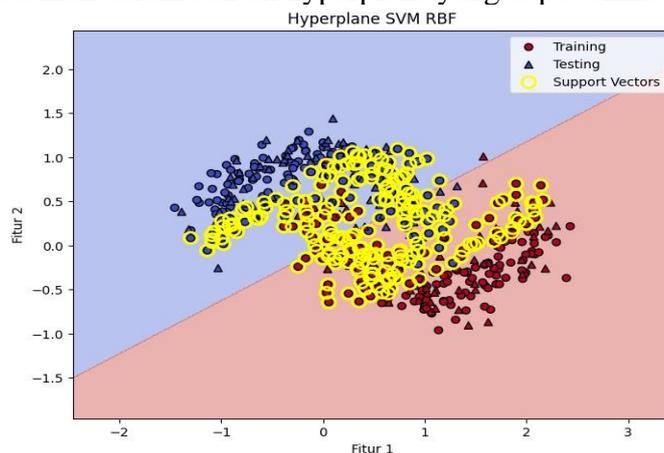
$$K(x, x_i) = \exp(-0.01\|x_i - x\|^2)$$

Dengan demikian, model dari fungsi keputusan $f(x)$ yang diperoleh yaitu:

$$f(x) = \sum_{i=1}^n \alpha_i y_i (\exp(-0.01\|x_i - x\|^2)) + 0.486978$$

Visualisasi Hyperplane SVM pada Kernel RBF

Untuk menampilkan visualisasi hyperplane dari model SVM dengan kernel RBF, penelitian dilakukan dengan pemilihan dua fitur utama yaitu Indeks Massa Tubuh dan AMH. Pemilihan ini dilakukan untuk menyederhanakan menjadi dua dimensi agar dapat divisualisasikan secara grafis. Selanjutnya, model SVM dilatih ulang menggunakan parameter terbaik yang telah diperoleh sebelumnya yaitu $C=1$ dan $\gamma=0.01$. Kemudian hasilnya divisualisasikan dalam bentuk hyperplane yang dapat dilihat pada Gambar 3:



Gambar 3 Hyperplane SVM RBF

Gambar 3 memperlihatkan visualisasi hyperplane yang dibentuk oleh model SVM dengan kernel RBF. Model SVM memisahkan dua kelas menggunakan garis keputusan (decision boundary) non-linier yang membagi area ruang fitur menjadi dua wilayah berwarna berbeda yaitu merah muda dan biru muda. Warna merah muda menunjukkan wilayah prediksi untuk kelas negatif (label = 0), sedangkan warna biru muda menunjukkan wilayah prediksi untuk kelas positif (label = 1). Data training divisualisasikan dalam bentuk titik bulat, sedangkan data testing divisualisasikan dalam bentuk segitiga. Support vector yang berperan penting dalam pembentukan hyperplane ditandai dengan lingkaran kuning yang mengelilingi titik-titik data. Dari visualisasi ini dapat diamati bahwa sebagian besar support vector terletak di dekat batas antara dua kelas, yang mencerminkan perannya dalam menentukan bentuk dan posisi hyperplane.

Prediksi Kelas dan Evaluasi Model (Confusion Matrix dan Akurasi)

Penelitian ini melakukan evaluasi terhadap model Support Vector Machine (SVM) dengan kernel polinomial dan kernel Radial Basis Function (RBF) menggunakan data training dan data testing. Tujuan evaluasi ini adalah untuk menilai

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

kemampuan model dalam mengklasifikasikan data secara akurat setelah proses pelatihan dilakukan menggunakan parameter terbaik.

Setelah model SVM untuk kedua kernel telah dilatih, model digunakan untuk memprediksi kelas pada data training dan data testing yang telah melalui proses normalisasi. Untuk menilai kualitas dari hasil prediksi tersebut, penelitian ini menggunakan confusion matrix sebagai alat evaluasi utama.

Training Data

a. Kernel Polinomial

Setelah model fungsi keputusan $f(x)$ diperoleh dengan bentuk sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i y_i (x_i \cdot 1)^2 - 0.33158804$$

Maka Langkah selanjutnya adalah menentukan kelas dari setiap data berdasarkan tanda dari nilai $f(x)$ tersebut. Proses klasifikasi ini dilakukan dengan menggunakan fungsi tanda (sign), yaitu:

- “Jika $f(x) > 0$, maka prediksi kelas = +1 (PCOS)”
- “Jika $f(x) < 0$, maka prediksi kelas = -1 (Tidak PCOS)”

Penentuan tanda (sign) ini diterapkan pada seluruh data training, sehingga diperoleh prediksi kelas untuk masing-masing sampel.

Berdasarkan hasil prediksi terhadap data training, penelitian ini memperoleh confusion matrix dari model SVM dengan kernel polinomial, sebagaimana ditunjukkan pada Tabel 4.:

Tabel 4. Confusion Matrix pada Kernel Polinomial (Data Training)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	202	52
Positif (+1)	59	65

Dari Tabel 4. tersebut, diketahui bahwa jumlah prediksi benar untuk kelas positif (TP) adalah 65, sedangkan untuk kelas negatif (TN) sebanyak 202. Sebaliknya, terdapat 52 kasus negatif yang diprediksi sebagai positif (FP), dan 59 kasus positif yang salah diprediksi sebagai negatif (FN). Berdasarkan nilai-nilai tersebut, dihitung akurasi dan metrik evaluasi lainnya seperti precision, recall, dan f1-score.

- Akurasi = $(TP+TN)/(TP+TN+FN+FP)$
 $Akurasi = (65+202)/(65+202+59+52)$
 $Akurasi = 267/378$
 $Akurasi = 0,706349 \approx 71\%$
- Precision = $TP/(TP+FP)$
 $Precision = 65/(65+52)$
 $Precision = 65/117$
 $Precision = 0,555555 \approx 56\%$
- Recall = $TP/(TP+FN)$
 $Recall = 65/(65+59)$
 $Recall = 65/124$
 $Recall = 0,524193 \approx 52\%$
- f1-score = $(2 \times Precision \times Recall)/(Precision + Recall)$
 $f1\text{-score} = (2 \times 0,56 \times 0,52)/(0,56 + 0,52)$

$$f1\text{-score}=0,5824/1,08$$

$$f1\text{-score}=0,539259 \approx 54\%$$

Model SVM dengan kernel polinomial mampu mengklasifikasikan data training dengan akurasi sebesar 71% yang menunjukkan bahwa sebagian besar prediksi model sesuai dengan label asli pada data pelatihan. Precision sebesar 56% menunjukkan bahwa lebih dari setengah prediksi positif yang diberikan model adalah benar, meskipun masih ada sejumlah kesalahan. Recall sebesar 52% menunjukkan bahwa model cukup mampu mengenali sebagian besar data yang benar-benar positif (PCOS). F1-score sebesar 54% menunjukkan keseimbangan yang relatif baik antara precision dan recall, sehingga model dinilai memiliki performa yang stabil pada data training.

b. Kernel RBF

Setelah penelitian memperoleh bentuk model fungsi keputusan $f(x)$ pada SVM kernel RBF yaitu:

$$f(x) = \sum_{i=1}^n \alpha_i y_i (\exp(-0.01 \|x_i - x\|^2)) + 0.486978$$

Tahap selanjutnya adalah menentukan tanda (sign) dari nilai $f(x)$ untuk setiap data training. Penentuan tanda (sign) bertujuan untuk mengklasifikasikan data ke dalam kelas target, dengan aturan sebagai berikut:

- Jika $f(x) > 0$, maka data diprediksi sebagai kelas +1 (PCOS).
- Jika $f(x) < 0$, maka data diprediksi sebagai kelas -1 (Tidak PCOS).

Sehingga diperoleh prediksi kelas untuk masing-masing data. Setelah prediksi kelas diperoleh, penelitian melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan confusion matrix seperti yang akan ditampilkan pada Tabel 4.5 berikut ini:

Tabel 5. Confusion Matrix pada Kernel RBF (Data Training)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	195	59
Positif (+1)	59	65

Berdasarkan Tabel 5, diketahui bahwa diperoleh jumlah prediksi benar untuk kelas negatif (TN) adalah sebanyak 195, sedangkan jumlah prediksi benar untuk kelas positif (TP) adalah sebanyak 65. Sebaliknya, terdapat 59 data negatif yang salah diprediksi sebagai positif (FP) dan 59 data positif yang salah diprediksi sebagai negatif (FN). Dengan diperoleh nilai TN, TP, FP dan FN tersebut, maka dapat dilakukan perhitungan metrik evaluasi lainnya.

- Akurasi = $(TP+TN)/(TP+TN+FN+FP)$
 Akurasi = $(65+195)/(65+195+59+59)$
 Akurasi = $260/378$
 Akurasi = $0,687830 \approx 69\%$
- Precision = $TP/(TP+FP)$
 Precision = $65/(65+59)$
 Precision = $65/124$
 Precision = $0,524193 \approx 52\%$
- Recall = $TP/(TP+FN)$
 Recall = $65/(65+59)$
 Recall = $65/124$

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

$$\text{Recall}=0,524193\approx 52\%$$

- $f1\text{-score}=(2 \times \text{Precision} \times \text{Recall})/(\text{Precision}+\text{Recall})$
 $f1\text{-score}=(2 \times 0,52 \times 0,52)/(0,52+0,52)$
 $f1\text{-score}=0,5408/1,04$
 $f1\text{-score}=0,52\approx 52\%$

Model SVM dengan kernel RBF menunjukkan kinerja klasifikasi yang cukup baik terhadap data training. Model menghasilkan akurasi sebesar 69% yang berarti sebagian besar prediksi model berhasil diklasifikasikan dengan benar. Selain itu, model mencapai precision sebesar 52% yang menunjukkan bahwa sekitar setengah dari prediksi positif yang dibuat model adalah benar, yang mencerminkan tingkat ketepatan model dalam mengidentifikasi kelas positif. Recall yang diperoleh sebesar 52% menunjukkan bahwa model berhasil mengenali setengah dari seluruh kasus positif yang sebenarnya, sehingga sensitivitas model tergolong sedang. Sebagai hasil akhirnya, f1-score sebesar 52% menggambarkan keseimbangan antara precision dan recall, yang menunjukkan bahwa performa model pada data training cukup stabil namun masih memiliki ruang untuk ditingkatkan.

c. Regresi Logistik

Setelah dilakukan pemodelan menggunakan Support Vector Machine (SVM) dengan kernel polinomial dan RBF, evaluasi model regresi logistik dilakukan untuk membandingkan performa kedua metode tersebut. Proses evaluasi ini melibatkan pengukuran akurasi, precision, recall, dan f1-score yang diterapkan pada training data untuk mengukur sejauh mana model dapat memprediksi hasil dengan baik. Hasil evaluasi ini menghasilkan confusion matrix pada Tabel 4.6 sebagai berikut:

Tabel 6 Confusion Matrix pada Regresi Logistik (Data Training)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	235	19
Positif (+1)	90	34

Pada Tabel 6 ditampilkan confusion matrix hasil evaluasi model regresi logistik pada training dataset. Berdasarkan tabel tersebut, diketahui bahwa jumlah data yang benar-benar negatif dan diprediksi sebagai negatif (TN) sebanyak 235. Selain itu, jumlah data yang benar-benar positif dan diprediksi sebagai positif adalah sebanyak 34. Sementara itu, jumlah data yang sebenarnya negatif tetapi diprediksi sebagai positif (FP) tercatat sebanyak 19, dan jumlah data yang sebenarnya positif tetapi diprediksi sebagai negatif (FN) sebanyak 90. Berdasarkan nilai-nilai ini, maka dapat dilakukan perhitungan metrik evaluasi model seperti akurasi, precision, recall, dan f1-score untuk menilai performa model secara menyeluruh.

- $\text{Akurasi}=(\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FN}+\text{FP})$
 $\text{Akurasi}=(34+235)/(34+235+90+19)$
 $\text{Akurasi}=269/378$
 $\text{Akurasi}=0,711640\approx 71\%$
- $\text{Precision}=\text{TP}/(\text{TP}+\text{FP})$
 $\text{Precision}=34/(34+19)$
 $\text{Precision}=34/53$
 $\text{Precision}=0,641509\approx 64\%$

- $Recall = TP / (TP + FN)$
 $Recall = 34 / (34 + 90)$
 $Recall = 34 / 124$
 $Recall = 0,274193 \approx 27\%$
- $f1\text{-score} = (2 \times Precision \times Recall) / (Precision + Recall)$
 $f1\text{-score} = (2 \times 0,64 \times 0,27) / (0,64 + 0,27)$
 $f1\text{-score} = 0,3456 / 0,91$
 $f1\text{-score} = 0,379780 \approx 38\%$

Model regresi logistik berhasil mengklasifikasikan data training dengan akurasi sebesar 71% yang menunjukkan bahwa sebagian besar prediksi model sesuai dengan label sebenarnya dalam data training. Precision sebesar 64% menunjukkan bahwa mayoritas prediksi positif yang dilakukan oleh model adalah benar, sehingga model cukup baik dalam menghindari kesalahan prediksi positif. Recall sebesar 27% bahwa model hanya mampu mengenali sebagian kecil dari total kasus positif yang sebenarnya, yang mencerminkan sensitivitas model masih rendah. F1-score sebesar 38% mengindikasikan bahwa keseimbangan antara precision dan recall masih kurang optimal, sehingga model cenderung precision daripada recall dalam prediksinya.

Data Testing

a. Kernel Polinomial

Sama seperti pada proses data training, model fungsi keputusan $f(x)$ yang telah dibentuk juga melakukan proses klasifikasi menggunakan fungsi tanda (sign) terhadap data testing untuk memperoleh prediksi kelas. Berdasarkan hasil prediksi tersebut, diperoleh confusion matrix yang ditunjukkan pada Tabel 7:

Tabel 7 Confusion Matrix pada Kernel Polinomial (Data Testing)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	91	19
Positif (+1)	28	25

Berdasarkan Tabel 7, diketahui bahwa jumlah prediksi untuk kelas negatif (TN) adalah sebanyak 91, sedangkan prediksi benar untuk kelas positif (TP) sebanyak 25. Sebaliknya, terdapat 19 data negatif yang salah diprediksi sebagai positif (FP) dan 28 data positif yang diprediksi sebagai negatif (FN). Dengan nilai-nilai tersebut, maka dapat dilakukan perhitungan terhadap akurasi, precision, recall, dan f1-score untuk mengetahui seberapa baik model dalam mengklasifikasikan data pada proses pengujian.

- $Akurasi = (TP + TN) / (TP + TN + FN + FP)$
- $Akurasi = (25 + 91) / (35 + 91 + 28 + 19)$
- $Akurasi = 116 / 163$
- $Akurasi = 0,711656 = 71\%$
- $Precision = TP / (TP + FP)$
- $Precision = 25 / (25 + 19)$
- $Precision = 25 / 44$
- $Precision = 0,568181 \approx 57\%$
- $Recall = TP / (TP + FN)$
- $Recall = 25 / (25 + 28)$
- $Recall = 25 / 53$
- $Recall = 0,471698 \approx 47\%$

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

- $f1\text{-score}=(2 \times \text{Precision} \times \text{Recall})/(\text{Precision}+\text{Recall})$
- $f1\text{-score}=(2 \times 0,57 \times 0,47)/(0,57+0,47)$
- $f1\text{-score}=0,5358/1,04$
- $f1\text{-score}=0,515192 \approx 52\%$

Model SVM dengan kernel polinomial berhasil menghasilkan data testing dengan akurasi sebesar 71%. Precision sebesar 57% menunjukkan bahwa lebih dari setengah prediksi positif yang dibuat oleh model adalah benar, meskipun masih terdapat kesalahan klasifikasi. Recall sebesar 47% menunjukkan bahwa model mampu mengenali hampir setengah dari seluruh total kasus positif yang sebenarnya tetapi masih ada yang terlewat. F1-score sebesar 52% menunjukkan bahwa keseimbangan sedang antara precision dan recall, sehingga performa model pada data testing tergolong cukup stabil namun masih dapat ditingkatkan.

Kernel RBF

Setelah model SVM dengan kernel RBF selesai dilatih menggunakan data training, penelitian ini menggunakan model tersebut untuk melakukan prediksi terhadap data testing. Model menghitung nilai fungsi keputusan $f(x)$ untuk setiap data testing, kemudian menentukan tanda (sign) dari nilai tersebut untuk memprediksi kelas target.

Setelah prediksi kelas diperoleh, penelitian ini mengevaluasi hasil klasifikasi model terhadap data testing dengan menggunakan confusion matrix. Maka, diperoleh confusion matrix yang ditunjukkan pada Tabel 4.8:

Tabel 8 Confusion Matrix pada Kernel RBF (Data Testing)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	90	20
Positif (+1)	26	27

Dari Tabel 8, bahwa jumlah prediksi benar untuk kelas negatif (TN) adalah 90 data, sedangkan jumlah prediksi benar untuk kelas positif (TP) adalah 27 data. Sementara itu, terdapat 20 data negatif yang salah diprediksi sebagai positif (FP) dan 26 data positif yang salah diprediksi sebagai negatif (FN). Dengan diketahui nilai TN, TP, FP dan FN tersebut, maka dilakukan perhitungan metrik evaluasi untuk menilai seberapa baik model dalam mengklasifikasikan data pada proses pengujian.

- $\text{Akurasi}=(\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FN}+\text{FP})$
- $\text{Akurasi}=(27+90)/(27+90+26+20)$
- $\text{Akurasi}=117/163$
- $\text{Akurasi}=0,717791 \approx 72\%$

- $\text{Precision}=\text{TP}/(\text{TP}+\text{FP})$
- $\text{Precision}=27/(27+20)$
- $\text{Precision}=27/47$
- $\text{Precision}=0,574468 \approx 57\%$

- $\text{Recall}=\text{TP}/(\text{TP}+\text{FN})$
- $\text{Recall}=27/(27+26)$
- $\text{Recall}=27/53$
- $\text{Recall}=0,509433 \approx 51\%$

- $f1\text{-score}=(2 \times \text{Precision} \times \text{Recall})/(\text{Precision}+\text{Recall})$
- $f1\text{-score}=(2 \times 0,57 \times 0,51)/(0,57+0,51)$
- $f1\text{-score}=0,5814/1,08$
- $f1\text{-score}=0,538333 \approx 54\%$

Model SVM dengan kernel RBF berhasil melakukan prediksi terhadap data testing dengan tingkat akurasi sebesar 72%. Nilai ini menunjukkan bahwa mayoritas prediksi model sesuai dengan label aslinya pada data testing. Precision sebesar 57% menandakan bahwa lebih dari separuh prediksi positif yang dihasilkan model adalah benar, sehingga tingkat ketepatan klasifikasinya cukup baik. Recall sebesar 51% menunjukkan bahwa model mampu mengenali lebih dari separuh kasus positif yang sebenarnya ada dalam data, yang mencerminkan sensitivitas model yang cukup seimbang. F1-score sebesar 54% mencerminkan keseimbangan antara precision dan recall, sehingga performa model pada data testing tergolong stabil dan layak untuk digunakan.

Regresi Logistik

Proses evaluasi terhadap model regresi logistik juga mencakup pengukuran metrik performa yang diterapkan pada testing dataset, dengan tujuan untuk menilai kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya. Hasil evaluasi tersebut kemudian disajikan dalam bentuk confusion matrix, sebagaimana ditampilkan pada Tabel 4.9 berikut:

Tabel 4.9 Confusion Matrix pada Regresi logistik (Data Testing)

Kelas	Prediksi Negatif (-1)	Prediksi Positif (+1)
Negatif (-1)	104	6
Positif (+1)	37	16

Pada tahap pengujian model terhadap testing dataset, diperoleh confusion matrix seperti yang ditampilkan pada Tabel 4.9. Berdasarkan tabel tersebut, dapat dijelaskan bahwa jumlah data yang diklasifikasikan dengan benar sebagai negatif (TN) adalah sebanyak 104. Selanjutnya, jumlah data yang diklasifikasikan dengan benar sebagai positif (TP) yaitu sebanyak 16. Adapun jumlah data yang sebenarnya positif namun salah diprediksi sebagai negatif (FN) adalah sebanyak 37, dan jumlah data yang sebenarnya negatif tetapi salah diprediksi sebagai positif (FP) tercatat sebanyak 6. Berdasarkan nilai-nilai tersebut, dilakukan perhitungan metrik evaluasi lebih lanjut untuk mengetahui performa model secara keseluruhan.

- $\text{Akurasi}=(\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FN}+\text{FP})$
- $\text{Akurasi}=(16+104)/(16+104+37+6)$
- $\text{Akurasi}=120/163$
- $\text{Akurasi}=0,736196 \approx 74\%$

- $\text{Precision}=\text{TP}/(\text{TP}+\text{FP})$
- $\text{Precision}=16/(16+6)$
- $\text{Precision}=16/22$
- $\text{Precision}=0,727272 \approx 73\%$

- $\text{Recall}=\text{TP}/(\text{TP}+\text{FN})$
- $\text{Recall}=16/(16+37)$
- $\text{Recall}=16/53$
- $\text{Recall}=0,301886 \approx 30\%$

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

- $f1\text{-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$
- $f1\text{-score} = (2 \times 0,73 \times 0,30) / (0,73 + 0,30)$
- $f1\text{-score} = 0,438 / 1,03$
- $f1\text{-score} = 0,425242 \approx 43\%$

Model regresi logistik mampu mengklasifikasikan data testing dengan akurasi sebesar 74% yang menunjukkan bahwa sebagian besar hasil prediksi sesuai dengan label sebenarnya pada data testing. Precision sebesar 73% menunjukkan bahwa sebagian besar prediksi positif yang dihasilkan model adalah benar, yang mencerminkan ketepatan model dalam mengidentifikasi kelas positif. Recall sebesar 30% menunjukkan bahwa model hanya berhasil mengenali sebagian kecil dari seluruh kasus positif yang sebenarnya, sehingga sensitivitas model masih tergolong rendah.

Perbandingan Kernel Polinomial, Kernel RBF, dan Regresi Logistik

Setelah dilakukan analisis terhadap performa model klasifikasi menggunakan kernel polinomial, kernel RBF dan regresi logistik, peneliti menyusun tabel yang menyajikan hasil evaluasi masing-masing model berdasarkan metrik akurasi, precision, recall, dan f1-score. Hasil yang diperoleh dari analisis ini diharapkan dapat memberikan gambaran yang jelas mengenai keunggulan dan kekurangan masing-masing model dalam konteks performa klasifikasi.

Tabel 4.10 Hasil Analisis Model SVM dan Regresi Logistik

Model	Dataset	Akurasi	Precision	Recall	F1-score
SVM (Kernel Polinomial)	Training	71%	56%	52%	54%
	Testing	71%	57%	47%	52%
SVM (Kernel RBF)	Training	69%	52%	52%	52%
	Testing	72%	57%	51%	54%
Regresi Logistik	Training	71%	64%	27%	38%
	Testing	74%	73%	30%	43%

Penelitian menganalisis performa tiga model klasifikasi yaitu SVM dengan kernel polinomial, SVM dengan kernel RBF, dan Regresi Logistik berdasarkan metrik evaluasi akurasi, precision, recall, dan f1-score pada data training dan testing. Model SVM dengan kernel polinomial menunjukkan kinerja yang konsisten, dengan akurasi sebesar 71% baik pada data training maupun testing, serta f1-score sebesar 54% dan 52% secara berturut-turut. Model SVM dengan kernel RBF memiliki f1-score yang sedikit lebih baik pada data testing (54%) dibandingkan data training (52%) serta mencatat akurasi tertinggi kedua pada data testing (72%). Sementara itu, Regresi Logistik mencapai akurasi tertinggi sebesar 74% pada data testing, namun memiliki recall dan f1-score yang relatif rendah yaitu 30% dan 43%. Secara keseluruhan, SVM dengan kernel RBF cenderung lebih seimbang dalam hal precision dan recall, sedangkan Regresi logistik unggul pada precision namun lemah dalam recall

KESIMPULAN

Air merupakan sumber daya alam yang sangat penting dan harus dijaga kelestariannya karena perannya yang vital dalam menunjang kehidupan manusia dan makhluk hidup lainnya. Penurunan kualitas dan kuantitas air menjadi salah satu masalah utama lingkungan yang perlu mendapat perhatian serius. Pengelolaan sumber daya air harus dilakukan secara terpadu dan berkelanjutan, melibatkan koordinasi antara masyarakat dan pemerintah sesuai dengan ketentuan hukum yang berlaku, seperti Undang-Undang No. 17 Tahun 2019 tentang Sumber Daya Air.

Kearifan lokal masyarakat, khususnya di Manggarai, Nusa Tenggara Timur, memainkan peran strategis dalam pelestarian sumber daya air. Ritual Barong Wae, sebagai bagian dari kearifan lokal, tidak hanya berfungsi sebagai upacara keagamaan tetapi juga sebagai strategi sosial dan ekologis untuk menjaga keberlangsungan sumber air dan keseimbangan ekosistem. Ritual ini mencerminkan hubungan harmonis antara manusia, alam, dan entitas spiritual, yang diwariskan secara turun-temurun dan menjadi landasan pengelolaan sumber daya air secara berkelanjutan.

Pelestarian lingkungan air tidak hanya dapat dilakukan melalui teknologi dan kebijakan pemerintah, tetapi juga harus didukung oleh nilai-nilai budaya, mitos, dan kepercayaan lokal yang menumbuhkan sikap arif dan bijaksana dalam menjaga lingkungan. Kesadaran masyarakat untuk bertanggung jawab terhadap kelestarian sumber air perlu terus ditumbuhkan, baik di masyarakat adat maupun masyarakat umum, agar sumber daya air tetap tersedia dalam kualitas dan kuantitas yang memadai untuk kebutuhan saat ini dan masa depan.

Dengan demikian, integrasi antara pendekatan ilmiah, kebijakan hukum, dan kearifan lokal menjadi kunci utama dalam pengelolaan dan pelestarian sumber daya air yang berkelanjutan di Manggarai dan wilayah lainnya.

DAFTAR PUSTAKA

AANasihuddin. KEARIFAN_LOKAL_DALAM_PERLINDUNGAN_DAN_PENGELOLAAN_LINGKUNGAN_HIDUP_STUDI_DI_DESA_JANGGOLAN_KEC_SUMPIUH_BAN

Perbandingan Tingkat Akurasi Antara Support Vector Machine (Svm) Menggunakan Kernel Polinomial & Kernel Radial Basis Function (Rbf) Untuk Klasifikasi Penyakit Polycystic Ovary Syndrome (Pcos).

YUMAS.v. 2 (1), 99-107, 2017 Journal Name Bina Hukum Lingkungan 2 (1), 99-107, 2017FANSERVASI DAN ER DY

AhmadMiftahudinThohari.LINGKUNGAN_TRADISI_MASYARAKAT_JAWA_DALAM_UPAYA_MENGHORMATI_ENTITAS_ALAM.June 2022Turast Jurnal Penelitian dan Pengabdian 10(1):36-49 DOI:10.15548/turast.v10i1.3966

Dewy mulianti KEARIFAN LOKAL MASYARAKAT TERHADAP SUMBER MATA AIR SEBAGAI UPAYA KONSERVASI DAN PENGELOLAAN SUMBER DAYA LINGKUNGAN.Volume 6, Nomor 3, Juni 2022 DOI: <https://doi.org/10.24970/bhl.v6i3.286>

Erna Nema Iman.KEARIFAN LOKAL DAN UPAYA PELEST A PELESTARIAN LINGKUNGAN AIR: AN LINGKUNGAN AIR:

Frederikus Iren. KEARIFAN LOKAL MASYARAKAT DALAM MENGELOLA SUMBERMATA AIR DI DESA CAONG KECAMATAN MEMPAWAH HULUKABUPATEN LANDAK. JURNAL HUTAN LESTARI (2020).Vol. 8 (1) : 120 – 135 JURNAL HUTAN LESTARI (2020)

STUDI E STUDI ETNOGRAFI M TNOGRAFI MASYARAKAT ADAT MANGGARAI, FL ANGGARAI, FLORES, NUSA TENGGARA TIMUR. Volume 13 1 Volume 13 No. 1, April 2023 1 .4-302023.

Tapung, Z. R. E. Ntelok, H. C. Darong, Kearifan Lokal dan Upaya Pelestarian 1 Paradigma Jurnal Kajian Budaya.