Terbit *online* pada laman: http://jurnal.utu.ac.id/JTI

Jurnal Teknologi Informasi

|ISSN (Online): 2829-8934|

Prediksi Tipe Kulit Wajah Menggunakan Algoritma Support Vector Machine (SVM)

Seli Karisma Putri[1], Nurtriana Hidayati[2*]
[1,2] Program Studi Sistem Informasi, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang, Indonesia
Email: [1]selikp01@gmail.com , [2]anna@usm.ac.id
* Nurtriana Hidayati (081325863292)

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima: 2 Mei 2026
Revisi: 17 September 2026
Diterbitkan: 30 Oktober 2026

Kata Kunci:
Prediksi
Radial Basis Function (RBF)
RapidMiner
Support Vector Machine (SVM)
Tipe Kulit

ABSTRAK

Kulit merupakan organ terbesar manusia yang befungsi penting. Secara umum, tipe kulit dibagi menjadi empat kategori, yaitu kulit normal, kering, berminyak, dan kombinasi. Kesalahan mengidentifikasi tipe kulit wajah secara mandiri berpotensi pada pemilihan produk perawatan kulit yang tidak sesuai, sehingga dapat memperburuk kondisi kulit. Oleh karena itu, penelitian ini bertujuan mengembangkan model prediksi tipe kulit wajah yang lebih akurat menggunakan algoritma *Support Vector Machine* (SVM), metode yang efektif dalam memprediksi data kompleks. Penelitian ini menggunakan data sekunder yang berasal dari Kaggle yang kemudian diolah dan dianalisis menggunakan *software RapidMiner*. Dalam proses pemodelan, optimasi kernel terbaik didapatkan melalui kernel *Radial Basis Function* (RBF) dengan parameter nilai *epsilon* sebesar 0.01, *cost* sebesar 1, dan *gamma* sebesar 1. Hasil evaluasi model menunjukkan performa yang sangat baik, dengan tingkat akurasi mencapai 98.20%, presisi sebesar 98.19%, *recall* sebesar 98.22%, serta nilai *Root Mean Square Error* (RMSE) yang rendah yaitu 12.4%. Dari 2.000 data yang dievaluasi, terdapat 36 data yang tidak sesuai dengan label asli Ini membuktikan bahwa model SVM dengan kernel RBF yang dikembangkan mampu melakukan prediksi tipe kulit wajah dengan tingkat keakuratan yang tinggi.

Copyright © 2026 Jurnal Teknologi Informasi UTU
All rights reserved

1. Pendahuluan

Kemajuan teknologi telah membawa perubahan signifikan dalam industri kecantikan, terutama dalam bidang perawatan kulit [1]. Melalui penggunaan peralatan canggih untuk menganalisis kondisi kulit, seperti alat pengukur kelembapan, elastisitas kulit, tingkat hidrasi, dan kadar minyak, dapat memudahkan dalam menentukan strategi perawatan yang paling efektif.

Manusia memiliki tipe kulit wajah bervariasi yang dipengaruhi oleh faktor genetik dan lingkungan [2]. Ada empat tipe kulit yaitu kulit normal, kering, berminyak, dan kombinasi [3]. Kulit kering adalah kondisi di mana kulit cenderung terasa kering, kasar, dan tertarik. Kulit berminyak adalah kondisi kulit terlihat mengkilap, licin, dan memiliki pori- pori yang besar. Kulit normal adalah kondisi peralihan di antara kulit berminyak dan kering yang seimbang, bersih, dan tidak sensitif. Kulit kombinasi merupakan kondisi di mana kulit terlihat berminyak pada bagian t-zone dan kering di bagian pipi.

Tren mengidentifikasi tipe kulit wajah secara mandiri sangat populer dalam beberapa tahun terakhir. Hal ini dikarenakan, biaya konsultasi dokter spesialis kulit yang relatif tinggi, adanya antrean

panjang, serta potensi kesalahan diagnosis akibat perbedaan pengalaman klinis antar dokter [4]. Kondisi tersebut mendorong individu untuk memilih metode sederhana dalam menentukan tipe kulit. Namun, hal ini sering kali mengakibatkan kesalahan dalam menentukan tipe kulit. Kesalahan dalam mengidentifikasi tipe kulit wajah secara mandiri berpotensi pada pemilihan produk perawatan kulit yang tidak sesuai, sehingga dapat memperburuk kondisi kulit. Hal ini menunjukkan pentingnya mengetahui tipe kulit sebagai dasar dalam memilih produk skincare yang tepat, mengingat beragamnya produk yang tersedia di pasaran yang diformulasikan untuk berbagai tipe kulit.

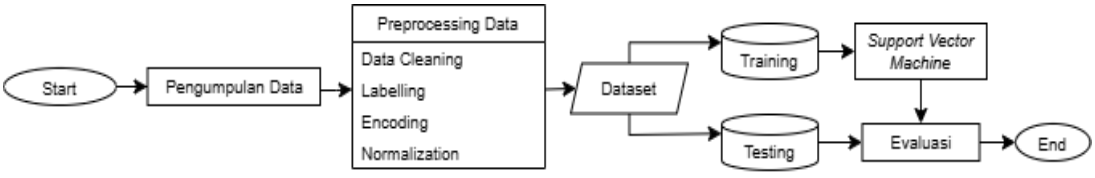
Pendekatan yang dapat digunakan untuk menyelesaikan permasalahan tersebut adalah *data mining*. *Data Mining* adalah proses menemukan pola dan tautan dalam kumpulan data besar yang sebelumnya tidak diketahui melalui teknik analisis data tingkat lanjut [5].. Dalam permasalahan identifikasi tipe kulit wajah, model diterapkan dengan memanfaatkan karakteristik kulit sebagai masukan dan tipe kulit sebagai kelas yang akan diprediksi. Salah satu algoritma *machine learning* yang dapat diterapkan adalah *Support Vector Machine* (SVM). Metode SVM dipilih karena kemampuannya dalam menemukan *hyperplane* optimal yang memisahkan kelas-kelas data dengan memaksimalkan margin antara titik data terdekat (*support vectors*) dari kelas yang berbeda [6]. SVM efektif dalam mengolah data dengan banyak variabel dan kompleksitas tinggi[7] , sehingga metode ini sangat sesuai untuk melakukan prediksi tipe kulit, seperti kulit normal, berminyak, kering, maupun kombinasi.

Penelitian oleh [8], membandingkan metode SVM dengan Naïve Bayes untuk memprediksi penyakit diabetes melitus menghasilkan akurasi SVM sebesar 78,04% sedangkan Naïve Bayes sebesar 76,98%. Pada penelitian [9], membandingkan metode SVM dengan KNN untuk memprediksi adanya seorang karyawan yang akan keluar dari perusahaan, menunjukkan bahwa algoritma SVM memiliki nilai rata-rata metrik *accuracy* 86%, sedangkan KNN nilai rata-rata metrik *accuracy* 74%. Penelitian [10], membandingkn metode SVM dengan Regresi Logistik untuk memprediksi *customer churn*, hasil prediksi SVM lebih tinggi dengan nilai *accurary* 91,56%, dibandingkan prediksi Regresi Logistik dengan nilai *accurary* 90.66%.

Tujuan dari penelitian ini adalah menerapkan metode *Support Vector Machine* (SVM) untuk memprediksi tipe kulit wajah berdasarkan karakteristik data yang tersedia. Selain itu, penelitian ini bertujuan untuk mengevaluasi performa algoritma SVM dengan mengukur metrik *accuracy*, *precision*, *recall*, dan *Root Mean Square Error* (RMSE) pada proses klasifikasi. Diharapkan model yang dihasilkan dapat membantu individu mengenali karakteristik kulit mereka secara tepat, sehingga risiko kesalahan identifikasi tipe kulit pada metode mandiri dapat diminimalkan

2. Metodologi Penelitian

Dengan alur yang disusun sistematis, penelitian dapat berjalan dengan efektif dan efisien, serta menghasilkan temuan yang valid dan reliabel. Penelitian ini diawali dengan pengumulan data, kemudian *preprocessing* data, hingga menghasilkan output. Alur penelitian disajikan pada Gambar 1. berikut



Gambar 1. Alur penelitian

Pada Gambar 1 proses penelitian dimulai dari pengumpulan data yang relevan. Data yang terkumpul kemudian melalui tahap *preprocessing* agar siap digunakan dalam analisis lebih lanjut. Setelah data siap, dilakukan pengolahan data menggunakan algoritma SVM untuk membangun model yang mampu memprediksi dengan akurat. Tahap terakhir adalah evaluasi model.

2.1. Pengumpulan Data

Penelitian ini merupakan penelitian kuantitatif yang menggunakan pengumpulan data sekunder dengan pendekatan studi pustaka. Data yang digunakan berasal dari *repository* Kaggle yang diunggah

tahun 2025. Dataset tersebut terdiri dari 2.000 data dengan 8 atribut. Tipe data yang terdapat dalam dataset penelitian ini terdiri dari data numerik dan data kategorik. Data yang diperoleh dapat dilihat pada Tabel 1.

Tabel 1. Data Asli

No	*Age*	*Gender*	*Hydration Level*	*Oil Level*	*Sensitivity*	*Humidity*	*Temperature*	*Skin Type*
1	36	*Male*	*Low*	*Low*	*High*	31.9	10.1	*Dry*
2	36	*Female*	*High*	*Low*	*High*	30.1	20.5	*Combination*
3	34	*Female*	*High*	*Medium*	*Low*	53.0	19.5	*Normal*
...	…	...	...	...	...	...	...	...
2000	27	*Male*	*High*	*Medium*	*Low*	43.8	19.5	*Normal*

Atribut usia (*age*), kelembaban (*humidity*), dan suhu (*temperature*) termasuk ke dalam data numerik karena nilainya berupa angka. Sedangkan, atribut jenis kelamin (*gender*), tingkat hidrasi (*hydration_level*), tingkat minyak (*oil_level*), sensitivitas (*sensitivity*), dan tipe kulit (*skin_type*) termasuk ke dalam data kategorik karena nilainya tidak dinyatakan dalam bentuk angka yang bisa dihitung secara matematis, melainkan dalam bentuk label atau nama yang menggambarkan karakteristik tertentu.

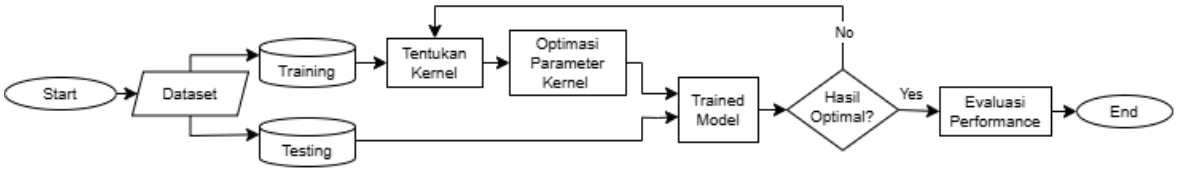
2.2. *Preprocessing* Data

Preprocessing data merupakan proses pengolahan awal terhadap data asli yang akan analisis, dengan tujuan memperbaiki kualitas data, menjamin keakuratan hasil, serta mengatasi berbagai masalah yang mungkin terdapat pada data asli [11]. Tahapan *preprocessing* meliputi:

1) *Data Cleaning*
   Dalam data *cleaning*, dilakukan pendeteksian dan penanganan terhadap *outlier* atau data yang menyimpang dari pola dan dapat memengaruhi hasil analisis secara negatif, serta penghapusan data duplikat yang dapat menyebabkan ketidakakuratan dalam hasil [11].
2) *Labelling*
   Proses *labelling* adalah langkah penting untuk menentukan atribut mana yang akan berperan sebagai label (kelas target) dalam analisis. Sebelum dataset dapat digunakan untuk membangun model, salah satu kolom data harus diset sebagai label yang menunjukkan kelas atau kategori dari setiap data[11].
3) *Encoding*
   Encoding dilakukan untuk mengonversi variabel kategori atau data teks menjadi representasi numerik agar algoritma pembelajaran mesin dapat memahami dan mengolah data tersebut [11]. Penelitian ini menggunakan teknik *One-Hot Encoding* untuk mengonversi variabel kategori menjadi vektor biner yang terdiri dari angka 0 dan 1, di mana setiap kategori hanya memiliki satu nilai 1 pada kolom yang sesuai kategorinya, sementara kolom lainnya bernilai 0 [11].
4) *Normalization*
   Untuk mengatasi masalah perbedaan skala, maka diterapkan proses normalisasi menggunakan metode standarisasi, agar dataset memiliki rentang yang seragam, sehingga perbedaan skala antar atribut tidak mempengaruhi model yang akan digunakan [12]. Metode proses normalisasi data pada penelitian ini adalah metode *Z-transformation* atau *Z-score normalization*.

2.3. Pemodelan Data

Tahap ini mengimplementasikan algoritma SVM yang digunakan sebagai metode utama dalam perhitungan data. Konsep SVM secara sederhana adalah menemukan *hyperplane* terbaik yang berperan sebagai pemisah antara dua kelas dalam ruang *input* [13]. Alur penerapan algoritma SVM dalam penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Alur penerapan algoritma SVM

Alur penerapan algoritma SVM pada Gambar 2 dalam penelitian ini menggunakan dataset tipe kulit wajah yang telah melalui tahap *preprocessing* sebagai data masukan untuk pemodelan menggunakan algoritma *Support Vector Machine* (SVM). Dataset kemudian dibagi menggunakan metode *10-Fold Cross Validation*, yaitu membagi data menjadi 10 bagian (*fold*) dengan sembilan *fold* digunakan sebagai data *training* dan satu *fold* sebagai data *testing* secara bergantian hingga seluruh data memperoleh kesempatan sebagai data *testing*.

Penelitian ini mengujikan kernel *Linear*, *Polynomial*, *Radial Basis Function* (RBF), dan *Sigmoid*. Dalam implementasi SVM pada *RapidMiner*, tersedia fitur untuk mengatur parameter penting yang berperan dalam mengontrol model dan kemampuan generalisasi seperti nilai C (*cost*), *gamma*, *degree* (untuk kernel polinomial), *coef* dan *epsilon* (untuk regresi).

1) Kernel *linear* adalah fungsi kernel yang paling sederhana dan digunakan ketika data yang akan diklasifikasikan dapat dipisahkan secara linier. Fungsi kernel *linear* pada persamaan 1.

$$K(x, xi) = (xi^T\ xj) \tag{1}$$

$K(x, xi) =$ merupakan nilai fungsi kernel antara data ke-i dan data ke-j, sedangkan x_i dan x_j merupakan vektor fitur dari masing-masing data.

2) Kernel *polynomial* merupakan fungsi kernel yang mewakili kesamaan vektor sampel pelatihan dalam ruang fitur yang ditunjukkan pada persamaan 2.

$$K(x, xi) = (\gamma xi^T\ xj + r)^d \tag{2}$$

γ merupakan parameter yang mengatur skala hasil perkalian *dot*. Parameter r merupakan konstanta dan d menunjukkan derajat polynomial.

3) Kernel *Radial Basis Function* (RBF) adalah konsep kernel yang paling banyak digunakan untuk memecahkan masalah klasifikasi prediksi pada kumpulan data yang tidak dapat dipisahkan secara linier. Fungsi kernel RBF pada persamaan 3.

$$\mathrm{K}\ (xi,\ xj) = exp(-\gamma \|xi - xj\|^2), \gamma > 0 \tag{3}$$

$\|xi - xj\|^2$ menunjukkan kuadrat jarak Euclidean antara kedua data.

4) Kernel *sigmoid* adalah pengembangan dari *Artificial Neural Network* (ANN). Kernel *sigmoid* dalam persamaan 4.

$$K(x, xi) = \tanh(\gamma xi^T\ xj\ + r \tag{4}$$

Fungsi tanh merupakan fungsi hiperbolik tangent yang digunakan untuk menghasilkan nilai keluaran dari fungsi kernel sigmoid.

Pengujian berbagai kernel pada algoritma *Support Vector Machine* (SVM) merupakan langkah penting dalam menentukan performa optimal model prediksi, khususnya dalam konteks prediksi tipe kulit wajah. Parameter yang digunakan disajikan pada Tabel 2.

Tabel 2. Optimasi Parameter

Kernel	Parameter
Linear	epsilon $= 0.01$, $C = 1$.
Polynomial	epsilon $= 0.01$, gamma $= 1$, $C = 1$, degree $= 3$, coef $= 0$
RBF	epsilon $= 0.01$, gamma $= 1$, $C = 1$.
Sigmoid	epsilon $= 0.01$, gamma $= 1$, $C = 1$.

2.4. Evaluasi

Hasil dari masing-masing model selanjutnya dievaluasi dan dibandingkan berdasarkan nilai performanya untuk menentukan model dengan akurasi tertinggi. Evaluasi dilakukan dengan menguji model menggunakan data pengujian dan mengukur performanya melalui metrik akurasi, presisi, dan *recall* yang diperoleh pada *confussion matrix*. *Confusion matrix* merupakan sebuah matriks yang digunakan untuk mengukur performa model *supervised learning*, khususnya dalam tugas klasifikasi pada *machine learning* [11]. Matriks ini menyajikan perbandingan antara hasil prediksi model dengan data aktual, sehingga memungkinkan penilaian yang lebih mendetail mengenai akurasi dan jenis kesalahan yang terjadi dalam proses klasifikasi. Proses ini bertujuan untuk mengetahui seberapa baik model mampu melakukan prediksi dan memenuhi tujuan penelitian.

3. Hasil dan Pembahasan

Proses pemodelan pada penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) yang diimplementasikan melalui perangkat lunak *RapidMiner*. Berdasarkan data yang telah diperoleh melalui *website* Kaggle sebanyak 2.000 baris data dan telah di *import* ke dalam *library RapidMiner,* berikut hasil statistik yang diperoleh ditunjukkan pada Gambar 3.

Name	Type	Missing	Statistics		Filter (8 / 8 attributes): Search for Attributes
Age	Integer	0	Min 18	Max 59	Average 38.721
Gender	Polynominal	0	Least Male (969)	Most Female (1031)	Values Female (1031), Male (969)
Hydration_Level	Polynominal	0	Least High (395)	Most Low (924)	Values Low (924), Medium (681), ...[1 more]
Oil_Level	Polynominal	0	Least Medium (653)	Most Low (689)	Values Low (689), High (658), ...[1 more]
Sensitivity	Polynominal	0	Least Low (648)	Most Medium (687)	Values Medium (687), High (665), ...[1 more]
Humidity	Real	0	Min 20	Max 80	Average 48.686
Temperature	Real	0	Min 10	Max 34.900	Average 20.636
Skin_Type	Polynominal	0	Least Oily (500)	Most Combination (500)	Values Combination (500), Dry (500), ...[2 more]

Showing attributes 1 - 8 — Examples: 2,000 Special Attributes: 0 Regular Attributes:

Gambar 3. Statistik data awal

Dalam Gambar 3 hasil yang diperoleh menunjukkan bahwa terdapat tiga tipe data yaitu *real*, *integer*, dan *polynominal*, serta tidak terdapat nilai *missing* pada seluruh atribut dalam data. Dengan kondisi data yang sudah lengkap tanpa nilai yang hilang, maka tidak diperlukan tahap data *cleaning* dan dapat dilanjutkan ke proses berikutnya yaitu memilih atribut *Skin_Type* dan mengubah perannya menjadi "label".

Kemudian dilakukan *encoding* terhadap tipe data kategorik, karena SVM hanya dapat memproses data numerik. Tanpa *encoding*, data kategorikal tidak dapat diolah dengan SVM. Hasil encoding disajikan pada Gambar 4.

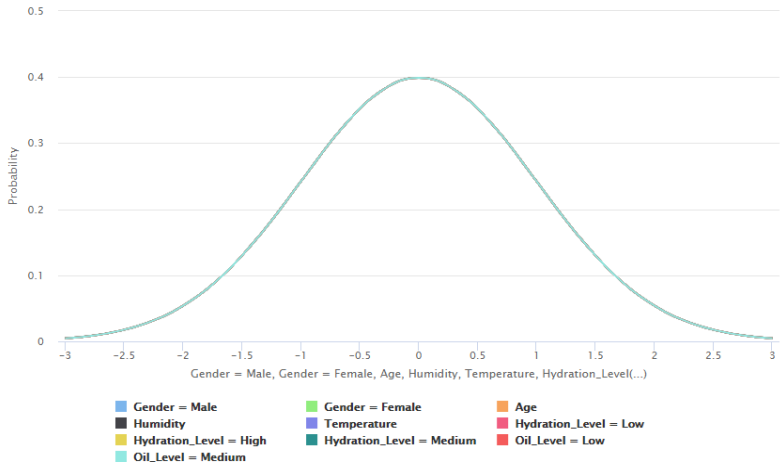
ExampleSet (2000 examples, 1 special attribute, 14 regular attributes) — Filter (2,000 / 2,000 examples): all

Row...	Skin_Type	Gend...	Gend...	Hydra...	Hydra...	Hydra...	Oil...	Oil...	Oil...	Sens...	Sens...	Sens...	Age	Humi...	Temper...
1	Dry	1	0	1	0	0	1	0	0	1	0	0	36	31.900	10.100
2	Combinat...	0	1	0	1	0	1	0	0	1	0	0	36	30.100	20.500
3	Normal	0	1	0	1	0	0	1	0	0	1	0	34	53	19.500
4	Oily	1	0	0	0	1	0	0	1	0	1	0	51	57.300	22.700
5	Normal	1	0	0	1	0	0	1	0	0	0	1	36	64.400	24
6	Normal	0	1	0	1	0	0	1	0	0	1	0	42	40.400	15.600
7	Oily	0	1	1	0	0	0	0	1	0	0	1	27	57.800	21.600
8	Dry	0	1	1	0	0	1	0	0	1	0	0	54	33.900	16.700
9	Combinat...	1	0	0	0	1	0	0	1	1	0	0	38	63.300	16.300
10	Normal	1	0	0	0	1	0	1	0	0	1	0	48	42.700	15.900
11	Combinat...	1	0	1	0	0	0	1	0	0	1	0	27	62.400	30.300
12	Normal	0	1	0	0	1	0	1	0	0	0	1	27	49	16.200
13	Normal	0	1	0	0	1	0	1	0	0	1	0	58	46	21.500
14	Oily	1	0	0	0	1	0	0	1	0	0	1	21	62.300	24
15	Combinat...	1	0	0	0	1	0	0	1	1	0	0	46	49.300	20.300
16	Combinat...	0	1	0	1	0	0	1	0	1	0	0	45	44.100	22.400
17	Oily	1	0	0	0	1	0	0	1	0	1	0	39	49.600	23

Gambar 4. Hasil Encoding

Gambar 4 memperlihatkan bahwa seluruh atribut kategorik (kecuali label) dalam dataset telah berhasil diubah menjadi bentuk numerik melalui teknik *one-hot encoding*. Setiap kategori hanya memiliki satu nilai 1 pada kolom yang sesuai dengan kategorinya, sementara kolom lainnya bernilai 0 [11]. Selain itu, atribut numerik tetap dipertahankan dalam bentuk aslinya.

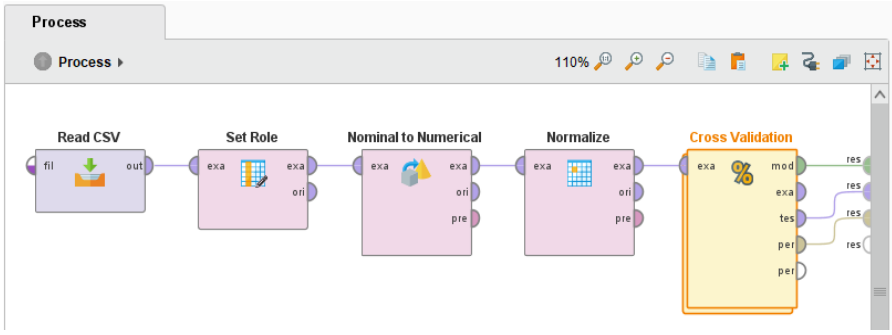
Tahap selanjutnya adalah melakukan analisis terhadap perbedaan grafik pengukuran antar atribut untuk menentukan karakteristik distribusi dan variasi masing-masing atribut dalam dataset. Untuk mengatasi masalah perbedaan skala tersebut, maka diterapkan proses normalisasi menggunakan metode standarisasi, agar dataset memiliki rentang yang seragam, sehingga perbedaan skala antar atribut tidak mempengaruhi model yang akan digunakan [12]. Setelah dilakukan normalisasi, maka hasilnya seperti pada Gambar 5 berikut.



Gambar 5. Hasil Normalisasi Data

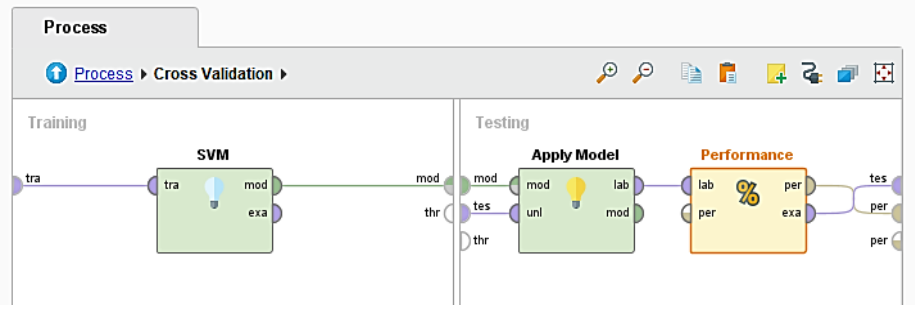
Berdasarkan grafik Gambar 5 hasil normalisasi di atas, dapat dilihat bahwa seluruh atribut telah terdistribusi mengikuti pola distribusi normal. Distribusi yang seragam seperti ini sangat penting untuk meningkatkan performa dan akurasi model, khususnya pada algoritma yang sensitif terhadap perbedaan skala data.

Setelah tahap normalisasi selesai, tahap selanjutnya adalah pengolahan data menggunakan algoritma *Support Vector Machine* (SVM). Pada tahap ini, dilakukan *10-Fold Cross Validation* untuk membagi data ke dalam 10 *fold*, yang digunakan secara bergantian sebagai data *training* dan *testing* dalam proses pembentukan serta pengujian model SVM. Proses ditunjukkan pada Gambar 6.



Gambar 6. Keseluruhan Proses

Dalam operator *Cross Validation* terdapat dua subproses utama, *training* dan *testing*. Pada bagian *training*, data yang ditentukan digunakan untuk membangun model klasifikasi menggunakan operator SVM sehingga menghasilkan *trained model*. Selanjutnya, pada bagian *testing*, data *testing* digunakan sebagai data masukan pada operator *Apply Model* untuk menghasilkan prediksi berdasarkan model SVM yang telah dilatih. Hasil prediksi tersebut kemudian dievaluasi menggunakan operator *Performance* untuk memperoleh nilai performa. Penerapan algoritma SVM pada Gambar 7.



Gambar 7. Penerapan algoritma SVM

Pengolahan data menggunakan *10-Fold Cross Validation* dengan menerapkan algoritma SVM serta empat jenis kernel yang parameternya telah ditentukan. Proses tersebut dilakukan untuk memperoleh performa dari setiap konfigurasi model. Performa hasil pengujian yang diperoleh ditampilkan pada Tabel 3.

Tabel 3. Hasil Pengujian Kernel

Kernel	***Accuracy***	***Mean Precision***	***Mean Recall***
Linear	94.90%	95.40%	94.87 %
Polynomial	96.20%	96.19%	96.13%
RBF	98.20%	98.19%	98.22%
Sigmoid	58.65%	61.82%	58.75%

Berdasarkan Tabel 3 hasil pengujian menggunakan 10-*Fold Cross Validation*, nilai akurasi tertinggi didapatkan melalui kernel *Radial Basis Function* (RBF) sebesar 98.20% dengan parameter epsilon $= 0.01$, gamma $= 1$ dengan nilai $C = 1$. Kernel ini menghasilkan performa terbaik dibandingkan dengan kernel lain yang diuji, yang menunjukkan bahwa kernel RBF paling optimal untuk menangani kasus penelitian ini.

Analisis hasil prediksi dilakukan dengan memanfaatkan *confusion matrix* untuk melihat kesesuaian antara hasil prediksi model dengan kelas sebenarnya pada data. *Confusion matrix* menunjukkan jumlah data pada setiap kelas yang berhasil diprediksi dengan benar maupun yang masih mengalami kesalahan prediksi. Hasil akurasi yang diperoleh dari metode *confusion matrix* ditampilkan pada Gambar 8 berikut.

accuracy: 98.20% +/- 1.14% (micro average: 98.20%)

	true Dry	true Combination	true Normal	true Oily	class precision
pred. Dry	500	2	0	0	99.60%
pred. Combination	0	470	2	4	98.74%
pred. Normal	0	13	498	0	97.46%
pred. Oily	0	15	0	496	97.06%
class recall	100.00%	94.00%	99.60%	99.20%	

Gambar 8. Hasil Akurasi

Pada Gambar 8 di atas, hasil evaluasi algortima SVM pada kernel RBF untuk prediksi tipe kulit dianalisis berdasarkan metrik akurasi yang diperoleh dari *confusion matrix*. Secara keseluruhan didapatkan nilai akurasi sebesar 98.20%, ini menandakan bahwa model tersebut mampu memprediksi data dengan tingkat ketepatan yang sangat tinggi.

Nilai presisi pada model SVM dengan kernel RBF diperoleh berdasarkan hasil prediksi pada masing-masing kelas. Presisi digunakan untuk mengukur ketepatan model dalam memprediksi suatu kelas, yaitu berdasarkan perbandingan antara data yang diprediksi benar pada suatu kelas dengan seluruh data yang diprediksi sebagai kelas tersebut. Hasil nilai presisi pada masing-masing tipe kulit wajah ditampilkan pada Gambar 9.

weighted_mean_precision: 98.19% +/- 1.10% (micro average: 98.22%), weights: 1, 1, 1, 1					
	true Dry	true Combination	true Normal	true Oily	class precision
pred. Dry	500	2	0	0	99.60%
pred. Combination	0	470	2	4	98.74%
pred. Normal	0	13	498	0	97.46%
pred. Oily	0	15	0	496	97.06%
class recall	100.00%	94.00%	99.60%	99.20%	

Gambar 9. Hasil Presisi

Berdasarkan Gambar 9, nilai metrik *weighted mean precision* sebesar 98.19%. Selain itu, nilai presisi untuk masing-masing kelas (c*lass precision*) ini bervariasi, pada prediksi tipe kulit kering mendapat nilai presisi sebesar 99.60%, prediksi tipe kulit kombinasi mendapat nilai presisi sebesar 98.74%, prediksi tipe kulit normal mendapat nilai presisi sebesar 97.46%, prediksi tipe kulit berminyak mendapat nilai presisi sebesar 97.06%. Hasil tersebut menunjukkan bahwa model SVM dengan kernel RBF memiliki tingkat ketepatan prediksi yang tinggi pada masing-masing kelas.

Metrik *recall* digunakan untuk mengetahui kemampuan model dalam mengenali seluruh data yang sebenarnya termasuk ke dalam masing-masing kelas. Nilai *recall* diperoleh dari hasil prediksi model pada proses pengujian menggunakan 10-*Fold Cross Validation*. Hasil nilai *recall* untuk masing-masing tipe kulit wajah ditampilkan pada Gambar 10.

weighted_mean_recall: 98.22% +/- 1.19% (micro average: 98.20%), weights: 1, 1, 1, 1					
	true Dry	true Combination	true Normal	true Oily	class precision
pred. Dry	500	2	0	0	99.60%
pred. Combination	0	470	2	4	98.74%
pred. Normal	0	13	498	0	97.46%
pred. Oily	0	15	0	496	97.06%
class recall	100.00%	94.00%	99.60%	99.20%	

Gambar 10. Hasil Recall

Berdasarkan Gambar 10, nilai metrik *weighted mean recall* sebesar 98,22%. Selain itu, nilai *recall* pada tipe kulit kering sebesar 100%, tipe kulit kombinasi sebesar 94.00%, tipe kulit normal sebesar 99.60%, tipe kulit berminyak sebesar 99.20%. Tingginya nilai *recall* pada setiap kelas menunjukkan bahwa kesalahan model dalam melewatkan data dari kelas yang sebenarnya relatif rendah.

Nilai *Root Mean Squared Error* (RMSE) yang diperoleh dari hasil analisis menggunakan *RapidMiner* adalah sebesar 0.124 atau 12.4%. Hal ini menunjukkan bahwa tingkat kesalahan rata-rata antara nilai prediksi dan nilai aktual pada model yang dibangun relatif kecil, sehingga akurasi prediksi yang dihasilkan dapat dikategorikan baik

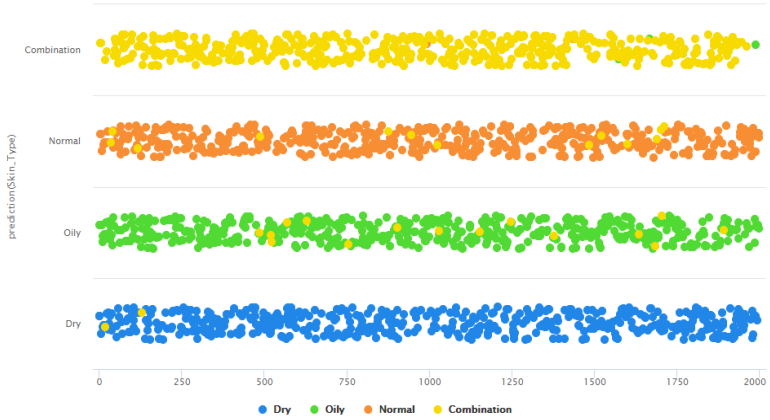
Pada model *Support Vector Machine* (SVM) yang digunakan dalam penelitian ini, jumlah total *support vectors* yang diperoleh adalah sebanyak 944 dengan nilai bias (*offset*) yang diperoleh adalah sebesar 0.915.

Setelah dilakukan proses perhitungan menggunakan model yang telah dibangun. Dari total dataset yang diuji, terdapat 1.964 data yang berhasil diprediksi tepat sesuai dengan kelas aslinya. Sementara itu, terdapat 36 data yang tidak sesuai dengan kelas data awal, dimana model memprediksi kelas yang berbeda dari kelas sebenarnya. Data prediksi yang tidak sesuai dari kelas sebenarnya ditampilkan pada Gambar 11 berikut.

No	Skin_Type	prediction (Skin_Type)	conf. (Normal)	conf. (Dry)	conf. (Oily)	conf. (Combi.)	Gender=Male	Gender=Female	Hydration_Level=Low	Hydration_Level=Medium	Hydration_Level=High	Oil_Level=Low
1	Combination	Dry	0	1	0	0	1,031	-1,031	1,079	-0,496	-0,718	1,379
2	Combination	Normal	0	0	1	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
3	Combination	Normal	0	1	0	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
4	Combination	Normal	1	0	0	0	-0,969	0,969	-0,926	2,015	-0,718	-0,725
5	Combination	Dry	0	0	0	1	-0,969	0,969	1,079	-0,496	-0,718	1,379
6	Combination	Oily	0	1	0	0	1,031	-1,031	1,079	-0,496	-0,718	-0,725
7	Combination	Normal	0	0	0	1	-0,969	0,969	-0,926	-0,496	1,391	-0,725
8	Combination	Oily	1	0	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
9	Combination	Oily	0	0	1	0	-0,969	0,969	1,079	-0,496	-0,718	-0,725
1	Combination	Oily	1	0	0	0	1,031	-1,031	1,079	-0,496	-0,718	-0,725
11	Combination	Oily	0	0	1	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
12	Combination	Oily	0	0	1	0	1,031	-1,031	1,079	-0,496	-0,718	-0,725
13	Combination	Normal	0	1	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
14	Combination	Oily	0	1	0	0	-0,969	0,969	1,079	-0,496	-0,718	-0,725
15	Combination	Normal	1	0	0	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
16	Normal	Combination	0	0	1	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
17	Combination	Normal	0	1	0	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
18	Combination	Oily	1	0	0	0	-0,969	0,969	1,079	-0,496	-0,718	-0,725
19	Combination	Oily	0	0	0	1	-0,969	0,969	-0,926	-0,496	1,391	-0,725
20	Combination	Oily	0	1	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
21	Combination	Oily	0	0	1	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
22	Normal	Combination	0	0	1	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
23	Oily	Combination	0	0	0	1	-0,969	0,969	-0,926	-0,496	1,391	-0,725
24	Combination	Normal	0	0	1	0	1,031	-1,031	-0,926	2,015	-0,718	-0,725
25	Combination	Normal	0	1	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
26	Oily	Combination	1	0	0	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
27	Combination	Normal	1	0	0	0	1,031	-1,031	-0,926	2,015	-0,718	-0,725
28	Combination	Oily	0	1	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
29	Oily	Combination	0	0	1	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
30	Combination	Oily	0	1	0	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
31	Combination	Normal	0	0	1	0	-0,969	0,969	-0,926	2,015	-0,718	-0,725
32	Combination	Normal	0	1	0	0	1,031	-1,031	-0,926	-0,496	1,391	-0,725
33	Combination	Oily	0	0	1	0	-0,969	0,969	-0,926	-0,496	1,391	-0,725
34	Combination	Normal	0	1	0	0	1,031	-1,031	-0,926	2,015	-0,718	-0,725
35	Combination	Oily	0	0	1	0	-0,969	0,969	1,079	-0,496	-0,718	-0,725
36	Oily	Combination	0	0	1	0	1,031	-1,031	1,079	-0,496	-0,718	-0,725

Gambar 11. Data Hasil Prediksi

Berdasarkan data prediksi yang tidak sesuai dengan kelas sebenarnya pada Gambar 11, visualisasi hasil prediksi disajikan dalam bentuk grafik pada Gambar 12.



Gambar 12. Grafik Hasil Prediksi

Berdasarkan Gambar 12, grafik *scatter* tersebut memperlihatkan sebaran prediksi terhadap empat kategori tipe kulit, yaitu kering, berminyak, normal dan kombinasi. Sumbu horizontal pada grafik merupakan jumlah dataset sebanyak 2.000 data, sedangkan sumbu vertikal pada grafik menunjukkan hasil prediksi model.

Warna setiap titik pada grafik menggambarkan label asli dari data tersebut. Titik-titik yang warnanya sesuai dengan posisi prediksinya menandakan prediksi yang benar, sementara titik-titik yang warnanya berbeda dari posisi prediksi menunjukkan adanya ketidaksesuaian prediksi. Berdasarkan sebaran titik yang terlihat, dapat diamati bahwa masih terdapat sejumlah data yang diprediksi ke dalam kategori yang tidak sesuai dengan label aslinya

4. Kesimpulan

Model prediksi tipe kulit berhasil dibangun menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) melalui pengujian 10-*Fold Cross Validation*.

Kernel RBF dengan parameter epsilon = 0.01, gamma = 1, dan cost = 1 menghasilkan akurasi 98,20%, precision 98,19%, recall 98,22%, dan RMSE 12,4%. Dari 2.000 data yang dievaluasi, terdapat 36 data yang tidak sesuai dengan label asli, sehingga model mampu melakukan prediksi tipe kulit kering, berminyak, normal dan kombinasi dengan tingkat ketepatan yang tinggi.

Daftar Pustaka

[1] N. Nawiyah, R. C. Kaemong, M. A. Ilham, and F. Muhammad, "Penyebab Pengaruhnya Pertumbuhan Pasar Indonesia Terhadap Produk Skin Care Lokal Pada Tahun 2022," *Armada : Jurnal Penelitian Multidisiplin*, vol. 1, no. 12, pp. 1390–1396, Dec. 2023, doi: 10.55681/armada.v1i12.1060.

[2] N. A. Sundari, R. Magladena, and S. Saidah, "Klasifikasi Jenis Kulit Wajah Menggunakan Metode Covolutional Neural Network (CNN) Efficientnet-B0," *eProceedings of Engineering*, vol. 9, no. 6, pp. 3180–3187, Dec. 2022.

[3] N. Andrini, "Karakteristik Dan Perawatan Kulit Untuk Orang Asia," *Jurnal Pandu Husada*, vol. 4, no. 3, pp. 14–23, Jul. 2023, [Online]. Available: https://jurnal.umsu.ac.id/index.php/JPH

[4] J. Ran, G. Dong, F. Yi, L. Li, and Y. Wu, "Automatic Measurement of Comprehensive Skin Types Based on Image Processing and Deep Learning," *Electronics (Switzerland)*, vol. 14, no. 1, Jan. 2025, doi: 10.3390/electronics14010049.

[5] A. D. W. M. Sidik, I. H. Kusumah, A. Suryana, M. Artiyasa, A. Junfithrana, and Edwinanto, "Gambaran Umum Metode Klasifikasi Data Mining," *Fidelity : Jurnal Teknik Elektro*, vol. 2, no. 2, pp. 34–38, 2020.

[6] A. Toha, P. Purwono, and W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, no. 1, pp. 12–21, May 2022, doi: 10.12928/biste.v4i1.6079.

[7] A. Permana *et al.*, *Machine Learning*, 1st ed. Padang: Global Eksekutif Teknologi, 2023. [Online]. Available: www.globaleksekutifteknologi.co.id

[8] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indonesian Journal on Software Engineering (IJSE)*, vol. 7, no. 1, pp. 63–68, 2021, [Online]. Available: http://ejournal.bsi.ac.id/ejurnal/index.php/ijse63

[9] M. A. Al Fatih and K. M. Lhaksmana, "Prediksi Employee Attrition menggunakan Algoritma Support Vector Machine (SVM)," *e-Proceeding of Engineering*, vol. 10, no. 2, pp. 1930–1938, Apr. 2023.

[10] X. Xiahou and Y. Harada, "B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 2, pp. 458–475, Jun. 2022, doi: 10.3390/jtaer17020024.

[11] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. Jambi: Sonpedia, 2024. [Online]. Available: https://www.researchgate.net/publication/377415198

[12] R. C. Prihandari and Mustakim, *Data Mining: Konsep Dan Apikasi Menggunakan Rapidminer (Series: Supervised Learning Dan Unsupervised Learning)*. Institut Riset Dan Publikasi Indonesia, 2021.

[13] Muttaqin *et al.*, *Pengenalan Data Mining*. Yayasan Kita Menulis, 2023. Accessed: Apr. 27, 2025. [Online]. Available: https://kitamenulis.id/2023/09/29/data-mining-teori-dan-implementasi/