

PERBANDINGAN KINERJA SUPPORT VECTOR MACHINE DALAM KLASIFIKASI OBESITAS DENGAN PENDEKATAN KERNEL LINEAR DAN RADIAL BASIS FUNCTION

Muhamad Rois.S¹⁾, Kusrini²⁾

^{1, 2)} Universitas AMIKOM Yogyakarta

Email : rois@students.amikom.ac.id¹⁾, kusrini@amikom.ac.id²⁾

ABSTRAK

Obesitas adalah kondisi medis yang ditandai dengan penumpukan lemak tubuh yang berlebihan hingga dapat menimbulkan risiko berbagai penyakit kronis, seperti diabetes, penyakit jantung, dan kanker. Di Indonesia, dalam kurun waktu 10 tahun terjadi peningkatan obesitas yang signifikan, dari 10,5% pada tahun 2007 menjadi 21,8% pada tahun 2018. Secara global, pada tahun 2030 diperkirakan 1 dari 5 wanita dan 1 dari 7 pria akan hidup dengan obesitas, yang setara dengan lebih dari 1 miliar orang di seluruh dunia. Untuk mengatasi permasalahan ini, teknologi kecerdasan buatan digunakan dalam prediksi obesitas guna mengidentifikasi faktor risiko secara lebih akurat. Penelitian ini membandingkan performa klasifikasi obesitas menggunakan algoritma *Support Vector Machine* (SVM) dengan dua pendekatan berbeda: SVM dengan kernel Linear tanpa *hyperparameter tuning* dan SVM dengan kernel *Radial Basis Function* (RBF) dengan *hyperparameter tuning*. Dataset yang digunakan bersumber dari Universitas Sinop yang tersedia di Kaggle, dengan total 1610 data. Hasil eksperimen menunjukkan bahwa model SVM Linear tanpa *hyperparameter tuning* memiliki akurasi 72% pada data uji, sedangkan model SVM-RBF dengan *hyperparameter tuning* C dan gamma mencapai akurasi 83%. Perbedaan performa ini menunjukkan bahwa pemilihan kernel dan penerapan *hyperparameter tuning* dapat meningkatkan akurasi serta keandalan prediksi obesitas.

Kata Kunci : obesitas, *Support Vector Machine*, SVM-RBF, *hyperparameter tuning*.

ABSTRACT

Obesity is a medical condition characterized by excessive fat accumulation, which can increase the risk of chronic diseases such as diabetes, heart disease, and cancer. In Indonesia, obesity has significantly increased over the past decade, rising from 10.5% in 2007 to 21.8% in 2018. Globally, by 2030, it is estimated that 1 in 5 women and 1 in 7 men will be living with obesity, totaling more than 1 billion people worldwide. To address this issue, artificial intelligence technology is utilized in obesity prediction to accurately identify risk factors. This study compares the performance of obesity classification using the *Support Vector Machine* (SVM) algorithm with two different approaches: Linear SVM without *hyperparameter tuning* and *Radial Basis Function* (RBF) SVM with *hyperparameter tuning*. The dataset used comes from Sinop University and is available on Kaggle, consisting of 1,610 data points. Experimental results show that the Linear SVM model without *hyperparameter tuning* achieves 72% accuracy on test data, while the RBF SVM model with optimized C and gamma parameters achieves 83% accuracy. This performance difference indicates that selecting the appropriate kernel and applying *hyperparameter tuning* can improve accuracy and reliability in obesity prediction.

Keywords: obesity, *Support Vector Machine*, SVM-RBF, SVM Linear, *hyperparameter tuning*, classification.

1. PENDAHULUAN

Obesitas merupakan kondisi medis yang ditandai dengan penumpukan lemak tubuh berlebih yang dapat meningkatkan risiko berbagai penyakit kronis, seperti diabetes, penyakit jantung, hipertensi, dan kanker (Toar et al., 2023). Obesitas tidak hanya berdampak pada penampilan fisik, tetapi juga menjadi salah satu penyebab utama penyakit tidak menular yang berkontribusi terhadap peningkatan angka morbiditas dan mortalitas (Dianingsih et al., 2022). Secara global, kasus obesitas terus meningkat secara signifikan. Pada tahun 2030, diperkirakan 1 dari 5 wanita dan 1 dari 7 pria akan mengalami obesitas, yang setara dengan lebih dari 1 miliar orang di seluruh dunia (Kemenkes, 2023).

Di Indonesia, kasus obesitas mengalami peningkatan yang cukup signifikan dalam kurun waktu sepuluh tahun. Berdasarkan data Kementerian Kesehatan (2023), kasus obesitas meningkat dari 10,5% pada tahun 2007 menjadi 21,8% pada tahun 2018. Obesitas juga berkontribusi terhadap penyebab kematian akibat penyakit kardiovaskular (5,87% dari total kematian) serta penyakit diabetes dan ginjal (1,84% dari total kematian). Kondisi ini menunjukkan bahwa obesitas telah menjadi permasalahan kesehatan yang serius dan memerlukan upaya pencegahan serta penanganan yang lebih efektif (Wiardani et al., 2023).

Untuk mengatasi permasalahan ini, pendekatan berbasis teknologi mulai banyak digunakan dalam dunia medis, salah satunya adalah kecerdasan buatan dengan *Machine Learning*. *Machine Learning* dapat membantu dalam menganalisis pola data kesehatan dan memberikan prediksi yang lebih akurat dalam mengidentifikasi faktor risiko obesitas (Arora et al., 2022). Salah satu algoritma dalam *Machine Learning* yang sering digunakan untuk klasifikasi adalah *Support Vector Machine* (SVM), yang mampu mengolah data dengan kompleksitas tinggi dan menghasilkan model prediksi dengan tingkat akurasi yang baik (Javeed et al., 2023).

SVM bekerja dengan membangun *hyperplane* yang memisahkan data dari

berbagai kelas dengan margin maksimum. Dalam kasus klasifikasi obesitas, pemilihan kernel sangat mempengaruhi performa model. Kernel Linear sering digunakan karena kesederhanaannya, tetapi tidak selalu optimal dalam menangani data yang tidak terdistribusi secara linear. Sebaliknya, kernel *Radial Basis Function* (RBF) lebih fleksibel dalam menangani pola data yang lebih kompleks (Reza et al., 2023). Selain itu, penerapan *hyperparameter tuning* pada SVM-RBF, seperti pengoptimalan parameter C dan gamma, dapat meningkatkan performa model dengan menyesuaikan keseimbangan antara margin pemisahan dan kompleksitas model (Reza et al., 2023).

Beberapa penelitian sebelumnya telah membandingkan performa SVM dengan berbagai pendekatan dalam klasifikasi obesitas. Misalnya, penelitian (Wong et al., 2022) membandingkan SVM dengan algoritma *Machine Learning* lainnya seperti Random Forest dan XGBoost dalam memprediksi obesitas di kalangan pekerja dewasa Malaysia. Hasilnya menunjukkan bahwa SVM memiliki akurasi sekitar 72%, yang masih dapat ditingkatkan dengan penerapan *hyperparameter tuning*. Penelitian lainnya (Koklu & Sulak, 2024) menunjukkan bahwa model SVM tanpa *hyperparameter tuning* hanya mencapai akurasi 74,03%, yang menunjukkan adanya keterbatasan dalam mengklasifikasikan data obesitas secara optimal.

Penelitian yang dilakukan oleh Koklu dan Sulak (Koklu & Sulak, 2024) melakukan analisis status obesitas berdasarkan aktivitas sosial dan fisik individu. Penelitian ini menggunakan data yang dikumpulkan melalui survei online dari 1.610 partisipan di Turki. Dalam studi ini, empat metode kecerdasan buatan diterapkan, yaitu *Artificial Neural Networks* (ANN), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), dan *Random Forest* (RF). Hasilnya menunjukkan bahwa *Random Forest* memiliki performa terbaik dengan *accuracy* tertinggi sebesar 87,82%, sementara SVM mencatatkan *accuracy* terendah, yaitu 74,03%. Meskipun algoritma ini cukup baik dalam

mengklasifikasikan beberapa kategori, terdapat tantangan dalam membedakan kategori *overweight* dan *obesity*. Berdasarkan analisis *Confusion Matrix*, SVM menunjukkan tingkat kesalahan klasifikasi yang cukup tinggi di antara kedua kelas tersebut, yang kemungkinan besar disebabkan oleh kemiripan karakteristik data atau ketidakseimbangan jumlah sampel di antara kategori tersebut karena dalam penelitian tersebut tidak disebutkan *preprocessing* datanya.

Selain itu *Support Vector Machine* (SVM) merupakan algoritma *Machine Learning* yang sensitif terhadap perbedaan skala. Oleh karenanya, min-max scaler membantu meningkatkan kinerja model dengan memastikan bahwa semua fitur memberikan kontribusi yang proporsional (Han et al., 2012). Dan *Support Vector Machine* memiliki kelemahan dalam menangani data yang tidak seimbang, oleh sebab itu teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang pertama kali dikenalkan oleh (Chawla et al., 2002) memberikan kontribusi dalam menangani data yang tidak seimbang.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membandingkan performa klasifikasi obesitas menggunakan SVM Linear tanpa *hyperparameter tuning* dan SVM-RBF dengan *hyperparameter tuning*. Penerapan *hyperparameter tuning* juga menjadi aspek penting dalam meningkatkan performa SVM-RBF. Parameter C dan gamma (γ) dioptimalkan dengan menggunakan teknik *Grid Search*, yang memungkinkan pemilihan kombinasi parameter terbaik untuk memperoleh hasil klasifikasi yang lebih akurat. Nilai C mengontrol kompleksitas model, sementara gamma (γ) menentukan seberapa jauh pengaruh satu titik data terhadap titik lain dalam ruang fitur. Eksperimen dilakukan dengan berbagai kombinasi parameter, dan hasil terbaik diperoleh dengan parameter C dan gamma (γ) tertentu yang memberikan akurasi tertinggi. Oleh karenanya, selain untuk membandingkan kedua model, penelitian ini juga bertujuan untuk mencari nilai parameter C dan gamma (γ) yang terbaik pada SVM-RBF yang

sebelumnya sudah dilakukan *pre-processing* data. Selanjutnya melakukan evaluasi terhadap metrik performa seperti akurasi, presisi, *recall*, dan F1-score.

2. METODE

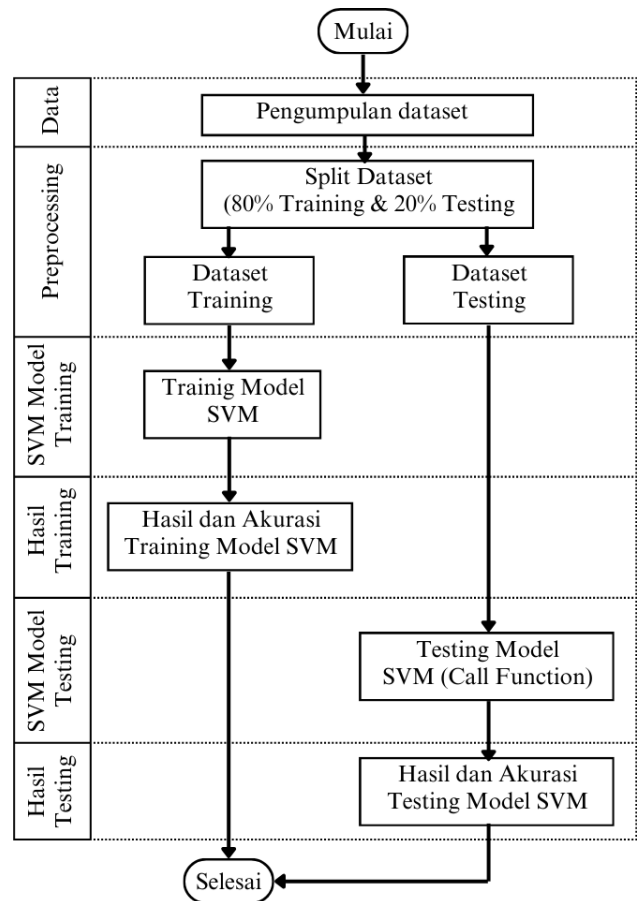
Pengumpulan data dilakukan untuk mendapatkan berbagai macam data yang diperlukan dalam penelitian. Dataset yang digunakan merupakan dataset dari hasil penelitian Niğmet Köklü dan Süleyman Alpaslan Sulak (2024) yang berasal dari Universitas Sinop, di Turki yang didapat melalui *Open Dataset Repository Kaggle*.

Dataset yang digunakan dalam penelitian ini sebagaimana yang terlihat pada tabel 1, terdiri dari 14 kolom dengan total 1610 data yang mencakup berbagai faktor yang berhubungan dengan obesitas. Data tersebut mencakup variabel kebiasaan makan, aktivitas fisik, serta faktor gaya hidup lainnya yang berkontribusi terhadap risiko obesitas. Berikut sebaran data dalam dataset obesitas:

Tabel 1. Distibusi data pada dataset

Atribut	Fitur	Data
Sex	1. Male	712
	2. Female	898
Age	Values in integers	1610
Height	Values in integers (cm)	1610
Overweight/ Obese Families	1. Yes	266
	2. No	1344
Consumption of Fast Food	1. Yes	436
	2. No	1174
Frequency of Consuming Vegetables	1. Rarely	400
	2. Sometimes	708
	3. Always	502
Number of Main Meals Daily	1. 1-2	444
	2. 3	928
	3. 3+	238
Food Intake Between Meals	1. Rarely	346
	2. Sometimes	564
	3. Usually	417
	4. Always	283
Smoking	1. Yes	492
	2. No	118

Liquid Intake Daily	1. Amount smaller than one liter	456
	2. Within the range of 1 to 2 liters	523
	3. In excess of 2 liters	631
Calculation Of Calorie Intake	1. Yes	286
	2. No	1324
Physical Exercise	1. No physical activity	206
	2. In the range of 1-2 days	290
	3. In the range of 3-4 days	370
	4. In the range of 5-6 days	358
	5. 6+ days	386
Schedule Dedicated to Technology	1. Between 0 and 2 hours	382
	2. Between 3 and 5 hours	826
	3. Exceeding five hours	402
Type of Transportation Used	1. Automobile	660
	2. Motorbike	94
	3. Bike	116
	4. Public transportation	602
	5. Walking	138
Class	1. Underweight	73
	2. Normal	658
	3. Overweight	592
	4. Obesity	287

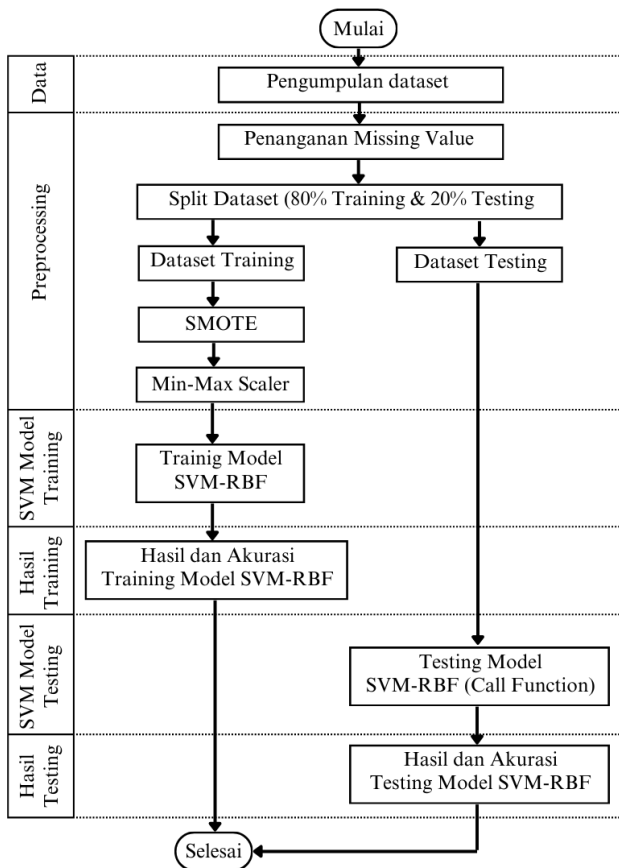


Gambar 1. Alur Model SVM Linear

Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dengan dua pendekatan berbeda untuk klasifikasi obesitas. Pendekatan pertama adalah SVM linear tanpa *hyperparameter tuning* dan tanpa *preprocessing* data melainkan model ini langsung diterapkan pada dataset sebagaimana alur pada gambar 1. SVM linear bekerja dengan mencari *hyperplane* terbaik yang memisahkan kelas-kelas dalam dataset berdasarkan margin maksimum. Pendekatan ini digunakan sebagai *baseline* untuk membandingkan efektivitas teknik *preprocessing* dan *hyperparameter tuning* serta pemilihan kernel pada SVM pada pendekatan lainnya.

Pendekatan kedua menggunakan SVM dengan kernel *Radial Basis Function* (RBF) yang dilengkapi dengan *hyperparameter tuning* untuk parameter C dan gamma (γ). Sebelum dilakukan pemodelan, dataset menjalani serangkaian *preprocessing*, termasuk penanganan *missing value* dengan metode imputasi, *balancing* dataset menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), dan normalisasi fitur dengan Min-Max Scaler.

Pada gambar 2 terlihat alur yang diterapkan pada penelitian ini pada model SVM-RBF yang terdapat *hyperparameter tuning* pada proses training model RBF setelah dataset melewati *preprocessing* data sebelumnya.



Gambar 2. Alur Model SVM-RBF

2.1 Penanganan *Missing value*

Dataset yang tidak memiliki nilai merupakan *missing value* dan dapat mengurangi kualitas analisis dan performa model SVM. *Missing value* dapat terjadi karena berbagai alasan, seperti kesalahan saat pengumpulan data atau kurangnya input dari responden. Pada dataset obesitas, penanganan *missing value* dilakukan dengan beberapa metode, yaitu untuk kolom numerik seperti *Height* dan *Age*, metode imputasi pengisian dengan mean (rata-rata) serta pada kolom kategorikal menggunakan modus.

2.2 Split Dataset

Proses membagi dataset menjadi beberapa bagian untuk data *training* dan data *testing*. Pada dataset obesitas, pembagian 80% untuk dataset *training* dan 20% untuk dataset *testing*. Diketahui dataset sejumlah 1610, maka data *training* sejumlah 1288 data dan data *testing* sejumlah 322 data.

2.3 Synthetic Minority Over-sampling Technique (SMOTE)

Proses ini memastikan bahwa distribusi kelas pada variabel target, yaitu status obesitas dalam dataset berada dalam proporsi yang seimbang. *Balancing* dataset ini dilakukan dengan teknik *oversampling*, yaitu membuat dan menambahkan data sintetis pada kelas minoritas. Metode SMOTE ini dilakukan pada dataset *training* agar data *testing* tetap *independent*. Diketahui bahwa dataset *training* sejumlah 1288 dengan pembagian kelas terbesar 526 data sedangkan kelas terkecil 58 data sehingga dibutuhkan penyeimbangan jumlah data, sehingga setelah proses ini dilakukan, maka jumlah data disemua kelas akan menjadi 526 data.

2.4 Min-Max Scaler

Algoritma Support Vector Machine (SVM) sangat sensitif terhadap perbedaan skala antar fitur. Fitur dengan rentang besar dapat mendominasi model dan menyebabkan bobot tertentu menjadi lebih berpengaruh dibandingkan fitur lainnya. Oleh karena itu, normalisasi menjadi langkah penting dalam preprocessing data sebelum diterapkan ke dalam model SVM.

Normalisasi Min-Max Scaler mengubah nilai fitur dalam dataset ke dalam rentang 0 sampai 1, sehingga semua fitur memiliki skala yang sama dan tidak ada fitur yang mendominasi proses pembelajaran model.

2.5 SVM-RBF

Modeling dataset obesitas menggunakan SVM dengan kernel RBF bertujuan untuk mengklasifikasikan data ke dalam empat kelas (*underweight*, *normal*, *overweight*, dan *obesity*). Dalam algoritma Support Vector Machine (SVM) dengan Radial Basis Function (RBF) kernel, keputusan klasifikasi didasarkan pada pemetaan non-linear data ke dalam dimensi yang lebih tinggi menggunakan fungsi kernel. Rumus umum untuk kernel RBF diberikan oleh:

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2)$$

dengan $K(x_i, x_j)$ sebagai fungsi kernel yang mengukur kesamaan antara dua vektor fitur x_i dan x_j , sedangkan γ (gamma) adalah hiperparameter yang mengontrol sejauh mana satu sampel data dapat memengaruhi keputusan klasifikasi. Nilai gamma yang lebih tinggi membuat model lebih kompleks dengan batas keputusan yang lebih fleksibel, sedangkan nilai gamma yang lebih rendah menghasilkan batas keputusan yang lebih sederhana. Dengan pemetaan ini, SVM dapat menemukan hyperplane optimal bahkan dalam kasus data yang tidak terpisahkan secara linear.

Hyperparameter-tuning untuk C dan gamma (γ) di definisikan dalam parameter grid yang difungsikan untuk menampung parameter C dan gamma (γ) *hyperparameter-tuning* untuk C dan gamma (γ). Nilai parameter C diantaranya, [0.1, 1, 10, 100] dan nilai parameter gamma (γ) diantaranya, [0.001, 0.01, 0.1, 1]. Namun nilai tersebut tidak terbatas pada yang sudah didefinisikan. Melainkan dapat dilakukan perubahan pada saat *training* model SVM-RBF. Selanjutnya akan dipilih parameter C dan gamma (γ) yang terbaik sesuai dengan data *training* yang telah di *preprocessing* SMOTE menjadi sejumlah 2104 data. Selanjutnya parameter C dan gamma (γ) yang terpilih, akan dijadikan sebagai parameter dalam pelatihan model SVM-RBF.

Parameter C berperan dalam menentukan kompromi antara margin yang lebar dan kesalahan klasifikasi, sementara gamma (γ) mengontrol pengaruh satu titik data terhadap pembentukan *hyperplane* dalam ruang dimensi tinggi. Pemilihan *hyperparameter* terbaik dilakukan berdasarkan evaluasi pada data *training* menggunakan teknik validasi silang untuk menghindari *overfitting*. Setelah model dengan parameter optimal diperoleh, evaluasi dilakukan pada data *testing* untuk mengukur performanya dalam klasifikasi obesitas.

Evaluasi model dilakukan dengan membandingkan performa kedua pendekatan menggunakan metrik akurasi, presisi, *recall*, dan F1-score. Model SVM linear tanpa

preprocessing dibandingkan dengan SVM-RBF yang telah melalui *preprocessing* dan *hyperparameter tuning* untuk menentukan dampak *preprocessing* dan *tuning* terhadap kinerja klasifikasi. Analisis hasil evaluasi memberikan gambaran mengenai efektivitas SVM dalam prediksi obesitas, serta sejauh mana *preprocessing* data dan *hyperparameter tuning* dapat meningkatkan akurasi dan kemampuan generalisasi model.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, dilakukan analisis perbandingan performa antara dua model klasifikasi obesitas menggunakan algoritma *Support Vector Machine* (SVM), yaitu SVM Linear tanpa *preprocessing* dan *hyperparameter tuning* serta SVM dengan kernel *Radial Basis Function* (RBF) yang telah melalui tahap *preprocessing* data dan *hyperparameter tuning*. Evaluasi model dilakukan berdasarkan metrik akurasi, presisi, *recall*, dan F1-score pada data *training* dan *testing* untuk mengukur sejauh mana masing-masing model mampu mengklasifikasikan obesitas dengan baik.

3.1 Evaluasi Model SVM Linear

Model SVM Linear diterapkan langsung pada dataset tanpa dilakukan *preprocessing* atau *hyperparameter tuning*. Hasil evaluasi menunjukkan bahwa pada data *training*, model ini memiliki akurasi sebesar 76% dengan nilai presisi, *recall*, dan F1-score yang masing-masing juga sebesar 76%. Namun, ketika diuji pada data *testing*, terjadi sedikit penurunan performa, dengan akurasi sebesar 72% dan nilai metrik lainnya yang juga sebesar 72% sebagaimana yang terlihat pada tabel 2. Hal ini menunjukkan bahwa model SVM Linear memiliki keterbatasan dalam menangkap pola kompleks dari data obesitas, terutama karena data tidak mengalami proses normalisasi dan *balancing*, sehingga kemungkinan besar model lebih rentan terhadap ketidakseimbangan kelas dan perbedaan skala fitur.

Tabel 2. Evaluasi Testing Model SVM Linear

Akurasi	72%
Presisi	72%
Recall	72%
F1 Score	72%

3.2 Evaluasi Model SVM-RBF dengan Hyperparameter tuning

Pada model kedua, yaitu SVM dengan kernel RBF, dilakukan beberapa tahap *preprocessing* sebelum pelatihan model. Proses *preprocessing* meliputi penanganan *missing value*, *balancing* dataset menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE) dimana dataset training terdiri dari 1288 data, dengan distribusi kelas yang tidak seimbang, di mana kelas terbesar memiliki 526 data dan kelas terkecil hanya 58 data. Oleh karena itu, diperlukan penyeimbangan jumlah data. Setelah proses penyeimbangan dilakukan, setiap kelas akan memiliki 526 data, sehingga data training menjadi 2104 data. Selain itu normalisasi fitur pada dataset dengan Min-Max Scaler. Setelah itu, dilakukan tuning hyperparameter pada parameter C dan gamma (γ) untuk memperoleh kombinasi terbaik yang memberikan performa optimal. Didapati bahwa nilai terbaik untuk parameter C adalah 10 dan parameter gamma (γ) adalah 1. Pemilihan parameter C dan gamma dioptimalkan dengan teknik *Grid Search* sebagaimana tertera pada tabel 3. Hasil evaluasi menunjukkan peningkatan yang signifikan dibandingkan dengan model SVM Linear. Pada data *training*, model ini mencapai akurasi sebesar 96% dengan nilai presisi, *recall*, dan F1-score yang juga sebesar 96%, menunjukkan bahwa model mampu mengklasifikasikan data *training* dengan sangat baik.

Tabel 3. Kode Program Pemilihan Parameter C dan Gamma

```
# Grid search untuk hyperparameter
tuning
param_grid = {
```

```
'C': [0.01, 0.1, 1, 10, 100,
1000],
'gamma': [0.0001, 0.001, 0.01,
0.1, 1, 10]
}

svc = SVC(kernel='rbf',
class_weight='balanced')
grid_search = GridSearchCV(svc,
param_grid, cv=10, verbose=1,
scoring='f1_weighted')
grid_search.fit(X_train_scaled,
y_train_smote)

print("Best Parameters:",
grid_search.best_params_)

# Model terbaik berdasarkan pemilihan
param C dan gamma pada Grid Search
best_svc = grid_search.best_estimator_
best_svc.fit(X_train_scaled,
y_train_smote)
```

Pada data *testing*, SVM-RBF tetap menunjukkan performa yang lebih unggul dibandingkan model SVM Linear, dengan akurasi sebesar 83% dan nilai presisi, *recall*, serta F1-score masing-masing juga sebesar 83% sebagaimana yang terlihat pada tabel 4. Meskipun terjadi sedikit penurunan performa dari data *training* ke data *testing*, hal ini masih dalam batas wajar dan menunjukkan bahwa model tidak mengalami *overfitting* yang berlebihan. Peningkatan performa model ini dapat dikaitkan dengan adanya proses *preprocessing*, yang membantu mengatasi masalah skala fitur dan ketidakseimbangan kelas, serta *hyperparameter tuning* yang memungkinkan model untuk lebih optimal dalam menangkap pola non-linear dalam dataset.

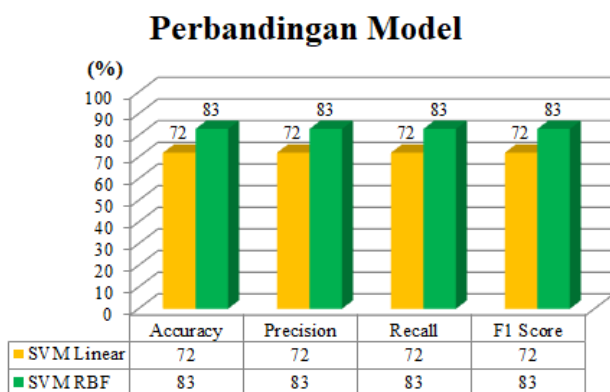
Tabel 4. Evaluasi Testing Model SVM-RBF

Akurasi	83%
Presisi	83%
Recall	83%
F1 Score	83%

3.3 Perbandingan Performa Model

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa model SVM-RBF dengan *preprocessing* dan *hyperparameter tuning* mampu memberikan hasil yang lebih baik dibandingkan dengan SVM Linear tanpa *preprocessing* sebagaimana yang terlihat pada gambar 3. Peningkatan ini terlihat jelas dari nilai akurasi, presisi, *recall*, dan F1-score yang lebih tinggi pada model SVM-RBF. Hal ini menunjukkan bahwa proses *preprocessing* dan *tuning* memiliki dampak signifikan dalam meningkatkan kinerja model klasifikasi.

Perbedaan signifikan dalam performa kedua model ini dapat dikaitkan dengan beberapa faktor. Pertama, penerapan metode *Synthetic Minority Over-sampling Technique* (SMOTE) membantu dalam menangani ketidakseimbangan data, sehingga model dapat belajar dengan lebih baik dari seluruh kelas yang tersedia. Selain itu, penggunaan Min-Max Scaler memastikan bahwa semua fitur memiliki skala yang seragam, yang dapat membantu dalam menghindari dominasi fitur dengan rentang nilai yang lebih besar. Kedua faktor ini berperan dalam meningkatkan efektivitas model dalam menggeneralisasi pola dari data yang diberikan.



Gambar 3. Perbandingan Evaluasi Model SVM-Linear dan SVM-RBF

Selain itu, pemilihan kernel RBF pada model SVM-RBF juga memberikan kontribusi besar terhadap peningkatan performa. Kernel RBF mampu menangkap hubungan non-linear dalam data, yang mungkin tidak dapat ditangani dengan baik oleh SVM Linear. Hal

ini sangat penting terutama dalam kasus klasifikasi yang kompleks, di mana hubungan antar variabel tidak selalu bersifat linear. Dengan tuning hyperparameter seperti nilai C dan gamma (γ), model SVM-RBF dapat menyesuaikan margin keputusan dan kompleksitas model secara optimal, sehingga menghasilkan prediksi yang lebih akurat.

Dengan demikian, pendekatan SVM-RBF dengan *preprocessing* dan tuning hyperparameter lebih direkomendasikan untuk tugas klasifikasi obesitas dibandingkan SVM Linear tanpa *preprocessing*. Proses *balancing* data, normalisasi fitur, dan pemilihan kernel yang tepat terbukti mampu meningkatkan performa model secara signifikan. Oleh karena itu, dalam implementasi sistem klasifikasi serupa, penting untuk mempertimbangkan tahapan *preprocessing* dan pemilihan hyperparameter yang optimal guna memperoleh hasil yang lebih akurat dan andal.

Merujuk pada penelitian terdahulu (Koklu & Sulak, 2024), maka terdapat peningkatan nilai akurasi model SVM dari 74.03% menjadi 83%. Peningkatan yang terjadi dikarenakan adanya proses *preprocessing* SMOTE dan Min-Max Scaler, yang membantu menyeimbangkan data dan meningkatkan distribusi fitur, serta pemilihan kernel RBF, yang lebih optimal dalam menangani data non-linear. Kedua teknik ini tidak diterapkan dalam penelitian sebelumnya, sehingga berkontribusi pada peningkatan akurasi model sebesar 8.97%.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa penggunaan algoritma Support Vector Machine (SVM) dengan kernel RBF yang telah melalui *preprocessing* data dan *hyperparameter tuning* memberikan hasil yang lebih akurat dibandingkan SVM Linear tanpa *preprocessing*. *Preprocessing* data seperti penanganan *missing value*, *balancing* dataset menggunakan SMOTE, dan normalisasi fitur dengan Min-Max Scaler terbukti berkontribusi dalam meningkatkan performa model. Selain itu, *hyperparameter*

tunning pada SVM-RBF dengan memilih nilai optimal untuk parameter C, yaitu 10 dan gamma (γ), yaitu 1 memungkinkan model menangkap pola kompleks dalam data obesitas dengan lebih baik. Dengan demikian, pendekatan SVM-RBF dapat menjadi solusi yang lebih efektif dalam klasifikasi obesitas dibandingkan dengan SVM Linear yang tidak dioptimalkan.

4.2. Saran

Dalam penelitian ini, pemrosesan data dan hyperparameter tuning telah terbukti meningkatkan performa model, namun masih terdapat beberapa aspek yang dapat diperbaiki. Salah satu tantangan yang dapat diperhatikan oleh peneliti selanjutnya adalah eksplorasi lebih lanjut terhadap parameter C dan gamma (γ) dengan metode optimasi lainnya. Selain itu, penggunaan dataset yang lebih besar dan lebih beragam dari berbagai populasi dapat membantu meningkatkan generalisasi model.

5. DAFTAR PUSTAKA

- Arora, N., Singh, A., Al-Dabagh, M. Z. N., & Maitra, S. K. (2022). A Novel Architecture for Diabetes Patients' Prediction Using K -Means Clustering and SVM. *Mathematical Problems in Engineering*, 2022. <https://doi.org/10.1155/2022/4815521>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Dianingsih, A., Dieny, F. F., Nuryanto, N., & Syauqy, A. (2022). Pengaruh Senam Aerobik Terhadap Kualitas Tidur Dan Sindrom Makan Malam Pada Mahasiswa Obesitas. *Gizi Indonesia*, 45(2), 197–208. <https://doi.org/10.36457/gizindo.v45i2.735>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. Elsevier. ISBN: 978-0-12-381479-1
- Javeed, A., Dallora, A. L., Berglund, J. S., Idrisoglu, A., Ali, L., Rauf, H. T., & Anderberg, P. (2023). Early Prediction of Dementia Using Feature Extraction Battery (FEB) and Optimized Support Vector Machine (SVM) for Classification. *Biomedicines*, 11(2). <https://doi.org/10.3390/biomedicines11020439>
- Kementerian Kesehatan Republik Indonesia. (2023). *Panduan Hari Obesitas Sedunia Tahun 2023*. Jakarta: Direktorat Pencegahan dan Pengendalian Penyakit Tidak Menular, Ditjen Pencegahan dan Pengendalian Penyakit.
- Koklu, N., & Sulak, S. A. (2024). Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities. *Sinop Üniversitesi Fen Bilimleri Dergisi*, 9(1), 217–239. <https://doi.org/10.33484/sinopfbid.1445215>
- Reza, M. S., Hafsha, U., Amin, R., Yasmin, R., & Ruhi, S. (2023). Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset. *Computer Methods and Programs in Biomedicine Update*, 4(August), 100118. <https://doi.org/10.1016/j.cmpbup.2023.100118>
- Toar, J., Telew, A., & Lumenta, G. (2023). The Perbedaan Tiga Kategori Aktifitas Fisik pada Status Obesitas dan Non Obesitas. *HIGEIA (Journal of Public Health Research and Development)*, 7(3), 458–467. <https://doi.org/10.15294/higeia.v7i3.61369>
- Wiardani, N. K., Agung, A., Kusumajaya, N., Kesehatan, P., & Denpasar, K. (2023). GIZI INDONESIA Journal of The Indonesian Nutrition Association PERILAKU MAKAN, AKTIVITAS FISIK, DAN PENGGUNAAN INTERNET PADA REMAJA SEKOLAH YANG MENGALAMI OBESITAS DI PROVINSI BALI Eating

Behavior, Physical Activity, and Internet Use in Junior High School. *Gizi Indon*, 2023(2), 207–220.
<https://doi.org/10.36457/gizindo.v46i2.794>

Wong, J. E., Yamaguchi, M., Nishi, N., Araki, M., & Wee, L. H. (2022). Predicting Overweight and Obesity Status among Malaysian Working Adults with Machine Learning or Logistic Regression: Retrospective Comparison Study. *JMIR Formative Research*, 6(12).
<https://doi.org/10.2196/40404>