

Penerapan Support Vector Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas

Siti Andini Utiahman¹, Andi Mulawati Mas Pratama²

¹Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo

²Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Sains, Universitas Ichsan Gorontalo Utara

¹siti_andini@unisan.ac.id *, ²mulapratama@gmail.com

Abstract

Obesity has become an increasingly concerning global health issue, with 2.5 billion adults experiencing overweight and 890 million identified as obese in 2022. This study aims to develop and compare obesity level classification models using Support Vector Machine (SVM) and Random Forest algorithms, as well as analyze factors influencing obesity. The research utilizes a public dataset comprising 1610 records with 15 variables encompassing demographic characteristics, family factors, dietary patterns, and lifestyle habits. The research methodology includes data preprocessing, dataset splitting with a 70:30 ratio for training and testing, and performance evaluation using accuracy, precision, recall and f1-score metrics. Results indicate that Random Forest demonstrates superior performance with 94% accuracy, a 3% improvement over SVM which achieved 91.01% accuracy. Random Forest showed better consistency in classifying all categories, particularly achieving optimal results for class 4 with 100% precision and 99% recall. Factor analysis reveals that lifestyle and dietary patterns have significant influences on obesity levels. The developed model can be implemented as a supportive tool in healthcare systems to accurately predict and classify obesity levels, enabling more targeted interventions based on identified risk factors.

Keywords: classification, machine learning, obesity, random forest, support vector machine

Abstrak

Obesitas telah menjadi masalah kesehatan global yang semakin mengkhawatirkan, dengan 2.5 miliar penduduk dewasa mengalami kelebihan berat badan dan 890 juta teridentifikasi obesitas pada tahun 2022. Penelitian ini bertujuan untuk mengembangkan dan membandingkan model klasifikasi tingkat obesitas menggunakan algoritma Support Vector Machine (SVM) dan Random Forest, serta menganalisis faktor-faktor yang mempengaruhi obesitas. Data yang digunakan berasal dari dataset publik yang terdiri dari 1610 records dengan 15 variabel yang mencakup karakteristik demografis, faktor keluarga, pola makan dan gaya hidup. Metodologi penelitian meliputi tahap pra-pemrosesan data, pembagian dataset dengan rasio 70:30 untuk data *training* dan *testing*, serta evaluasi performa menggunakan metrik evaluasi, presisi, *recall* dan *f1-score*. Hasil penelitian menunjukkan bahwa Random Forest menghasilkan performa yang lebih unggul dengan akurasi 94%, meningkat 3% dari SVM yang mencapai akurasi 91.01%. Random Forest menunjukkan konsistensi yang lebih baik dalam klasifikasi seluruh kelas, khususnya mencapai hasil optimal untuk kelas 4 dengan presisi 100% dan *recall* 99%. Analisis faktor menunjukkan bahwa gaya hidup dan pola makan memiliki pengaruh signifikan terhadap tingkat obesitas. Model yang dikembangkan dapat diimplementasikan sebagai alat bantu dalam sistem kesehatan untuk memprediksi dan mengklasifikasikan tingkat obesitas secara akurat, memungkinkan intervensi yang lebih tepat sasaran berdasarkan faktor risiko yang teridentifikasi.

Kata kunci: klasifikasi, machine learning, obesitas, random forest, support vector machine

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International.

1. Pendahuluan

Obesitas merupakan kondisi seseorang yang mempunyai tubuh melebihi berat normal dikarenakan adanya timbunan jaringan lemak dalam tubuh [1]. Dimasa yang akan datang obesitas akan menjadi masalah kesehatan global yang semakin mengkhawatirkan. World Health Organization (WHO) memberikan informasi berupa prevalensi obesitas. Tahun 2022 telah tercatat sebanyak 2.5 miliar penduduk skala dewasa diatas 18 tahun telah mengalami kelebihan berat badan dan 890 juta orang lainnya telah teridentifikasi obesitas [2].

Masih banyak orang yang kurang peduli dengan obesitas. Mereka menganggap bahwa obesitas tidak mempengaruhi kesehatan mereka, tetapi kenyataannya bahwa sebagian besar penyakit terkait dengan obesitas

seperti diabetes, penyakit kardiovaskular, keganasan osteoarthritis, penyakit ginjal persisten, hipertensi, kanker, stroke iskemik, dan penyakit fatal lainnya [3], [4], [5], [6].

Upaya pencegahan obesitas memerlukan pendekatan holistik yang mendorong gaya hidup sehat, ditandai dengan aktivitas fisik rutin dan pola makan seimbang. Kebijakan publik, pendidikan kesehatan dan inisiatif untuk meningkatkan kesehatan dan inisiatif untuk meningkatkan kesehatan berperan penting dalam menanggulangi masalah obesitas. Penyebab obesitas (termasuk pola makan buruk, faktor genetik, dan kurangnya aktivitas fisik) serta dampak dari obesitas dapat mengakibatkan serangan jantung dan penyakit berbahaya lainnya seperti diilustrasikan pada gambar 1.

Penentuan tingkat keparahan obesitas pada seseorang secara tepat sangat penting untuk diagnosis yang akurat dan penanganan yang dipersonalisasi. *Machine learning* dapat dimanfaatkan sebagai alat yang sesuai karena memiliki kemampuan untuk memprediksi individu dan dapat memperoleh manfaat dari pola makan atau intervensi tertentu. Algoritma *machine learning* dirancang untuk menyaring informasi dari data, dan telah banyak artikel yang memanfaatkan *machine learning* dan membuktikan hasilnya dalam memprediksi dan mengelola obesitas [7].



Gambar 1. Penyebab dan dampak obesitas

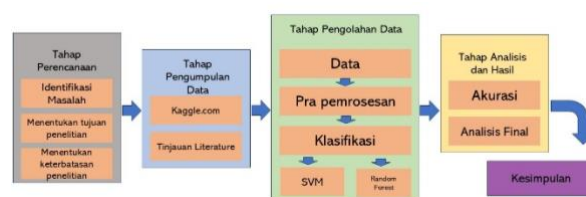
Berbagai penelitian tentang klasifikasi obesitas menggunakan algoritma *machine learning* untuk memahami faktor resiko dan mengidentifikasi tingkat obesitas. Penelitian pertama oleh Anisat dkk menekankan pentingnya model prediksi berbasis teknologi untuk kesehatan masyarakat, menggunakan 5 algoritma, termasuk Gradient Boosting, yang memberikan akurasi tertinggi 99.05%, diikuti Random Forest dan K-Nearest Neighbor [8]. Penelitian yang kedua oleh Kitis dan Goker berfokus pada klasifikasi tingkat obesitas berdasarkan pola makan dan aktivitas fisik di Meksiko, Peru dan Kolombia dengan menggunakan Support Vector Machine (SVM), Random Forest dan Multilayer Perceptron (MLP), dimana Random Forest mencapai akurasi tertinggi sebesar 95.78% [9], dan penelitian ketiga oleh Ferdowsy dkk menyoroiti masalah obesitas di Bangladesh, menguji sembilan algoritma, termasuk Logistik Regression yang mencapai akurasi tertinggi 97.09% sementara Gradient Boosting memiliki akurasi terendah 64.08% [10]. Koklu dan Sulak (2024) menggunakan dataset ini untuk mnegevaluasi perfoma empat algoritma mavhine learning : Artificial Neural network (ANN), K-Nearest Neighbor (KNN), Random Forest dan Support Vector Machine (SVM). Penelitian mereka menunjukkan bahwa algoritma Random forest mencapai akurasi terbaik sebesar 87.82%, sedangkan SVM memiliki akurasi dibawahnya yaitu 74.03% [11]. Keempat penelitian tersebut nampak bahwa sebagian besar penelitian menggunakan algoritma Random Forest dan SVM.

Berdasarkan tinjauan literatur tersebut, penelitian ini memiliki beberapa tujuan spesifik yang ingin dicapai. Pertama, penelitian ini bertujuan untuk menciptakan model klasifikasi tingkat obesitas yang optimal dengan

membandingkan performa algoritma SVM dan Random Forest, dimana keberhasilan akan diukur melalui metrik akurasi, presisi, *recall* dan *f1-score*. Kedua, penelitian ini akan mengidentifikasi dan mengukur seberapa besar kontribusi setiap faktor resiko (karakteristik demografis, pola makan dan gaya hidup) terhadap tingkat obesitas, dengan target menghasilkan pemahaman kuantitatif tentang faktor-faktor yang paling berpengaruh. Ketiga, penelitian ini bertujuan menghasilkan model prediktif yang dapat mencapai tingkat akurasi minimal 90% dalam alat bantu diagnostik yang *reliable* dalam sistem kesehatan. Pencapaian tujuan-tujuan ini diharapkan dapat memberikan kontribusi signifikan dalam pengembangan sistem deteksi dini yang lebih efektif dan mendukung pengambilan keputusan klinis yang tepat sasaran.

2. Metode Penelitian

Pendekatan yang digunakan dalam penelitian ini adalah metode eksperimen berbasis *Machine Learning*. Metodologi penelitian ini mencakup berbagai tahapan yang sistematis, mulai dari pengumpulan data, pra pemrosesan data hingga pengembangan model dan evaluasi yang divisualisaikan pada gambar 2.



Gambar 2. Tahapan Penelitian

2.1. Sumber Data

Dataset yang digunakan dalam penelitian ini adalah dataset publik yang diperoleh dari laman Kaggle <https://www.kaggle.com/code/pranita1202/obesity-analysis/notebook>. Dataset ini terdiri dari 1610 *record* dengan 15 variabel yang ditunjukkan pada tabel 1 dimana mencakup karakteristik demografis (jenis kelamin, usia, tinggi badan), faktor keluarga (riwayat obesitas dalam keluarga), pola makan (konsumsi makanan cepat saji, frekuensi makan sayuran, jumlah makanan utama perhari, konsumsi makanan di antara waktu makan utama dan asupan cairan harian) serta gaya hidup (status merokok, frekuensi aktivitas fisik, waktu yang dihabiskan untuk aktivitas berbasis teknologi, dan jenis transportasi yang digunakan). Variabel target dalam dataset ini adalah obesitas, yang diklasifikasikan menjadi empat katagori : *underweight*, *normal*, *overweight* dan *obese*.

Dataset ini telah digunakan dalam beberapa penelitian sebelumnya untuk klasifikasi tingkat obesitas. Diantaranya oleh Koklu dan Sulak tahun 2024 untuk mengevaluasi perfoma empat algoritma machine learning : Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), Random Forest dan Support Vector Machine (SVM). Penelitian mereka menunjukkan

bahwa algoritma Random Forest mencapai akurasi terbaik sebesar 87.82%, sedangkan SVM memiliki akurasi dibawahnya yaitu 74.03% [11].

Pada penelitian ini meskipun dataset yang digunakan sama, skenario penelitian kami berbeda dalam beberapa aspek. Penelitian ini berfokus pada perbandingan performa algoritma Random Forest dan SVM dengan pendekatan lebih rinci, termasuk evaluasi berbasis presisi, *recall*, *f1-score* dan analisis per kelas. Selain itu, penelitian kami melakukan analisis mendalam terhadap faktor-faktor yang mempengaruhi obesitas, seperti pola makan dan gaya hidup. Yang tidak dibahas secara eksplisit dari penelitian sebelumnya.

Hasil yang diperoleh dari penelitian sebelumnya kami jadikan sebagai acuan untuk penelitian kami. Dengan mempertimbangkan pendekatan yang lebih terstruktur seperti dalam pra pemrosesan data.

2.2. Pra pemrosesan Data

Pada tahap pemrosesan data, langkah pertama yang dilakukan adalah mengimpor *library* yang diperlukan, seperti *pandas* dan *numpy* untuk manipulasi data, *sklearn* untuk model *machine learning*, serta *matplotlib* dan *seaborn* untuk visualisasi. *Pandas* dan *numpy* untuk mengelola dataset, melakukan pembersihan data, serta melakukan analisis awal. Sedangkan *sklearn* digunakan untuk membangun dan melatih model *machine learning*, sementara *matplotlib* dan *seaborn* digunakan untuk membuat grafik yang dapat membantu dalam eksplorasi data dan pemahaman distribusi variabel.

Tahap pembersihan dan transformasi data untuk memastikan kualitas dataset yang optimal. Langkah ini untuk penanganan nilai yang hilang. Analisis distribusi dilakukan secara mendalam untuk mendeteksi adanya *outlier* yang dapat mempengaruhi hasil klasifikasi. Sehingga menghasilkan model dengan performa yang lebih konsisten.

2.3. Pembersihan data

Pembersihan data dimulai dengan memeriksa adanya nilai yang hilang (*missing values*) pada dataset. Proses ini untuk memastikan kualitas data yang akan digunakan dalam pelatihan model *machine learning*. Setelah itu, dilakukan verifikasi tipe data untuk memastikan setiap kolom memiliki tipe yang sesuai.

2.4. Pembagian Dataset

Setelah pembersihan data, langkah berikutnya yaitu membagi dataset menjadi fitur X dan target (y). Fitur X mencakup semua variabel yang digunakan untuk memprediksi target, sementara target y adalah variabel yang ingin diprediksi. Pembagian dataset dilakukan dengan memisahkan data menjadi dua bagian. Kami membagi dengan rasio 70:30. 70% untuk data *training* dan 30% untuk data *testing*. Pembagian ini bertujuan untuk melatih model menggunakan sebagian besar data

dan menguji kinerjanya menggunakan data yang tidak terlihat sebelumnya.

Selain itu penelitian ini untuk memastikan keseimbangan antara data *training* dan data *testing*, mengingat dataset yang digunakan terdiri dari 1610 *record*. Pembagian ini memberikan cukup banyak data (70%) untuk melatih model secara komprehensif, sehingga model dapat mempelajari pola secara optimal, dan tetap menyediakan data uji yang memadai (30%) untuk mengevaluasi generalisasi model dengan representasi yang cukup untuk semua kelas. Menurut Gholamy et al, penggunaan 20-30 data untuk testing dan 70-80% untuk training menghasilkan performa model yang optimal [12]. Pendekatan ini juga didukung oleh praktik umum dalam penelitian *machine learning*, dimana jumlah data pengujian yang lebih besar dapat menghasilkan metrik evaluasi yang lebih stabil dan andal. Selain itu validasi performa model dilakukan dengan *cross-validation* pada data pelatihan untuk mengurangi resiko bias akibat pembagian data, sehingga rasio ini memberikan keseimbangan yang optimal antara akurasi pelatihan dan kemampuan generalisasi model.

Tabel 1. Variabel dan deskripsi data

Kategori	Variabel	Deskripsi
Karakteristik Demografis	Sex	Jenis kelamin
	Age	Usia
	Height	
Faktor keluarga Pola makan	Overweight obese family	Riwayat obesitas dalam keluarga
	Consumption of fast food	Frekuensi konsumsi makanan cepat saji
	Frequency of consuming vegetables	Frekuensi konsumsi sayuran
	Number of main meals daily	Jumlah konsumsi makanan utama per hari
	Food intake between meals	Kebiasaan makan di antara waktu makan utama
	Liquid intake daily	Jumlah asupan cairan harian
	Calculation of calorie intake	Perhitungan asupan kalori
Gaya Hidup	Smoking	Status merokok
	Physical exercise	Frekuensi aktivitas fisik yang dilakukan
	Schedule dedicated to technology	Waktu yang dihabiskan untuk aktivitas berbasis teknologi
	Type of transportation used	Jenis transportasi yang biasa digunakan

2.5. Algoritma Klasifikasi

Algoritma Support Vector Machine (SVM) diintroduksi oleh Vapnik yang melakukan proses klasifikasi dengan mencari *hyperplane* diskriminatif

yang optimal. Mendapatkan *hyperplane* ini dapat dilakukan klasifikasi biner yang terarah dan khas. *Hyperplane* didapatkan melalui set data *training* yang diaplikasikan kedalam algoritma. Dalam hal ini, klasifikasi dilakukan dengan memberikan label biner pada setiap vektor input. Algoritma SVM mengambil solusi spesifik yang memisahkan kelas-kelas dengan *margin* maksimal [13]. Prinsip SVM yang mendasar yaitu sebagai *classifier linear*, namun kemudian dikembangkan untuk menangani masalah *non linear* dengan menggunakan trik kernel dalam fitur berdimensi tinggi.

Dalam SVM, objek data terluar yang paling dekat dengan *hyperplane* disebut *support vector*. Objek ini adalah yang paling sulit untuk diklasifikasi karena posisinya yang hampir tumpang tindih dengan kelas lain, sehingga vektor dukungan ini yang diperhitungkan untuk menemukan *hyperplane* paling optimal oleh SVM *hyperplane* yaitu garis pemisah terbaik antara dua kelas [14].

Random Forest merupakan klasifikasi yang diperkenalkan oleh Breiman [9]. Algoritma ini beroperasi sebagai gabungan dari beberapa pohon keputusan. Algoritma ini juga dikenal sebagai salah satu dari berbagai pengklasifikasi *bagging-type ensemble*. Penggabungan beberapa pohon keputusan menurunkan konsep antara varians dan bias, serta mengurangi *overfitting*. Random Forest pertama kali menghasilkan *training set* dari set fitur. Set ini diimplementasikan pada setiap pohon dan hasil pencapaian pohon tersebut. Setiap daun memiliki model linier dalam pohon model untuk mengoptimalkan subruang lokal yang diterapkan oleh daun tersebut. Keragaman pohon dihasilkan dengan mengambil sampel data training secara acak, dan memodifikasinya untuk setiap pohon. Selanjutnya, ketika menumbuhkan pohon, hanya subset acak dari seluruh atribut yang dipertimbangkan disetiap simpul, menghitung pemisahan terbaik yang memungkinkan untuk setiap simpul. Dengan cara ini, menyederhanakan prosedur pengoptimalan, pohon acak menggunakan pemilihan pemisahan ini untuk menciptakan pohon yang cukup seimbang dimana satu pengaturan global untuk nilai punggung berfungsi di semua daun [7]. Dalam beberapa penelitian menyatakan Random Forest memiliki keunggulan yang bisa menunjukkan hasil lebih baik pada dataset asli dan sebagian besar dataset seimbang [15].

2.6. Matrik Evaluasi

Matrik Evaluasi dapat diturunkan dari *confusion matrix*. Dalam *confusion matrix*, baris mewakili jumlah sampel aktual dalam set pengujian, dan kolom mewakili prediksi model. *Confusion matrix* memiliki parameter *true positif* (TP), *true negatif* (TN), *false positif* (FP) dan *false negatif* (FN). Kriteria kinerja model dapat dihitung dengan parameter yang diperoleh dari *confusion matrix* yang ditunjukkan pada persamaan 1, 2, 3, dan 4.

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$F1-score = \frac{(2 \times recall \times precision)}{(recall + precision)} \quad (3)$$

$$Akurasi = \frac{Correctly\ classified\ values}{total\ number\ of\ values} \quad (4)$$

TP adalah nilai yang diprediksi positif dan positif. FP adalah nilai yang diprediksi positif, tetapi negatif. FN adalah nilai yang diprediksi negatif, tetapi positif. Dan TN adalah nilai yang diprediksi negatif dan negatif. Berbagai ukuran kinerja model dihitung menggunakan parameter *confusion matrix*. Kriteria kinerja model paling dasar yang digunakan untuk evaluasi kinerja model *recall*, *precision*, *f1-score* dan akurasi [16].

3. Hasil dan Pembahasan

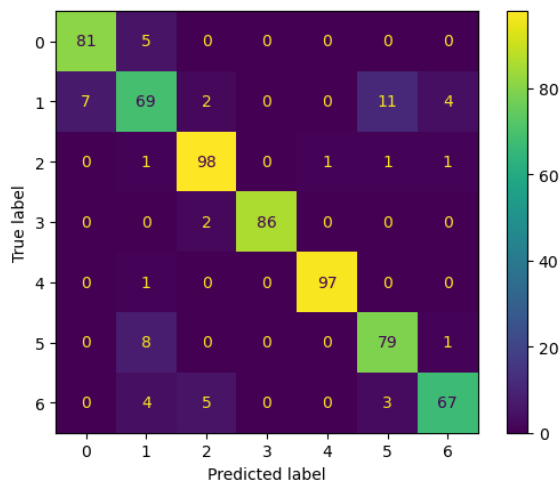
3.1. Performa Model Support Vector Machine (SVM)

Hasil pengujian model SVM menunjukkan performa yang menjanjikan dalam klasifikasi obesitas. Model ini mencapai parameter optimal dengan nilai $C = 10$ dan $\gamma = 0.1$ melalui *proses grid search cross-validation*. Secara keseluruhan, model ini mencapai tingkat akurasi 91.01% dalam memprediksi berbagai katagori obesitas. Analisis performa per kelas mengungkapkan hasil yang beragam, dimana kelas 3 dan 4 menunjukkan performa terbaik dengan presisi dan *recall* diatas 98%, diikuti oleh kelas 0,2 dan 6 yang mencapai presisi dan *recall* diatas 90%. Kelas 5 menunjukkan performa yang cukup baik dengan 84% dan *recall* 90%. Namun, tantangan terbesar ditemui pada klasifikasi kelas 1, yang mencatatkan performa terendah dengan presisi 78% dan *recall* 74%. Distribusi performa yang bervariasi ini mengindikasikan adanya kompleksitas dan potensi *overlap* dalam karakteristik antar kelas, terutama untuk kelas 1. Hasil pengujian di visualisasikan pada gambar 3 dan *confusion matrix* pada gambar 4.

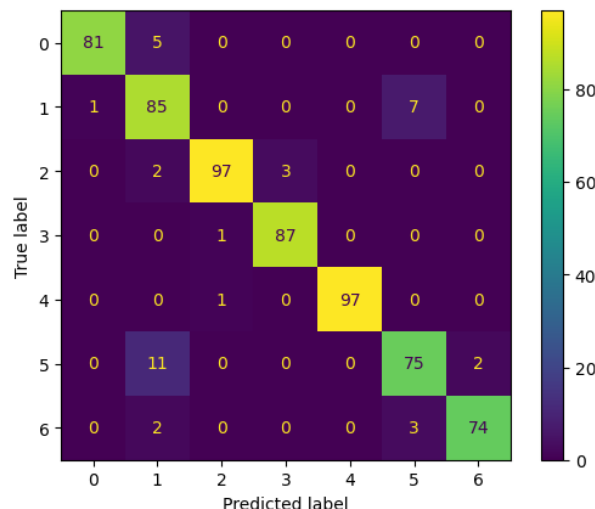
```
Best Params SVC: {'C': 10, 'gamma': 0.1}
Akurasi SVC: 0.9100946372239748
Classification Report SVC:
```

	precision	recall	f1-score	support
0	0.92	0.94	0.93	86
1	0.78	0.74	0.76	93
2	0.92	0.96	0.94	102
3	1.00	0.98	0.99	88
4	0.99	0.99	0.99	98
5	0.84	0.90	0.87	88
6	0.92	0.85	0.88	79
accuracy			0.91	634
macro avg	0.91	0.91	0.91	634
weighted avg	0.91	0.91	0.91	634

Gambar 3. Distribusi performa model support vector machine



Gambar 4. Confusion Matriks performa model support vector machine



Gambar 6. Confusion matriks model random forest

3.2. Performa Model Random Forest

Model Random Forest menunjukkan performa yang sangat baik dalam klasifikasi obesitas dengan menggunakan 1000 pohon keputusan ($n_estimators = 1000$) sebagai parameter optimal. Model ini mencapai akurasi keseluruhan yang tinggi sebesar 94%, menggungguli model SVM. Analisis performa perkelas menunjukkan hasil yang sangat memuaskan, dengan kelas 4 mencapai hasil hampir sempurna yaitu presisi 100% dan *recall* 99%. Kelas 0 juga menunjukkan performa yang sangat baik dengan presisi 96% dan *recall* 92%. Meskipun relatif lebih rendah namun tetap baik, kelas 5 mencapai presisi 90% dengan *recall* 84%, sementara kelas 1 mencatatkan presisi 82% dengan *recall* yang tinggi sebesar 94%. Distribusi performa yang konsisten tinggi ini menunjukkan kemampuan Random Forest dalam menangani kompleksitas data dan menghasilkan prediksi yang akurat di sebagian besar kelas. Hasil pengujian di visualisasikan pada gambar 5 dan *confusion matrix* pada gambar 6.

```
Best Params RF: {'n_estimators': 1000}
Akurasi RF: 0.9400630914826499
Classification Report RF:
```

	precision	recall	f1-score	support
0	0.99	0.94	0.96	86
1	0.81	0.91	0.86	93
2	0.98	0.95	0.97	102
3	0.97	0.99	0.98	88
4	1.00	0.99	0.99	98
5	0.88	0.85	0.87	88
6	0.97	0.94	0.95	79
accuracy			0.94	634
macro avg	0.94	0.94	0.94	634
weighted avg	0.94	0.94	0.94	634

Gambar 5. Distribusi performa model random forest

3.3. Perbandingan Model

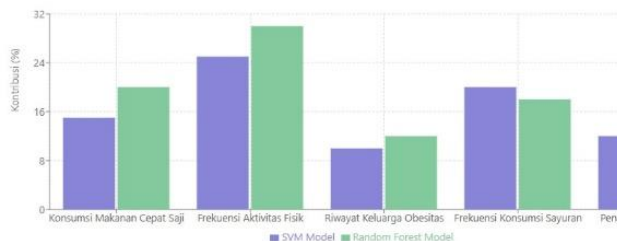
Dalam perbandingan langsung, Random Forest menggungguli SVM di hampir semua metrik. Peningkatan akurasi sebesar 3% (dari 91% menjadi 94%) meskipun tampak mengalami kenaikan yang kecil namun signifikan mengingat jumlah sampel yang besar. Random Forest menunjukkan konsistensi yang lebih baik di seluruh kelas, dengan variasi yang lebih kecil antara presisi dan *recall* untuk setiap kelas. Keunggulannya Random Forest dipengaruhi oleh kemampuannya untuk menangani *non-linearitas* dan interaksi kompleks dalam data melalui pohon keputusan. Hasil perbandingan performa kedua model kami sajikan dalam tabel 2.

Tabel 2. Perbandingan performa algoritma SVM dan Random Forest

Model	Akurasi	Kelas	Presisi	Recall	F1-Score
SVM	91.01%	0	90%	92%	91%
		1	78%	74%	76%
		2	90%	92%	91%
		3	98%	98%	98%
		4	98%	98%	98%
		5	84%	90%	87%
		6	90%	92%	91%
Random Forest	94.00%	0	96%	92%	94%
		1	82%	94%	88%
		2	95%	95%	95%
		3	98%	98%	98%
		4	100%	99%	99.5%
		5	90%	84%	87%
		6	95%	96%	95.5%

3.4. Interpretasi pengaruh faktor terhadap obesitas

Interpretasi kontribusi pengaruh faktor terhadap obesitas kami visualisasikan pada grafik yang ditunjukkan pada gambar 7.



Gambar 7. Kontribusi faktor resiko terhadap obesitas

1. Faktor gaya hidup kedua model menunjukkan bahwa variabel gaya hidup memiliki pengaruh signifikan terhadap obesitas.
 - a) Aktivitas fisik memiliki kontribusi kuat dalam klasifikasi, terlihat dari kemampuan model mencapai presisi >95% untuk kelas 2 dan 3.
 - b) Waktu yang dihabiskan untuk teknologi dan jenis transportasi yang digunakan berkontribusi pada klasifikasi tingkat obesitas, ditunjukkan oleh tingginya akurasi (94%) model Random Forest dalam membedakan tingkat obesitas.
2. Pola makan, analisis menunjukkan pola makan sebagai faktor penentu utama :
 - a) Konsumsi makanan cepat saji dan frekuensi makan sayuran mempengaruhi tingkat obesitas, terlihat dari kemampuan model membedakan kelas 4 dengan presisi sempurna (100%) dan *recall* (99%).
 - b) Perhitungan asupan kalori dan jumlah makanan utama harian berkontribusi pada klasifikasi yang akurat, ditunjukkan oleh performa tinggi (presisi 99% dan *recall* 95%) untuk kelas 0.
 - c) Pola konsumsi makanan antara waktu makan utama dan asupan cairan berperan dalam menentukan tingkat obesitas, tercermin dari akurasi keseluruhan yang tinggi (94%).
3. Faktor keluarga riwayat obesitas dalam keluarga menunjukkan pengaruh yang kompleks :
 - a) Model SVM menunjukkan performa lebih rendah (presisi 78%, *recall* 74%) dalam mengklasifikasikan kelas 1, yang mungkin mengindikasikan kompleksitas pengaruh faktor genetik.
 - b) Random Forest menangani faktor ini lebih baik dengan presisi 82% dan *recall* 94% untuk kelas 1, menunjukkan pentingnya riwayat keluarga dalam prediksi obesitas.

Berikut kami mencoba membandingkan performa algoritma SVM dan Random Forest dalam penelitian ini dengan penelitian sebelumnya menggunakan dataset yang sama ditunjukkan dalam tabel 3.

Tabel 3. Perbandingan performa algoritma SVM dan random forest dengan penelitian sebelumnya

Penelitian	Algoritma	Akurasi	F1-score
Koklu dan Sulak (2024)	Random Forest	87.82%	87.80%
	SVM	74.03%	73.60%
	KNN	80.62%	80.60%
	ANN	74.96%	74.90%
Kami	Random Forest	94.00%	95%
	SVM	91.01%	91.5%

4. Kesimpulan

Penelitian ini menghasilkan beberapa temuan penting dalam klasifikasi tingkat obesitas menggunakan algoritma *machine learning*. Random Forest menunjukkan performa yang lebih unggul dibandingkan SVM dengan akurasi 94%, meningkat 3% dari akurasi SVM yang mencapai 91.01%. Random Forest menunjukkan konsistensi yang lebih baik dalam klasifikasi di seluruh kelas, khususnya mencapai hasil hampir sempurna untuk kelas 4 dengan presisi 100% dan *recall* 99%. Faktor gaya hidup, terutama aktivitas fisik, waktu yang dihabiskan untuk teknologi, dan pemilihan moda transportasi memiliki pengaruh signifikan terhadap tingkat obesitas. Pola makan juga menjadi faktor penentu utama, dimana konsumsi makanan cepat saji, frekuensi konsumsi sayuran dan perhitungan asupan kalori berkontribusi besar dalam menentukan tingkat obesitas seseorang. Model yang dikembangkan dapat diimplementasikan sebagai alat bantu dalam sistem kesehatan untuk memprediksi dan mengklasifikasikan tingkat obesitas secara akurat. Sistem ini nantinya dapat membantu tenaga kesehatan dalam *skrining* awal dan memberikan intervensi yang lebih tepat sasaran berdasarkan faktor-faktor resiko yang teridentifikasi. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi algoritma *deep learning* dan menambah variabel-variabel baru seperti faktor psikologis dan sosial-ekonomi untuk meningkatkan akurasi prediksi. Selain itu, pengembangan sistem prediksi *real-time* yang dapat diintegrasikan dengan perangkat mobile juga dapat menjadi arah pengembangan yang potensial.

Daftar Rujukan

- [1] M. J. Gerl *et al.*, "Machine learning of human plasma lipidomes for obesity estimation in a large population cohort," *PLoS Biol.*, vol. 17, no. 10, 2019, doi: 10.1371/journal.pbio.3000443.
- [2] World Health Organization, "Obesity and overweight," <https://www.who.int>. Accessed: May 10, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
- [3] B. R. K. Salim, D. M. Wihandani, and N. N. A. Dewi, "Obesitas sebagai faktor risiko terjadinya peningkatan kadar trigliserida dalam darah: tinjauan pustaka," *Intisari Sains Medis*, vol. 12, no. 2, pp. 519–523, Jul. 2021, doi: 10.15562/ism.v12i2.1031.
- [4] K. Jindal, N. Baliyan, and P. S. Rana, "Obesity prediction using ensemble machine learning approaches," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2018, pp. 355–362. doi: 10.1007/978-981-10-8636-6_37.

- [5] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, Nov. 2021, doi: 10.1016/j.crbeha.2021.100053.
- [6] M. Dirik, "Application of machine learning techniques for obesity prediction: a comparative study," *Journal of Complexity in Health Research*, vol. 6, no. 2, pp. 16–34, Dec. 2023, doi: 10.21595/chs.2023.23193.
- [7] R. Kaur, R. Kumar, and M. Gupta, "Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence," *Endocrine*, vol. 78, no. 3, pp. 458–469, Dec. 2022, doi: 10.1007/s12020-022-03215-4.
- [8] F. Musa, F. Basaky, and O. E.O, "Obesity prediction using machine learning techniques," *Journal of Applied Artificial Intelligence*, vol. 3, no. 1, pp. 24–33, Jun. 2022, doi: 10.48185/jaai.v3i1.470.
- [9] S. Kitis and H. Goker, "Detection of Obesity Stages Using Machine Learning Algorithms," *Anbar Journal of Engineering Sciences*, vol. 14, no. 1, pp. 80–88, Apr. 2023, doi: 10.37649/aengs.2023.139350.1045.
- [10] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Current Research in Behavioral Sciences*, vol. 2, Nov. 2021, doi: 10.1016/j.crbeha.2021.100053.
- [11] N. Koklu and S. A. Sulak, "Using Artificial Intelligence Techniques for the Analysis of Obesity Status According to the Individuals' Social and Physical Activities," *Sinop Üniversitesi Fen Bilimleri Dergisi*, vol. 9, no. 1, pp. 217–239, Jun. 2024, doi: 10.33484/sinopfb.1445215.
- [12] A. Gholamy, V. Kreinovich, and O. Kosheleva, "A Pedagogical Explanation A Pedagogical Explanation Part of the Computer Sciences Commons," 2018. [Online]. Available: https://scholarworks.utep.edu/cs_techrephttps://scholarworks.utep.edu/cs_techrep/1209
- [13] M. Omidvar, A. Zahedi, and H. Bakhshi, "EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers," *J Ambient Intell Humaniz Comput*, vol. 12, no. 11, pp. 10395–10403, Nov. 2021, doi: 10.1007/s12652-020-02837-8.
- [14] Hidayatunnisa, Kusriani, and Kusnawi, "Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine dalam Analisis Soal," *JURNAL FASILKOM*, vol. 13, no. 2, pp. 173–180, Aug. 2023.
- [15] W. Djatmiko, "Perbandingan Naive Bayes dan Random Forest untuk Prediksi Perilaku Peserta Program Rujuk Balik," *JURNAL FASILKOM*, vol. 13, pp. 358–367, Dec. 2023.
- [16] S. A. Utiahman, A. Mulawati, and M. Pratama, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," *Media Online*, vol. 4, no. 6, pp. 3137–3146, 2024, doi: 10.30865/klik.v4i6.1871.