



Analisis Faktor Penentu Tingkat Obesitas Berdasarkan *Feature Importance* pada *Random Forest*

Muhammad Fauzan Rifnandy

Program Studi Ilmu Komputer, Universitas Islam Sumatera Utara, Indonesia

*Penulis Korespondensi: fauzanrifnandy111@gmail.com

Abstract. Obesity is a growing health concern that is closely associated with various chronic diseases. Machine learning has been widely applied to classify obesity levels, yet most studies focus on improving predictive performance rather than identifying the features contributing to classification outcomes. This study aims to analyze the determining factors of obesity levels using Feature Importance in a Random Forest model. The research utilized the Estimation of Obesity Levels Based on Eating Habits and Physical Condition dataset from the UCI Machine Learning Repository. After preprocessing, 2,087 records were analyzed. Hyperparameter tuning was performed using GridSearchCV, and model performance was evaluated using Accuracy, Precision, Recall, and F1-Score. The optimized Random Forest achieved an Accuracy of 94.26%, Precision of 94.60%, Recall of 94.26%, and F1-Score of 94.34%. Feature Importance analysis identified Weight as the most influential feature, followed by Age, FCVC, Height, and Gender. The results demonstrate that Random Forest not only provides high classification performance but also helps identify the features that contribute most to obesity level classification.

Keywords: Age; Feature Importance; Obesity Classification; Random Forest; Weight.

Abstrak. Obesitas menjadi salah satu tantangan kesehatan yang semakin banyak ditemukan serta berhubungan dengan berbagai penyakit tidak menular. Berbagai penelitian telah memanfaatkan machine learning untuk mengklasifikasikan tingkat obesitas, sebagian besar studi yang telah dilakukan lebih berorientasi pada evaluasi kinerja model, sementara analisis mengenai pengaruh setiap fitur terhadap hasil klasifikasi masih relatif terbatas. Penelitian ini bertujuan mengidentifikasi faktor-faktor yang berkontribusi terhadap klasifikasi tingkat obesitas menggunakan Feature Importance pada algoritma Random Forest. Dataset yang digunakan adalah Estimation of Obesity Levels Based on Eating Habits and Physical Condition dari UCI Machine Learning Repository. Setelah tahap preprocessing, sebanyak 2.087 data digunakan dalam analisis. Model dioptimasi menggunakan GridSearchCV dan dievaluasi dengan nilai Akurasi, Presisi, Recall, dan F1-Score. Hasil pengujian memperlihatkan bahwasannya Random Forest mendapatkan Akurasi 94,26%, Presisi 94,60%, Recall 94,26%, dan F1-Score 94,34%. Analisis Feature Importance memperlihatkan bahwasannya fitur Weight memiliki kontribusi terbesar dengan nilai 0,289, diikuti oleh Age, FCVC, Height, dan Gender. Temuan penelitian memperlihatkan bahwasannya Random Forest tidak hanya mampu menunjukkan kemampuan klasifikasi yang baik, tetapi juga memberikan gambaran mengenai fitur-fitur yang berperan penting dalam membedakan kategori tingkat obesitas.

Kata kunci: Berat; Hutan Acak; Klasifikasi Obesitas; Pentingnya Fitur; Usia.

1. LATAR BELAKANG

Obesitas semakin menjadi perhatian global seiring dengan meningkatnya jumlah kasus yang terjadi di berbagai wilayah dunia. Kondisi ini tidak hanya berkaitan dengan perubahan komposisi tubuh, tetapi juga berhubungan dengan meningkatnya risiko berbagai penyakit kronis yang dapat menurunkan kualitas hidup masyarakat. Oleh karena itu, obesitas tidak lagi dipandang sebagai masalah penampilan semata, melainkan sebagai tantangan kesehatan yang memerlukan perhatian berkelanjutan (Fitria & Berawi, 2019).

Obesitas diidentifikasi oleh Organisasi Kesehatan Dunia (WHO) sebagai sebuah penyakit kronis yang bisa kambuh, muncul akibat dari interaksi yang rumit antara faktor genetik, neurobiologi, pola makan, akses terhadap makanan sehat, kekuatan pasar, dan konteks

lingkungan yang lebih luas. Dalam beberapa dekade terakhir, obesitas telah menyebar secara global seiring dengan meningkatnya ketahanan pangan dan perkembangan sosial-ekonomi di berbagai negara, serta transformasi pola makan, tingkat aktivitas fisik, dan perilaku masyarakat maupun individu yang dipengaruhi oleh globalisasi dan sistem pangan industri (WHO, 2025).

Kementerian Kesehatan Republik Indonesia menyatakan bahwa obesitas sering kali dianggap sebagai kondisi yang tidak berbahaya, khususnya pada anak, padahal kondisi tersebut dapat meningkatkan risiko gangguan kesehatan pada masa dewasa. Peningkatan prevalensi obesitas yang terjadi dalam beberapa tahun terakhir menunjukkan bahwa obesitas tidak hanya menjadi permasalahan individu, tetapi juga tantangan kesehatan masyarakat yang memerlukan perhatian lebih lanjut (Kemenkes, 2024).

Seiring berkembangnya teknologi machine learning, berbagai penelitian telah dilakukan untuk mengklasifikasikan tingkat obesitas berdasarkan karakteristik individu, pola konsumsi, dan aktivitas fisik. Penelitian yang dilakukan oleh Ismail & Atim (2024) memakai Decision Tree C4.5 untuk mengkategorikan tingkat obesitas dan menghasilkan akurasi sebesar 95,2%. Penelitian yang dikembangkan oleh Valerian et al. (2025) menerapkan Gradient Boosting Machine (GBM) dan memperoleh akurasi sebesar 95% dalam klasifikasi tingkat obesitas. selanjutnya, riset yang di kembangkan Utiahman et al. (2024) turut membuktikan bahwa Random Forest mampu memberikan hasil klasifikasi yang memuaskan pada kasus tingkat obesitas dengan tingkat akurasi mencapai 94% (Utiahman & Pratama, 2024).

Meskipun performa klasifikasi merupakan aspek yang penting, informasi mengenai faktor-faktor yang paling berkontribusi terhadap hasil klasifikasi juga memiliki nilai yang tidak kalah penting. Analisis terhadap kontribusi setiap faktor dapat memberikan pemahaman yang lebih baik mengenai karakteristik yang berkaitan dengan tingkat obesitas, sehingga hasil penelitian tidak hanya mampu menghasilkan model prediksi, tetapi juga bisa memberikan informasi yang lebih mudah diinterpretasikan (Fitriani et al., 2025).

Padahal, identifikasi faktor-faktor yang berkontribusi terhadap tingkat obesitas memiliki peranan penting dalam memberikan pemahaman yang lebih baik mengenai karakteristik individu yang berkaitan dengan obesitas. Informasi tersebut dapat digunakan sebagai dasar dalam pengembangan penelitian lanjutan maupun sebagai bahan pertimbangan dalam upaya pencegahan dan pengendalian obesitas. Oleh karena itu, selain menghasilkan model klasifikasi yang baik, diperlukan pendekatan yang mampu memberikan interpretasi terhadap kontribusi setiap faktor yang digunakan dalam proses klasifikasi (Daka et al., 2025).

Keterbatasan tersebut menunjukkan bahwa evaluasi performa model saja belum cukup untuk memperoleh gambaran yang komprehensif mengenai proses klasifikasi yang dilakukan

algoritma. Diperlukan analisis tambahan yang mampu mengidentifikasi kontribusi setiap fitur sehingga hasil prediksi yang diperoleh menjadi lebih mudah dipahami dan dijelaskan. Karena itu, penelitian ini tidak hanya mengevaluasi performa Random Forest, tetapi juga menganalisis tingkat kontribusi fitur menggunakan *Feature Importance*.

2. KAJIAN TEORITIS

Obesitas

Obesitas merupakan kondisi kesehatan yang terjadi ketika akumulasi lemak tubuh berada pada tingkat yang melebihi kebutuhan fisiologis normal. Kondisi tersebut berkembang akibat ketidakseimbangan antara energi yang masuk melalui konsumsi makanan dan energi yang dikeluarkan melalui aktivitas tubuh dalam jangka waktu yang panjang. Dampak obesitas tidak hanya berkaitan dengan peningkatan berat badan, tetapi juga berhubungan dengan meningkatnya risiko berbagai gangguan kesehatan seperti diabetes, hipertensi, penyakit kardiovaskular, serta gangguan metabolik lainnya. Faktor yang memengaruhi obesitas bersifat kompleks karena melibatkan karakteristik biologis, perilaku individu, pola makan, aktivitas fisik, dan faktor lingkungan (Felndity et al., 2025).

Machine Learning

Machine learning dapat diartikan sebagai pendekatan dalam kecerdasan buatan yang memungkinkan komputer membangun pola pembelajaran berdasarkan data masa lalu tanpa harus diprogram secara eksplisit untuk setiap kondisi (Andhika & Bayunanda, 2025). Melalui proses pelatihan, model dapat mengenali hubungan antarvariabel dan memanfaatkan informasi tersebut untuk melakukan prediksi maupun klasifikasi terhadap data baru. Dalam bidang kesehatan, pendekatan ini banyak digunakan untuk membantu proses analisis data yang berukuran besar dan memiliki karakteristik yang beragam, termasuk dalam identifikasi faktor risiko dan klasifikasi kondisi kesehatan (Wardhana et al., 2023).

Random Forest

Random Forest dapat diartikan sebagai teknik pembelajaran *ensemble* yang beroperasi dengan membangun beberapa pohon keputusan secara terpisah, lalu mengolah hasil ramalan dari setiap pohon untuk mencapai kesimpulan akhir (Pandu et al., 2025). Pendekatan ini mampu meningkatkan stabilitas model karena keputusan tidak ditentukan oleh satu pohon saja, melainkan berdasarkan agregasi banyak pohon. Dengan mekanisme tersebut, Random Forest dikenal memiliki kapabilitas generalisasi yang lebih baik dan juga lebih tahan terhadap *overfitting* jika dibandingkan dengan model pohon keputusan yang hanya satu (Ernawati et al., 2025).

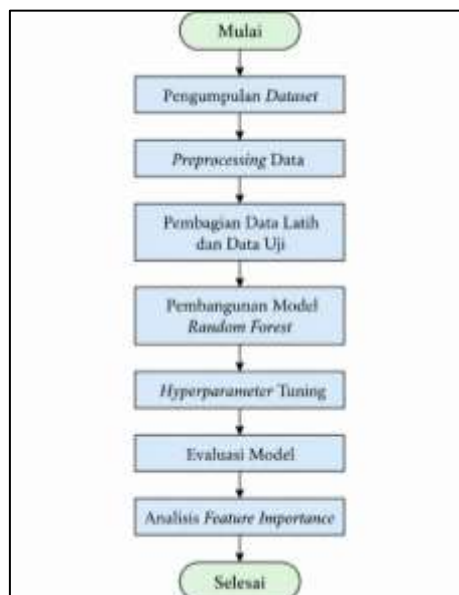
Hyperparameter Tuning

Hyperparameter tuning merupakan proses optimasi parameter model yang dilakukan sebelum proses pelatihan berlangsung. Tujuan utama tahap ini adalah memperoleh konfigurasi parameter yang mampu menghasilkan performa terbaik pada data yang digunakan (Saputri et al., 2025). Pada metode Random Forest, sejumlah parameter yang sering dioptimalkan mencakup total pohon keputusan, kedalaman maksimum setiap pohon, jumlah minimum sampel yang diperlukan untuk memisahkan *node*, serta jumlah minimum sampel yang terdapat pada *node* daun. Memilih kombinasi parameter yang sesuai mampu meningkatkan kemampuan model dalam melakukan prediksi pada data baru (Handyanto et al., 2026).

Feature Importance

Feature Importance merupakan teknik yang dipakai untuk menilai kontribusi setiap fitur terhadap proses pengambilan keputusan pada model *machine learning*. Pada Random Forest, nilai *feature importance* dihitung berdasarkan akumulasi penurunan impurity yang dihasilkan suatu fitur pada seluruh pohon keputusan yang terbentuk (Alfarisi et al., 2025). Semakin tinggi nilai yang diperoleh suatu fitur, semakin besar pula perannya dalam membantu model membedakan kategori target. Informasi ini dapat digunakan untuk mengidentifikasi faktor-faktor yang memiliki kontribusi terbesar terhadap hasil klasifikasi serta meningkatkan interpretasi terhadap model yang dibangun (Juliani & Voutama, 2025).

3. METODE PENELITIAN



Gambar 1. Alur Penelitian.

Pengumpulan Data

Dataset yang diterapkan dalam studi ini adalah Penilaian Tingkat Obesitas Berdasarkan Kebiasaan Makan dan Kondisi Fisik yang diperoleh dari UCI Machine Learning Repository. Dataset terdiri atas 2.111 data dengan 16 atribut prediktor dan 1 atribut target (NObesidad). Variabel target terdiri atas tujuh kategori tingkat obesitas, yaitu berat kurang, berat normal, kelebihan berat tingkat I, kelebihan berat tingkat II, obesitas tipe I, obesitas tipe II, dan obesitas tipe III. Atribut prediktor yang digunakan meliputi karakteristik individu, kebiasaan makan, serta aktivitas fisik, seperti usia (*Age*), tinggi badan (*Height*), berat badan (*Weight*), konsumsi sayuran (FCVC), aktivitas fisik (FAF), konsumsi air (CH2O), dan riwayat keluarga dengan obesitas (*family_history_with_overweight*).

Preprocessing Data

Sebelum proses pembangunan model dilakukan, data terlebih dahulu melalui tahap preprocessing untuk memastikan kualitas dan kesiapan data yang akan dianalisis. Tahap ini mencakup pengecekan terhadap *missing value* serta data duplikat. Hasil analisis menunjukkan bahwa kumpulan data tidak memiliki nilai yang hilang, namun teridentifikasi 24 data yang sama yang selanjutnya dihapus, sehingga total data yang dipakai dalam riset kini berjumlah 2.087 data. Setelah proses pembersihan data selesai, dilakukan pemisahan antara atribut prediktor dan atribut target. Variabel kategorikal selanjutnya diubah dengan metode *One Hot Encoding* agar bisa diterima oleh algoritma Random Forest. Kemudian, dataset tersebut dipisah menjadi data pelatihan dan data pengujian dengan rasio 80:20 menggunakan *strategi stratified sampling* agar proporsi setiap kelas dalam kedua grup data seimbang.

Model Random Forest

Pada studi ini, klasifikasi tingkat obesitas dilaksanakan menggunakan Random Forest, yang dibangun berdasarkan data pelatihan yang sudah menjalani proses preprocessing. Supaya memperoleh hasil yang lebih baik, dilakukan *hyperparameter tuning* dengan GridSearchCV dan mengevaluasi berbagai kombinasi parameter, seperti *n_estimators*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*. Setiap kombinasi parameter dievaluasi menggunakan teknik *cross validation*, lalu kombinasi yang menunjukkan performa terbaik dipilih sebagai model final yang digunakan pada tahap pengujian dan analisis *Feature Importance*.

Evaluasi Model

Evaluasi terhadap model dilakukan dengan tujuan menilai efektivitas Random Forest saat menentukan tingkat obesitas melalui data yang telah diuji. Dalam studi ini, performa model

dinilai dengan memanfaatkan indikator Akurasi, Presisi, *Recall*, dan F1-Score. Akurasi dipakai untuk menilai seberapa besar bagian prediksi yang tepat dibanding dengan keseluruhan data yang diuji. Nilai Akurasi dihitung menggunakan Persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \dots\dots\dots (i)$$

Keterangan:

TP (*True Positive*) yaitu total data yang diprediksi positif dengan akurat.

TN (*True Negative*) yaitu total data yang diprediksi negatif dengan akurat.

FP (*False Positive*) adalah jumlah data yang diprediksi positif tetapi sebenarnya negatif.

FN (*False Negative*) adalah jumlah data yang diprediksi negatif tetapi sebenarnya positif.

Semakin besar nilai akurasi yang diperoleh, semakin unggul kemampuan model dalam mengklasifikasikan tingkat obesitas.

Analisis Feature Importance

Tahap akhir melibatkan analisis *Feature Importance* untuk menentukan seberapa besar sumbangan setiap fitur dalam proses pengklasifikasian yang dilaksanakan oleh model Random Forest. Nilai *Feature Importance* didapat dari total penurunan impuritas yang dibawa oleh setiap fitur di seluruh pohon keputusan yang terbentuk. Semakin tinggi nilai *Feature Importance*, semakin signifikan peran fitur tersebut dalam mempengaruhi hasil klasifikasi. Temuan dari analisis ini dimanfaatkan untuk mengenali elemen-elemen terpenting yang berkontribusi terhadap tingkat kelebihan berat badan berdasarkan model yang telah dikembangkan.

4. HASIL DAN PEMBAHASAN

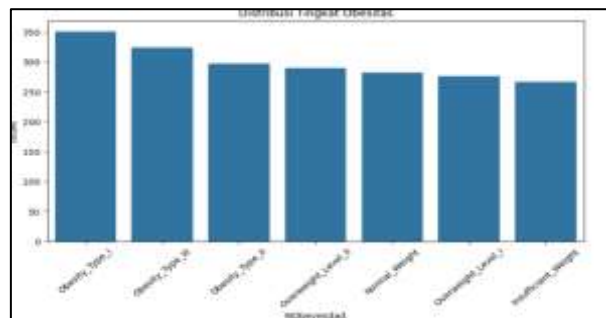
Karakteristik Dataset

Penelitian ini memanfaatkan dataset Estimasi Tingkat Obesitas Berdasarkan Kebiasaan Makan dan Kondisi Fisik yang didapat dari UCI Machine Learning Repository. Dataset awal mencakup 2.111 entri yang memiliki 16 atribut untuk prediksi dan 1 atribut sebagai target (NObesidad). Setelah proses pra-pemrosesan, teridentifikasi 24 entri yang merupakan duplikat, yang kemudian dihilangkan, sehingga total entri yang digunakan dalam penelitian berkurang menjadi 2.087.

Tabel 1. Karakteristik Dataset.

Keterangan	Jumlah
Jumlah data awal	2111
Data duplikat	24
Jumlah data akhir	2087
Jumlah atribut prediktor	16
Jumlah kelas target	7

Variabel target pada dataset terdiri atas tujuh kategori tingkat obesitas, yaitu berat kurang, berat normal, kelebihan berat tingkat I, kelebihan berat tingkat II, obesitas tipe I, obesitas tipe II, dan obesitas tipe III. Setelah proses pembersihan data, seluruh data digunakan dalam proses pelatihan dan pengujian model.

**Gambar 2.** Distribusi Kelas Tingkat Obesitas.

Berdasarkan Gambar 2, distribusi data pada setiap kategori obesitas relatif seimbang. Kondisi ini membantu model dalam mempelajari karakteristik masing-masing kelas tanpa adanya dominasi yang terlalu besar dari kelas tertentu.

Hasil Hyperparameter Tuning

Untuk memperoleh setelan Random Forest yang paling baik, dilakukan penyesuaian *hyperparameter* melalui GridSearchCV dengan metode validasi silang. Parameter yang diuji mencakup jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), jumlah minimal sampel untuk memisahkan *node* (*min_samples_split*), dan jumlah minimal sampel di *node* daun (*min_samples_leaf*) (Haris et al., 2026).

Tabel 2. Parameter Terbaik Random Forest.

Parameter	Nilai Terbaik
<i>n_estimators</i>	200
<i>max_depth</i>	None
<i>min_samples_split</i>	5
<i>min_samples_leaf</i>	1

Evaluasi Model Random Forest

Model terbaik hasil proses *hyperparameter tuning* kemudian kemampuannya dites ke data uji. Evaluasi dilakukan menggunakan indikator Akurasi, Presisi, *Recall*, dan F1-Score.

Table 3. Hasil Evaluasi Model Random Forest.

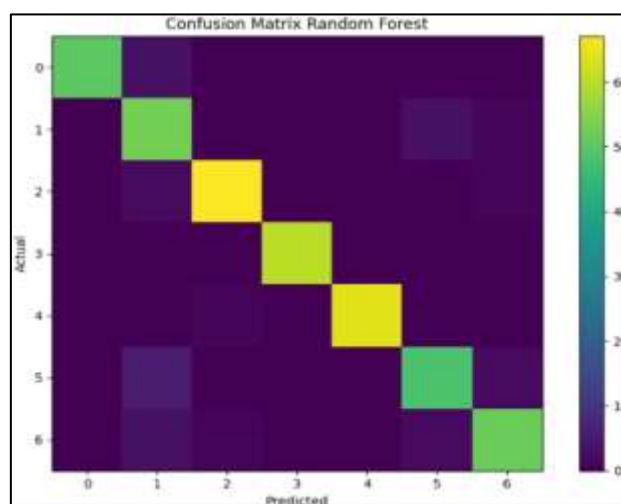
Indikator	Nilai
Akurasi	94,26%
Presisi	94,60%
Recall	94,26%
F1-Score	94,34%

Berdasarkan Tabel 3, Random Forest mencatat tingkat Akurasi sebesar 94,26%. Selain itu, indikator Presisi, *Recall*, dan F1-Score nya adalah 94,60%, 94,26%, dan 94,34%. Temuan ini mengindikasikan bahwa model memiliki kemampuan yang baik dalam mengkategorikan tujuh tingkat obesitas yang terdapat dalam dataset yang dianalisis. Untuk mendapatkan pemahaman yang lebih mendalam tentang hasil klasifikasi di setiap kategori, digunakan laporan klasifikasi yang disajikan dalam Tabel 4.

Tabel 4. Classification Report Random Forest.

Kelas	Presisi	Recall	F1-Score
Berat Kurang	1,00	0,94	0,97
Berat Normal	0,80	0,93	0,86
Obesitas Tipe I	0,97	0,96	0,96
Obesitas Tipe II	1,00	1,00	1,00
Obesitas Tipe III	1,00	0,98	0,99
Kelebihan Tingkat I	0,91	0,87	0,89
Kelebihan Tingkat II	0,93	0,90	0,91

Berdasarkan Tabel 4, model ini memperlihatkan hasil yang sangat memuaskan dalam kategori Obesitas Tipe II dengan indikator Presisi, *Recall*, dan F1-Score mencapai 1,00. Sementara itu, untuk kategori Berat Badan Normal, nilai Presisi-nya tampak lebih rendah jika dibandingkan dengan kategori lainnya. Ini menunjukkan bahwa beberapa data dalam kategori tersebut masih memiliki kesamaan karakteristik dengan kategori berat badan yang berdekatan, yang mengakibatkan adanya kesalahan dalam klasifikasi.

**Gambar 3.** Confusion Matrix Random Forest.

Berdasarkan Gambar 3, nilai pada diagonal utama matriks mendominasi, yang menunjukkan bahwa sebagian besar data berhasil diprediksi secara tepat. Kesalahan klasifikasi masih ditemukan pada beberapa kelas yang memiliki karakteristik serupa, terutama antara kategori Berat Normal, Kelebihan Tingkat I, dan Kelebihan Tingkat II. Namun demikian, jumlah prediksi benar masih jauh lebih besar dibandingkan prediksi yang salah sehingga menghasilkan performa klasifikasi yang tinggi.

Analisis Feature Importance

Untuk mengidentifikasi fitur-fitur yang paling berperan dalam klasifikasi tingkat obesitas, dilakukan analisis *Feature Importance* pada model yang telah dibangun. Nilai *Feature Importance* diperoleh dari model Random Forest terbaik yang dihasilkan pada tahap sebelumnya (Wahyudi et al., 2025).

Tabel 5. Nilai Feature Importance.

Rank	Fitur	Importance
1	Weight	0,289
2	Age	0,088
3	FCVC	0,085
4	Height	0,081
5	Gender	0,076
6	NCP	0,051
7	Family History	0,048
8	CALC	0,047

Berdasarkan Tabel 5, fitur Weight memiliki nilai *Feature Importance* tertinggi sebesar 0,289. Hasil ini menunjukkan bahwa berat badan merupakan faktor yang paling berkontribusi terhadap proses klasifikasi tingkat obesitas pada model yang dibangun. Nilai tersebut jauh lebih tinggi dibandingkan fitur lainnya, sehingga menunjukkan bahwa berat badan menjadi indikator utama dalam membedakan kategori tingkat obesitas. Selain *Weight*, fitur *Age*, *FCVC* (*Frequency of Consumption of Vegetables*), *Height*, dan *Gender* juga menunjukkan kontribusi yang cukup besar. Temuan ini menunjukkan bahwa klasifikasi tingkat obesitas bukan cuma dipengaruhi berdasarkan faktor antropometri, tetapi juga berkaitan dengan karakteristik individu dan pola konsumsi makanan.

Pembahasan

Hasil pengujian memperlihatkan bahwa Random Forest yang telah melalui proses optimasi parameter mampu memberikan performa klasifikasi cukup tinggi pada data obesitas yang digunakan. Nilai Akurasi sebesar 94,26% menunjukkan kalau mayoritas data uji berhasil dikategorikan dengan sesuai. Hasil tersebut mengindikasikan bahwa karakteristik individu, pola konsumsi makanan, dan aktivitas fisik yang tersedia pada dataset memiliki informasi yang cukup untuk membedakan tingkat obesitas.

Analisis *Feature Importance* memberikan informasi tambahan yang tidak dapat diperoleh hanya dari nilai akurasi model. Berdasarkan hasil yang diperoleh, fitur *Weight* menjadi paling dominan dalam proses klasifikasi. Temuan ini menunjukkan bahwa perubahan berat badan memiliki hubungan yang sangat kuat terhadap perbedaan kategori obesitas. Selain itu, fitur *Age*, *FCVC*, *Height*, dan *Gender* juga memberikan kontribusi yang relatif besar, sehingga klasifikasi tingkat obesitas tidak hanya dipengaruhi oleh ukuran tubuh tetapi juga karakteristik individu dan kebiasaan konsumsi makanan.

Berbeda dengan riset sebelumnya yang lebih fokus pada peningkatan performa dan perbandingan algoritma, penelitian ini juga melakukan analisis *Feature Importance* untuk mengidentifikasi faktor yang berkontribusi terhadap hasil klasifikasi. Hasil analisis menunjukkan bahwa *Weight* merupakan faktor dengan kontribusi terbesar, diikuti oleh *Age*, *FCVC*, *Height*, dan *Gender*. Temuan ini menunjukkan bahwa faktor antropometri masih menjadi indikator utama dalam proses klasifikasi tingkat obesitas, namun faktor pola konsumsi dan karakteristik individu juga memberikan kontribusi yang cukup signifikan.

Secara umum, temuan dari penelitian ini mengindikasikan bahwa penggunaan Random Forest tidak hanya menghasilkan tingkat akurasi yang tinggi, namun juga bisa digunakan untuk menemukan elemen-elemen yang memengaruhi prevalensi obesitas melalui *analisis Feature Importance*. Data tersebut bisa memberikan wawasan yang lebih mendalam tentang atribut yang digunakan oleh model untuk membedakan antara kategori tingkat obesitas dan menjadi landasan bagi penelitian berikutnya.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma Random Forest yang dioptimasi menggunakan hyperparameter tuning mampu mengklasifikasikan tingkat obesitas dengan baik pada dataset yang digunakan. Model menghasilkan nilai Akurasi sebesar 94,26%, Presisi sebesar 94,60%, *Recall* sebesar 94,26%, dan *F1-Score* sebesar 94,34%. Hal tersebut memperlihatkan bahwa Random Forest memiliki performa yang baik dalam membedakan tujuh kategori tingkat obesitas berdasarkan karakteristik individu, pola konsumsi makanan, dan aktivitas fisik.

Hasil analisis *Feature Importance* memperlihatkan bahwa fitur *Weight* merupakan fitur yang memiliki kontribusi terbesar terhadap proses klasifikasi dengan nilai importance sebesar 0,289. Selain itu, fitur *Age*, *FCVC*, *Height*, dan *Gender* juga memiliki kontribusi yang relatif tinggi dalam proses klasifikasi tingkat obesitas. Temuan ini menunjukkan bahwa model tidak

hanya memanfaatkan faktor antropometri, tetapi juga mempertimbangkan karakteristik individu dan pola konsumsi makanan dalam membedakan kategori tingkat obesitas.

Penelitian ini tentunya mempunyai keterbatasan karena menggunakan dataset yang bersifat observasional dan sebagian data dibangkitkan secara sintetis, sehingga hasil yang diperoleh terbatas pada identifikasi fitur yang berkontribusi terhadap klasifikasi dan tidak dapat digunakan untuk menyimpulkan hubungan sebab-akibat. Karena itu, penelitian berikutnya bisa menggunakan dataset yang jumlahnya lebih banyak dan beragam serta menerapkan metode interpretasi model lainnya untuk memperoleh gambaran yang lebih jelas mengenai kontribusi setiap fitur terhadap hasil klasifikasi.

DAFTAR REFERENSI

- Alfarisi, M. A., Sumanto, B., & Putu, F. R. (2025). Pengembangan Model Machine Learning untuk Deteksi Penyakit Diabetes Menggunakan Analisis Gini Importance. *Journal of Internet and Software Engineering*, 127-134.
- Andhika, A. R., & Bayunanda, E. (2025). Studi Literature Review Pemanfaatan Machine Learning Dalam Multisektor. *Integrative Perspectives of Social and Science Journal*, 6865-6880.
- Daka, R., Martha, E., & Putu, L. (2025). Faktor-Faktor Yang Mempengaruhi Kejadian Obesitas Pada Remaja Di Indonesia. *Jurnal NERS*, 3190-3205.
- Ernawati, A., Sitorus, Z., Iqbal, M., & Nasution, D. (2025). Penerapan Data Mining Untuk Klasifikasi Penduduk Miskin Di Kabupaten Labuhanbatu Menggunakan Random Forest Dan K-Nearest Neighbors. *Bulletin of Information Technology (BIT)*, 23-35.
- Felndity, B., Sahsun, D., Saiful, S., Musa, S., & Lestari, T. (2025). Upaya Pencegahan Obesitas Pada Remaja Melalui Pendidikan Kesehatan. *Indonesian Journal of Community Empowerment*, 653-659.
- Fitria, D. A., & Fitria, D. A. (2019). Hubungan Obesitas Terhadap Keseimbangan Postural. *JIMKI*, 76-89.
- Fitriani, S., Budiman, E., & Fadli, M. (2025). Optimalisasi Metode Random Forest menggunakan Particle Swarm Optimization dalam Prediksi Prestasi Mahasiswa. *Prosiding Seminar Nasional Teknologi Komputer dan Sains*, 406-415.
- Handyanto, L., Arifin, W. A., & Syafri, H. (2026). Perbandingan Metode Hyperparameter Tuning Pada Model XGBoost Untuk Prediksi Multioutput. *JURTI*, 50-59.
- Haris, M., Anam, K., & Kurnianingtyas, D. (2026). Implementasi Algoritma Random Forest Untuk Prediksi Churn Pada Pelanggan Retail Online. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1-10.
- Ismail, M. Y., & Atim, S. B. (2025). Klasifikasi Level Obesitas Menggunakan Decision Tree C45 Dalam Menentukan Akurasi Pada Kriteria Information Gain, Gain Ratio, Gini Index. *JIKI*, 169-178.
- Juliani, E., & Voutama, A. (2025). Optimasi Algoritma Random Forest Untuk Meningkatkan Akurasi Prediksi Indeks Massa Tubuh (Bmi). *JuSIN*, 11-20.

- Kemenkes. (2024). *Sering Dianggap Menggemaskan, Obesitas Membahayakan Masa Depan Anak*. Retrieved from kemkes.go.id: <https://kemkes.go.id/id/sering-dianggap-menggemaskan-obesitas-membahayakan-masa-depan-anak>
- Pandu, M. A., Saputro, R. E., & Purwadi, P. (2025). Analisis Tingkat Akurasi Metode Naive Bayes dan Random Forest dalam Prediksi Penjualan Emas. *Jurnal Pendidikan dan Teknologi Indonesia*, 1809-1821.
- Saputri, I., Arsi, P., & Isnaini, K. N. (2025). Effectiveness Hyperparameter Tuning On Random Forest, Linear Discriminant Analysis, Logistic Regression And Naïve Bayes Algorithms For Detecting Dos Network Attacks. *Jurnal Teknik Informatika*, 87-104.
- Utiahman, S. A., Mulawati, A., & Pratama, M. (2024). Penerapan Support Vector Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas. *Jurnal FASILKOM*, 754-760.
- Valerian, F. R., Syarief, M., & Fatah, D. A. (2025). Klasifikasi Tingkat Obesitas Menggunakan Metode Gbm Dan Confusion Matrix. *JATI*, 2242-2249.
- Wahyudi, A., Al Baihaqi, M. A., & Febrinanda, I. V. (2025). Penerapan Random Forest untuk Klasifikasi Risiko Penyakit Stroke Pada Rentang Usia 40-85 Tahun. *JUKOMIKA*, 118-124.
- Wardhana, R. G., Wang, G., & Sibuea, F. (2023). Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia. *JOISM*, 40-45.
- WHO. (2025). *Obesity and Overweight*. Retrieved from who.int: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>