Vol. 7, No. 2, Agustus 2026, hal. 243-251

Jurnal Computer Science and Information Technology (CoSciTech)

p-ISSN: 2723-567X
e-ISSN: 2723-5661

http://ejurnal.umri.ac.id/index.php/coscitech/index

Analisis Komparatif *Decision Tree* dan Random *Forest* untuk Klasifikasi Tingkat Obesitas

Akhmad Zulkifli[1], Yulisman[*2], Edriyansyah[3], M.Riscky Firdaus[4]
Email: [1]zulkifli.akhmad@gmail.com, [2]yulisman@htp.ac.id, [3]edriyansyah75@gmail.com, [4]mhdriskyfirdaus07@gmail.com

[1,2,4]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Hang Tuah Pekanbaru
[3]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Hang Tuah Pekanbaru

Diterima: 27 Juni 2026 | Direvisi: - | Disetujui: 18 Agustus 2026

Abstrak

Obesitas merupakan salah satu masalah kesehatan global yang terus mengalami peningkatan dan berkontribusi terhadap berbagai penyakit tidak menular, seperti diabetes melitus, hipertensi, dan penyakit kardiovaskular. Faktor gaya hidup, pola konsumsi makanan, serta aktivitas fisik memiliki peran penting dalam menentukan tingkat obesitas seseorang. Perkembangan teknologi machine learning memungkinkan pengembangan model prediksi yang dapat membantu mengidentifikasi tingkat obesitas berdasarkan karakteristik individu. Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Decision Tree* dan Random *Forest* dalam klasifikasi tingkat obesitas berdasarkan faktor gaya hidup dan kondisi fisik. Dataset yang digunakan terdiri atas 2.111 data dengan 17 atribut yang mencakup informasi demografi, kebiasaan makan, aktivitas fisik, konsumsi air, penggunaan perangkat teknologi, serta riwayat keluarga. Tahapan penelitian meliputi prapemrosesan data, transformasi atribut kategorikal, pembagian data latih dan data uji, pelatihan model, serta evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma Random *Forest* memperoleh performa terbaik dengan nilai *accuracy* sebesar 95,27%, sedangkan *Decision Tree* memperoleh *accuracy* sebesar 91,73%. Selain itu, analisis *feature importance* menunjukkan bahwa atribut *Weight*, *Height*, *Age*, FCVC, dan *Gender* merupakan faktor yang memberikan kontribusi terbesar dalam proses klasifikasi tingkat obesitas. Hasil penelitian menunjukkan bahwa pendekatan *ensemble learning* pada Random *Forest* mampu meningkatkan kualitas klasifikasi dibandingkan model pohon keputusan tunggal.

Kata kunci: *obesitas, machine learning, decision tree, random forest, klasifikasi*

Comparative Analysis of Decision Tree and Random Forest for Classification of Obesity Levels

Abstract

Obesity is a global health problem that continues to increase and contributes to various non-communicable diseases, such as diabetes mellitus, hypertension, and cardiovascular disease. Lifestyle factors, food consumption patterns, and physical activity play a significant role in determining a person's obesity level. The development of machine learning technology allows the development of predictive models that can help identify obesity levels based on individual characteristics. This study aims to compare the performance of the Decision Tree and Random Forest algorithms in classifying obesity levels based on lifestyle factors and physical conditions. The dataset used consists of 2,111 data points with 17 attributes covering demographic information, eating habits, physical activity, water consumption, use of technological devices, and family history. The research stages include data preprocessing, categorical attribute transformation, dividing training and test data, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The results show that the Random Forest algorithm achieved the best performance with an accuracy value of 95.27%, while the Decision Tree algorithm achieved an accuracy of 91.73%. Furthermore, feature importance analysis showed that Weight, Height, Age, FCVC, and Gender were the most significant contributing factors to obesity classification. The results showed that the ensemble learning approach in Random Forest improved classification quality compared to a single decision tree model.

Keywords: *obesity, machine learning, decision tree, random forest, classification*

1. PENDAHULUAN

Obesitas merupakan salah satu permasalahan kesehatan yang terus meningkat di berbagai negara dan menjadi faktor risiko utama berbagai penyakit tidak menular seperti diabetes, hipertensi, dan penyakit kardiovaskular [1]. Perubahan pola hidup, tingginya konsumsi makanan berkalori tinggi, rendahnya aktivitas fisik, serta faktor genetik merupakan beberapa penyebab yang berkontribusi terhadap peningkatan tingkat obesitas pada masyarakat [2], [3].

Perkembangan teknologi komputasi dan ketersediaan data kesehatan telah mendorong pemanfaatan metode *machine learning* untuk membantu proses identifikasi dan prediksi berbagai kondisi kesehatan [4] [5]. Berbagai penelitian menunjukkan bahwa algoritma klasifikasi mampu menghasilkan prediksi yang baik terhadap tingkat obesitas berdasarkan karakteristik individu, kondisi fisik, dan kebiasaan hidup sehari-hari [6], [7], [8], [9], [10]. Pendekatan ini memungkinkan proses analisis dilakukan secara lebih cepat dibandingkan metode konvensional serta dapat digunakan sebagai dasar pengembangan sistem pendukung keputusan pada bidang kesehatan [11].

Salah satu algoritma yang banyak digunakan dalam klasifikasi adalah *Decision Tree* karena memiliki struktur yang mudah dipahami dan mampu menghasilkan aturan keputusan yang sederhana [12], [13]. Meskipun demikian, algoritma ini memiliki keterbatasan dalam menangani kompleksitas data sehingga berpotensi mengalami *overfitting* [14]. Untuk mengatasi kelemahan tersebut, metode berbasis *ensemble learning* seperti *Random Forest* dikembangkan dengan memanfaatkan kombinasi sejumlah pohon keputusan untuk meningkatkan kemampuan generalisasi model dan performa klasifikasi [15], [16].

Dataset obesitas yang digunakan dalam penelitian ini berisi informasi mengenai kebiasaan makan, aktivitas fisik, kondisi fisik, dan karakteristik individu dari beberapa negara di Amerika Latin [17]. Dataset tersebut terdiri atas 2.111 data dan 17 atribut yang dapat dimanfaatkan untuk berbagai tugas klasifikasi dan analisis data berbasis *machine learning*. Karakteristik tersebut menjadikan dataset ini sesuai untuk mengevaluasi kemampuan berbagai algoritma klasifikasi dalam memprediksi tingkat obesitas [18].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma *Decision Tree* dan *Random Forest* dalam klasifikasi tingkat obesitas. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan algoritma yang memberikan performa terbaik [19]. Selain itu, penelitian ini juga menganalisis tingkat kepentingan atribut guna mengidentifikasi faktor-faktor yang paling berpengaruh terhadap klasifikasi tingkat obesitas [20]. Kontribusi utama penelitian ini adalah melakukan evaluasi komparatif dua algoritma klasifikasi populer pada dataset obesitas yang mengintegrasikan faktor gaya hidup dan kondisi fisik, serta mengidentifikasi atribut dominan melalui analisis *feature importance*.

2. METODE PENELITIAN

2.1 *Dataset*

Penelitian ini menggunakan dataset obesitas yang terdiri atas 2.111 data dan 17 atribut. *Dataset* mencakup informasi demografi, kebiasaan makan, aktivitas fisik, kondisi fisik, serta tingkat obesitas individu. Variabel target yang digunakan adalah *NObeyesdad* yang terdiri atas tujuh kategori tingkat obesitas [21]. *Dataset* ini banyak digunakan dalam penelitian klasifikasi obesitas dan menyediakan kombinasi atribut numerik serta kategorikal yang sesuai untuk penerapan algoritma *machine learning*. Tabel 1 menunjukkan atribut yang digunakan dalam penelitian.

Tabel 1. Deskripsi Atribut Dataset

No	Atribut	Deskripsi
1	*Gender*	Jenis kelamin
2	*Age*	Usia responden
3	*Height*	Tinggi badan (meter)
4	*Weight*	Berat badan (kilogram)
5	*family history with overweight*	Riwayat keluarga dengan obesitas
6	FAVC	Konsumsi makanan berkalori tinggi
7	FCVC	Frekuensi konsumsi sayuran
8	NCP	Jumlah makan utama per hari
9	CAEC	Kebiasaan makan di antara waktu makan
10	SMOKE	Kebiasaan merokok
11	CH2O	Konsumsi air harian
12	SCC	Pemantauan konsumsi kalori
13	FAF	Frekuensi aktivitas fisik
14	TUE	Durasi penggunaan perangkat teknologi
15	CALC	Frekuensi konsumsi alkohol
16	MTRANS	Moda transportasi yang digunakan
17	*NObeyesdad*	Tingkat obesitas (*target class*)

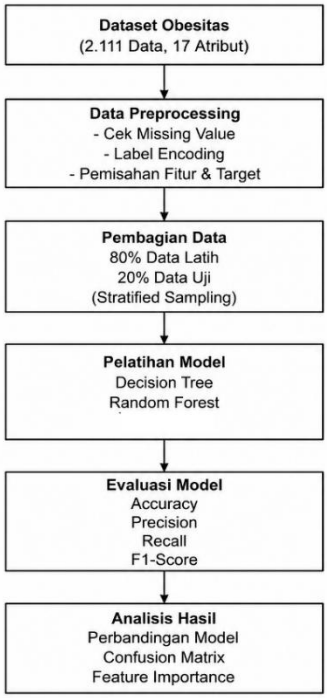
Dataset ini mengandung atribut numerik dan kategorikal sehingga dapat digunakan untuk berbagai pendekatan klasifikasi berbasis *machine learning* [22] [23]. Kombinasi atribut tersebut memungkinkan proses analisis dilakukan secara komprehensif dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap tingkat obesitas individu.

2.2. *Data Preprocessing*

Tahap *preprocessing* dilakukan untuk mempersiapkan data sebelum proses pelatihan model. Langkah pertama adalah pemeriksaan struktur data dan identifikasi nilai yang hilang (missing values). Berdasarkan hasil pemeriksaan awal, dataset tidak memiliki nilai kosong sehingga seluruh data dapat digunakan pada proses analisis.

Selanjutnya, atribut kategorikal ditransformasikan menjadi bentuk numerik menggunakan metode *Label Encoding*. Proses ini diperlukan karena algoritma klasifikasi yang digunakan tidak dapat memproses data kategorikal secara langsung. Setelah proses transformasi selesai, dataset dipisahkan menjadi variabel fitur (X) dan variabel target (y).

Tahap berikutnya adalah pembagian dataset menjadi data latih dan data uji menggunakan metode train-test split dengan proporsi 80% data latih dan 20% data uji. Untuk menjaga proporsi setiap kelas tetap seimbang pada kedua kelompok data, digunakan teknik stratified sampling.



Gambar 1. Alur Penelitian

Alur penelitian mulai dari akuisisi data, prapemrosesan, pelatihan model, evaluasi model, hingga analisis faktor dominan.

2.3. *Classification Models*

Penelitian ini menggunakan dua algoritma klasifikasi yang banyak digunakan dalam bidang *machine learning*, yaitu *Decision Tree* dan Random *Forest*.

2.3.1. *Decision Tree*

Decision Tree merupakan algoritma klasifikasi berbasis struktur pohon yang membagi data ke dalam beberapa cabang berdasarkan atribut yang dianggap paling informatif [22], [24]. Setiap simpul (*node*) merepresentasikan atribut tertentu, sedangkan daun (*leaf*) menunjukkan hasil klasifikasi [25]. Keunggulan metode ini adalah kemudahan interpretasi dan visualisasi aturan keputusan [26].

2.3.2. Random *Forest*

Random *Forest* merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan secara independen dan menghasilkan prediksi berdasarkan mekanisme majority *voting* [22]. Pendekatan ini mampu mengurangi risiko *overfitting* yang sering ditemukan pada *Decision Tree* tunggal sehingga menghasilkan performa yang lebih stabil [27].

2.4. *Evaluation Metrics*

Evaluasi performa model dilakukan menggunakan empat metrik klasifikasi, yaitu *accuracy, precision, recall*, dan *F1-score* [22].

2.4.1. *Accuracy*

Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data pengujian [22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2.4.2. *Precision*

Precision menunjukkan tingkat ketepatan model dalam mengidentifikasi suatu kelas [22].

$$Precision = \frac{TP}{TP + FP}$$

2.4.3. *Recall*

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang termasuk ke dalam kelas tertentu [22].

$$Recall = \frac{TP}{TP + FN}$$

2.4.4. *F1-Score*

F1-score merupakan rata-rata harmonik antara *precision* dan *recall* [22].

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

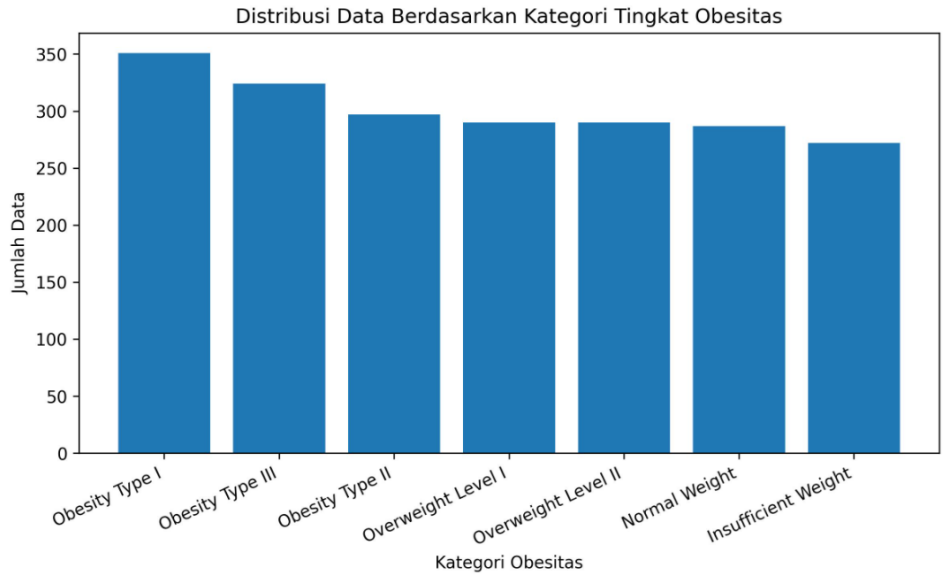
Metrik-metrik tersebut digunakan untuk membandingkan performa *Decision Tree* dan Random *Forest*, sehingga dapat ditentukan model yang paling efektif dalam klasifikasi tingkat obesitas.

3. HASIL DAN PEMBAHASAN

3.1. *Dataset Characteristics*

Dataset yang digunakan dalam penelitian ini terdiri atas 2.111 data dengan 17 atribut yang merepresentasikan karakteristik individu, kebiasaan makan, aktivitas fisik, serta tingkat obesitas. Variabel target (*NObeyesdad*) terdiri atas tujuh kategori tingkat obesitas, yaitu *Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II,* dan *Obesity Type III*.

Berdasarkan hasil analisis distribusi kelas, jumlah data pada setiap kategori relatif seimbang dengan rentang antara 272 hingga 351 data. Kondisi tersebut menunjukkan bahwa *dataset* memiliki distribusi kelas yang baik sehingga dapat digunakan untuk proses klasifikasi tanpa dominasi yang berlebihan pada kelas tertentu. Distribusi data yang seimbang berkontribusi terhadap peningkatan kemampuan model dalam mempelajari karakteristik setiap kelas secara lebih representatif.



Gambar 2. Distribusi data berdasarkan kategori tingkat obesitas

Selain distribusi kelas, dilakukan analisis statistik deskriptif terhadap atribut numerik utama yang terdiri atas usia (*Age*), tinggi badan (*Height*), dan berat badan (*Weight*). Hasil statistik deskriptif ditunjukkan pada Tabel 2.

Tabel 2. Statistik Deskriptif Atribut Numerik

Atribut	Minimum	Maksimum	Rata-rata	Standar Deviasi
Age	14	61	24.31	6.35
Height	1.45	1.98	1.7	0.09
Weight	39	173	86.59	26.19

Berdasarkan Tabel 2, usia responden berada pada rentang 14 hingga 61 tahun dengan rata-rata 24,31 tahun. Tinggi badan responden berkisar antara 1,45 meter hingga 1,98 meter dengan rata-rata 1,70 meter. Sementara itu, berat badan memiliki variasi yang cukup besar dengan rentang 39 hingga 173 kilogram dan rata-rata 86,59 kilogram. Variasi berat badan yang tinggi menunjukkan keberagaman tingkat obesitas dalam dataset sehingga memberikan kondisi yang sesuai untuk pengujian berbagai algoritma klasifikasi.

Karakteristik *dataset* tersebut menunjukkan bahwa data yang digunakan memiliki keragaman atribut yang cukup untuk mendukung proses klasifikasi tingkat obesitas menggunakan algoritma *Decision Tree* dan Random *Forest*.

3.2. *Model Performance Comparison*

Tahap evaluasi dilakukan untuk membandingkan kemampuan algoritma *Decision Tree* dan Random *Forest* dalam mengklasifikasikan tingkat obesitas berdasarkan atribut gaya hidup dan kondisi fisik yang tersedia pada *dataset*. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Tabel 3. Perbandingan Performa Model Klasifikasi

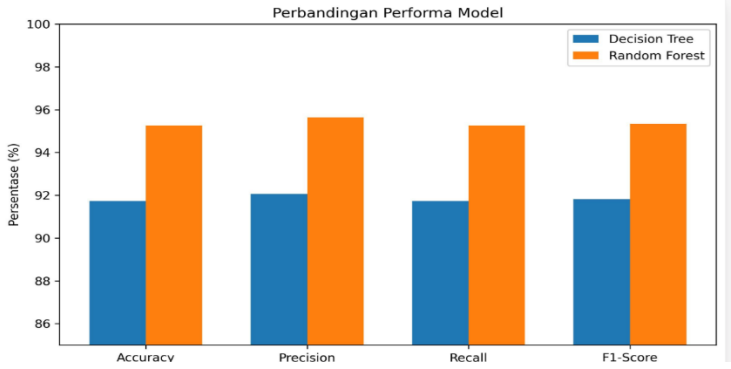
Model	***Accuracy***	***Precision***	***Recall***	***F1-Score***
Decision Tree	91,73%	92,06%	91,73%	91,82%
Random Forest	95,27%	95,63%	95,27%	95,34%

Hasil pengujian yang ditunjukkan pada Tabel 3 memperlihatkan bahwa kedua algoritma mampu menghasilkan performa klasifikasi yang sangat baik dengan nilai *accuracy* di atas 90%. Namun demikian, algoritma Random *Forest* menunjukkan performa yang lebih unggul dibandingkan Decision Tree pada seluruh metrik evaluasi.

Algoritma *Decision Tree* memperoleh nilai *accuracy* sebesar 91,73%, *precision* sebesar 92,06%, *recall* sebesar 91,73%, dan *F1-score* sebesar 91,82%. Hasil tersebut menunjukkan bahwa model mampu mengidentifikasi sebagian besar kategori obesitas dengan tingkat ketepatan yang tinggi. Meskipun demikian, model masih memiliki keterbatasan dalam menangani variasi pola data yang kompleks karena hanya menggunakan satu struktur pohon keputusan.

Sementara itu, algoritma Random *Forest* menghasilkan nilai *accuracy* sebesar 95,27%, *precision* sebesar 95,63%, *recall* sebesar 95,27%, dan *F1-score* sebesar 95,34%. Peningkatan performa ini menunjukkan bahwa pendekatan *ensemble learning* mampu meningkatkan kemampuan generalisasi model melalui kombinasi sejumlah pohon keputusan yang dibangun secara independen. Mekanisme *majority voting* yang digunakan oleh Random *Forest* juga membantu mengurangi risiko *overfitting* yang sering ditemukan pada *Decision Tree* tunggal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa Random *Forest* merupakan model yang paling efektif dalam klasifikasi tingkat obesitas pada *dataset* yang digunakan. Peningkatan *accuracy* sebesar 3,54% dibandingkan *Decision Tree* mengindikasikan bahwa penggunaan pendekatan *ensemble* memberikan kontribusi positif terhadap kualitas prediksi. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa metode *ensemble learning* cenderung menghasilkan performa yang lebih stabil dan akurat dibandingkan model pohon keputusan tunggal [22].

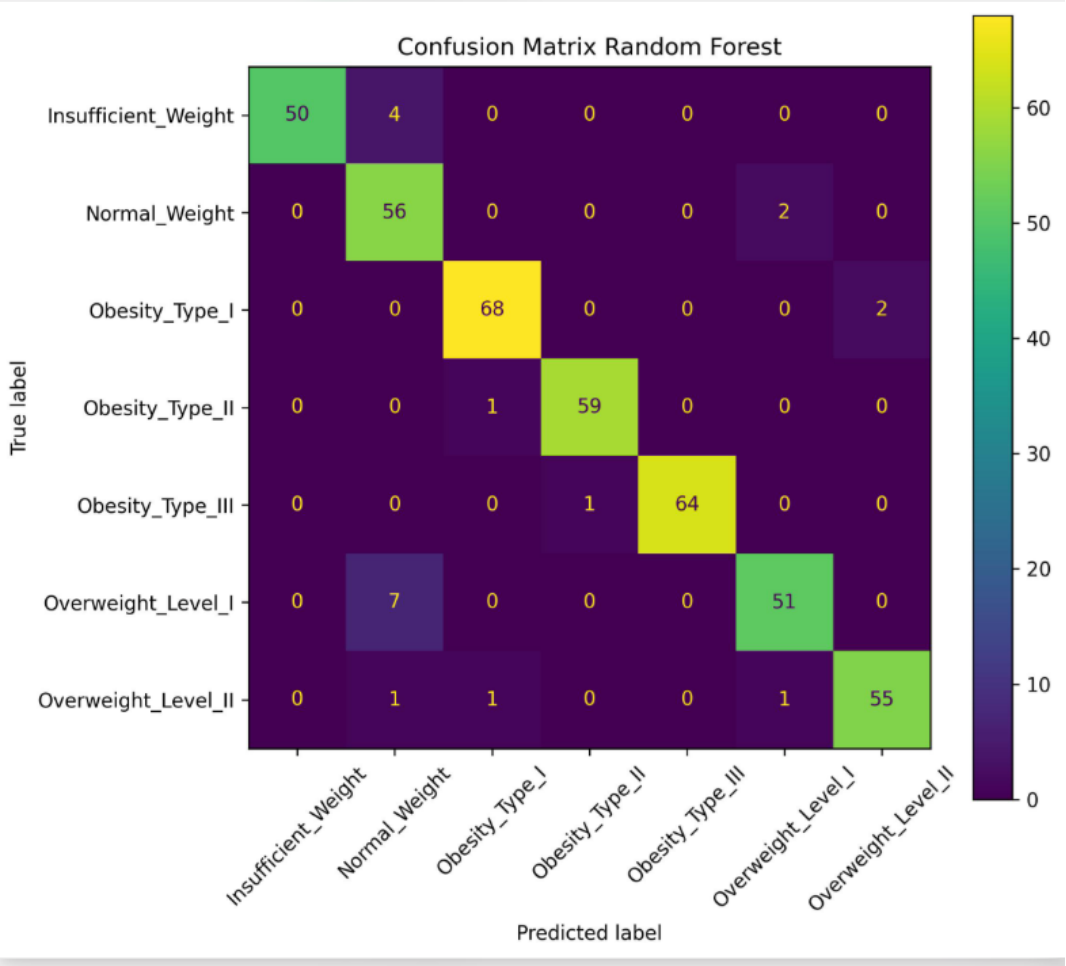


Gambar 3. Perbandingan performa algoritma *Decision Tree* dan Random *Forest* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

3.3. *Confusion Matrix Analysis*

Selain menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, evaluasi model juga dilakukan melalui analisis *confusion matrix*. Analisis ini digunakan untuk mengetahui kemampuan model dalam mengklasifikasikan setiap kategori tingkat obesitas secara lebih rinci. Dengan menggunakan *confusion matrix*, dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi pada masing-masing kelas.

Berdasarkan hasil evaluasi, algoritma Random *Forest* dipilih sebagai model terbaik karena menghasilkan nilai *accuracy* tertinggi dibandingkan *Decision Tree*. Oleh karena itu, analisis confusion matrix difokuskan pada model Random *Forest* untuk mengidentifikasi pola prediksi yang dihasilkan model terhadap tujuh kategori tingkat obesitas.



Gambar 4. *Confusion Matrix* algoritma Random *Forest*

Berdasarkan Gambar 4, sebagian besar data berhasil diklasifikasikan pada kategori yang sesuai, yang ditunjukkan oleh nilai yang tinggi pada diagonal utama matriks. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali karakteristik masing-masing kategori obesitas. Kesalahan klasifikasi yang terjadi umumnya ditemukan pada kelas-kelas yang memiliki karakteristik serupa, terutama pada kategori yang berdekatan seperti *Overweight Level I* dan *Overweight Level II*, atau antara *Obesity Type I* dan *Obesity Type II*.

Fenomena tersebut dapat dipahami karena individu pada kategori yang berdekatan sering memiliki pola gaya hidup dan kondisi fisik yang tidak jauh berbeda. Akibatnya, beberapa atribut yang digunakan dalam proses klasifikasi memiliki nilai yang saling tumpang tindih sehingga meningkatkan kemungkinan terjadinya kesalahan prediksi. Meskipun demikian, jumlah kesalahan klasifikasi relatif kecil dibandingkan jumlah prediksi yang benar, yang tercermin dari tingginya nilai *accuracy* yang diperoleh model.

Hasil analisis ini menunjukkan bahwa algoritma Random *Forest* tidak hanya memiliki performa yang baik secara keseluruhan, tetapi juga mampu mempertahankan tingkat klasifikasi yang konsisten pada berbagai kategori obesitas. Kemampuan tersebut menjadi salah satu alasan mengapa model berbasis *ensemble learning* sering digunakan pada permasalahan klasifikasi kesehatan yang melibatkan banyak atribut dan kelas target.

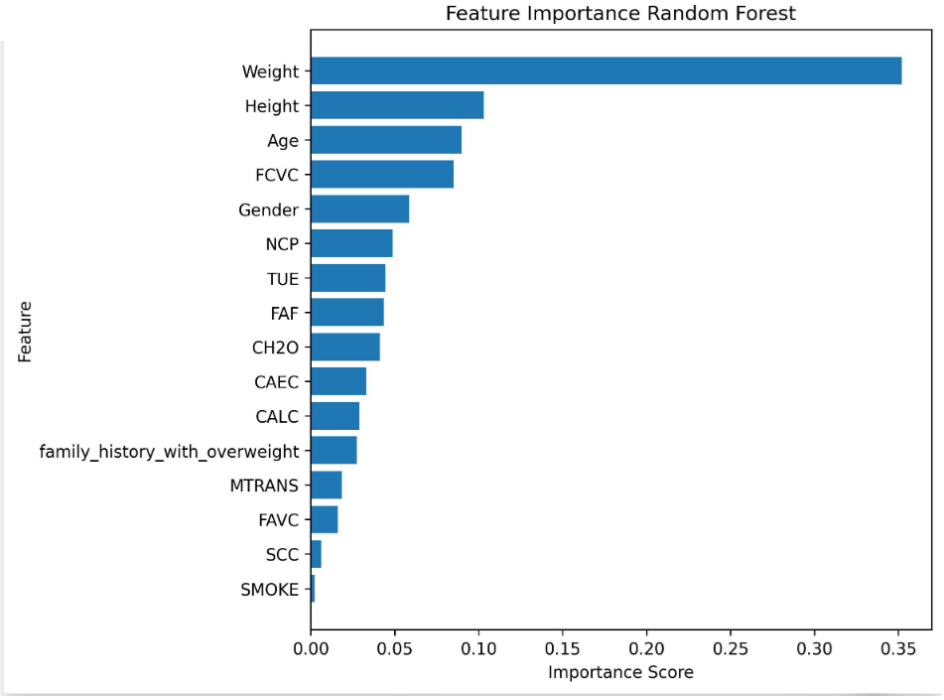
3.4. *Feature Importance Analysis*

Selain mengevaluasi performa klasifikasi, penelitian ini juga melakukan analisis *feature importance* untuk mengetahui atribut yang paling berpengaruh dalam proses prediksi tingkat obesitas. Analisis ini dilakukan menggunakan nilai kepentingan fitur yang dihasilkan oleh algoritma Random *Forest*.

Feature importance memberikan informasi mengenai kontribusi relatif setiap atribut terhadap proses pembentukan keputusan pada model. Semakin tinggi nilai kepentingan suatu atribut, semakin besar pengaruh atribut tersebut terhadap hasil klasifikasi.

Tabel 4. Lima Atribut dengan Nilai *Feature Importance* Tertinggi

Peringkat	Atribut	Deskripsi Ilmiah
1	*Weight*	Berat badan merupakan faktor utama dalam penentuan tingkat obesitas dan berhubungan langsung dengan perhitungan *Body Mass Index* (*BMI*).
2	*Height*	Tinggi badan berperan dalam perhitungan *BMI* sehingga berkontribusi terhadap penentuan kategori obesitas seseorang.
3	*Age*	Usia memengaruhi metabolisme tubuh, kebutuhan energi, dan kecenderungan peningkatan berat badan yang berkaitan dengan risiko obesitas.
4	FCVC	Frekuensi konsumsi sayuran mencerminkan kualitas pola makan dan berhubungan dengan pengendalian berat badan serta risiko obesitas.
5	*Gender*	Perbedaan jenis kelamin dapat memengaruhi distribusi lemak tubuh, kebutuhan energi, dan karakteristik metabolisme individu.



Gambar 5. Nilai *feature importance* pada algoritma Random *Forest*.

Berdasarkan hasil analisis feature importance, atribut *Weight* memiliki kontribusi terbesar dalam proses klasifikasi tingkat obesitas. Temuan ini menunjukkan bahwa faktor fisik masih menjadi indikator utama dalam penentuan kategori obesitas. Selain itu, atribut *Height* dan *Age* juga memiliki kontribusi yang tinggi karena berkaitan dengan karakteristik antropometri individu yang berpengaruh terhadap status gizi dan tingkat obesitas seseorang.

Atribut FCVC (*Frequency of Consumption of Vegetables*) termasuk ke dalam lima atribut terpenting yang digunakan model dalam proses klasifikasi. Hasil ini menunjukkan bahwa pola konsumsi sayuran memiliki hubungan dengan tingkat obesitas, di mana kualitas pola makan menjadi salah satu faktor yang memengaruhi kondisi kesehatan individu. Selain itu, atribut Gender juga memberikan kontribusi yang cukup besar terhadap proses klasifikasi. Temuan tersebut mengindikasikan adanya perbedaan karakteristik obesitas berdasarkan jenis kelamin yang dapat dipengaruhi oleh faktor fisiologis maupun perilaku.

Secara keseluruhan, hasil analisis *feature importance* menunjukkan bahwa atribut yang berkaitan dengan karakteristik fisik individu memiliki pengaruh yang lebih dominan dibandingkan atribut perilaku. Meskipun demikian, faktor gaya hidup seperti pola konsumsi makanan tetap memberikan kontribusi terhadap proses klasifikasi tingkat obesitas. Temuan ini mendukung

penelitian sebelumnya yang menyatakan bahwa obesitas merupakan kondisi multifaktorial yang dipengaruhi oleh kombinasi faktor biologis, demografis, dan perilaku [28]. Hasil tersebut menunjukkan bahwa upaya pencegahan obesitas perlu mempertimbangkan aspek fisik dan perilaku secara bersamaan untuk memperoleh hasil yang lebih efektif.

4. KESIMPULAN

Penelitian ini telah melakukan analisis komparatif terhadap algoritma *Decision Tree* dan Random *Forest* untuk klasifikasi tingkat obesitas berdasarkan faktor gaya hidup dan kondisi fisik individu. *Dataset* yang digunakan terdiri atas 2.111 data dengan 17 atribut yang mencakup informasi demografi, kebiasaan makan, aktivitas fisik, dan karakteristik fisik responden.

Hasil evaluasi menunjukkan bahwa kedua algoritma mampu menghasilkan performa klasifikasi yang baik. Namun demikian, algoritma Random *Forest* memberikan performa yang lebih unggul dibandingkan *Decision Tree* pada seluruh metrik evaluasi yang digunakan. Random *Forest* memperoleh nilai *accuracy* sebesar 95,27%, sedangkan *Decision Tree* memperoleh nilai *accuracy* sebesar 91,73%. Hasil tersebut menunjukkan bahwa pendekatan *ensemble learning* mampu meningkatkan kemampuan generalisasi model dan menghasilkan klasifikasi yang lebih akurat.

Analisis *confusion matrix* menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar pada masing-masing kategori tingkat obesitas. Sementara itu, hasil analisis *feature importance* mengidentifikasi bahwa atribut *Weight*, *Height*, *Age*, FCVC, dan *Gender* merupakan faktor yang memberikan kontribusi terbesar dalam proses klasifikasi. Temuan ini menunjukkan bahwa karakteristik fisik individu masih menjadi faktor dominan dalam penentuan tingkat obesitas, meskipun faktor gaya hidup juga berperan dalam mendukung proses klasifikasi.

Berdasarkan hasil penelitian, algoritma Random *Forest* direkomendasikan sebagai metode yang lebih efektif untuk klasifikasi tingkat obesitas pada *dataset* yang digunakan. Penelitian selanjutnya dapat mengembangkan model dengan melakukan optimasi parameter, menerapkan teknik *feature selection*, atau membandingkan model *ensemble learning* lain seperti *Gradient Boosting*, *AdaBoost*, dan *Extra Trees* untuk memperoleh performa klasifikasi yang lebih baik [29] [30]. Selain itu, penggunaan *dataset* dengan jumlah sampel yang lebih besar dan karakteristik responden yang lebih beragam dapat meningkatkan kemampuan generalisasi model pada berbagai populasi.

DAFTAR PUSTAKA

[1] World Health Organization, “Obesity and overweight.” Accessed: Jun. 24, 2026. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[2] S. A. Nabila, E. Sunarsih, Novrikasari, and A. Rahmiwati, “Dietary Behavior and Physical Activity on the Problem of Obesity: Systematic Review,” *Media Publikasi Promosi Kesehatan Indonesia*, vol. 7, no. 3, pp. 498–505, Mar. 2024, doi: 10.56338/mppki.v7i3.4533.

[3] R. Alfarisi, D. W. Nurfadilah, and D. Shakira, “Tingkat Pengetahuan Tentang Obesitas pada Masyarakat Dusun Sumber Sari Desa Taman Sari Kecamatan Gedong Tataan,” *Jurnal Ilmu Kedokteran dan Kesehatan*, vol. 9, no. 4, pp. 2549–4864, 2022, doi: 10.33024/jikk.v9i4.8384.

[4] L. Setiyani, A. N. Indahsari, and R. Roestam, “Analisis Prediksi Level Obesitas Menggunakan Perbandingan Algoritma Machine Learning dan Deep Learning,” *JTERA (Jurnal Teknologi Rekayasa)*, vol. 8, no. 1, pp. 139–146, 2023, doi: 10.31544/jtera.v8.i1.2023.139-146.

[5] H. Manik, “Integrasi Kecerdasan Buatan dalam Diagnostik Medis: Peluang dan Tantangan di Era Kesehatan Digital,” *Indonesian Journal of Medicine, Health and Nursing*, vol. 1, no. 1, pp. 45–57, 2026, doi: 10.70742/ijmhn.v1i1.577.

[6] A. Abdullah, D. U. Khairah, and M. W. Pangestika, “Comparison of random forest and xgboost algorithms in credit card fraud classification,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 3, pp. 392–398, Dec. 2025, doi: 10.37859/coscitech.v6i3.10470.

[7] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[8] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. “ O’Reilly Media, Inc.,” 2022.

[9] R. R. Pratama, “Analisis Model Machine Learning Terhadap Pengenalan Aktifitas Manusia,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 302–311, May 2020, doi: 10.30812/matrik.v19i2.688.

[10] E. Rohaeti, A. Andriyati, and M. Edy Rizal, “Klasifikasi Obesitas Menggunakan Domain-Based Decision Tree, XGBoost, dan SHAP,” *Limits: Journal of Mathematics and Its Applications*, vol. 23, no. 1, pp. 77–96, Apr. 2026, doi: 10.12962/limits.v23i1.7970.

[11] D. Fakhrizal, A. A. Sutrisno, N. A. Zain, G. A. Mukti, and A. Setiawan, “Implementasi Algoritma Machine Learning Menggunakan Model Random Forest Untuk Klasifikasi Obesitas,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 5, pp. 7579–7584, 2025, doi: 10.36040/jati.v9i5.14667.

[12] R. Faizal, A. Abdullah, and M. W. Pangestika, “Perbandingan Random Forest Regressor Dan Decision Tree Regressor Untuk Prediksi Hasil Panen,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 2, pp. 247–253, Sep. 2025, doi: 10.37859/coscitech.v6i2.9966.

[13] A. P. Sabrina and Z. Fatah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Resiko Diabetes,” *JAMASTIKA*, vol. 4, no. 2, pp. 273–279, 2025, doi: 10.35473/jamastika.v4i2.4543.

[14] K. Syaban and Mardiawati, “Evaluasi Model Ensemble Learning pada Identifikasi Faktor Risiko Diabetes Mellitus,” *Jurnal Teknologi dan Informasi*, vol. 15, no. 2, pp. 121–130, Sep. 2025, doi: 10.34010/jati.v15i2.16238.

[15] H. Fauzan, E. Haerani, F. Kurnia, and N. Yanti, “Implementasi Algoritma Random Forest pada Web-App Sebagai Instrumen Deteksi Dini Penyakit Diabetes,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 7, no. 1, pp. 75–83, Apr. 2025, doi: 10.37859/coscitech.v7i1.11261.

[16] W. Amanda and A. Voutama, “Klasifikasi Pendapatan Menggunakan Algoritma Random Forest: Studi Kasus Dataset Adult Income,” *Jurnal Ilmiah Informatika Global*, vol. 16, no. 2, pp. 79–84, 2025.

[17] Y. E. I. Noor, E. Sugiarto, and A. S. Fatimah, “The Description of Obesity Among Housewifes in The World,” *Jurnal Gizi Dan Kesehatan*, vol. 14, no. 1, pp. 34–42, 2022, Accessed: Jun. 25, 2026. [Online]. Available: https://storage.nu.or.id/storage/files/jurnal-kesehatan-obesitas-jgk_1721379451.pdf

[18] R. N. Alifah *et al.*, “Perbandingan Metode Tree Based Classification untuk Masalah Klasifikasi Data Body Mass Index,” *Indonesian Journal of Mathematics and Natural Sciences*, vol. 47, no. 1, pp. 49–65, 2024, doi: 10.15294/m2k97436.

[19] O. Septa, N. Triyasri, M. A. Permata, A. Salsabila, and F. Firdani, “Analisis Klasifikasi Multikelas Obesitas Menggunakan Algoritma Decision Tree Classifier, Random Forest Classifier, dan Support Vector Classifier (SVC),” *Jurnal of Data Science Methods and Applications*, vol. 02, no. 01, pp. 12–21, 2026, doi: 10.30873/jodmapps.v2i1.pp12-21.

[20] S. A. Utiarahman and M. A. M. Pratama, "Penerapan Support Vector Machine dan Random Forest Classifier Untuk Klasifikasi Tingkat Obesitas," *Jurnal FASILKOM (teknologi inFormASi dan ILmu KOMputer)*, vol. 14, no. 3, pp. 754–760, 2024, doi: 10.37859/jf.v14i3.8104.

[21] A. B. Alpiansah and Y. Ramdhani, "Optimasi Fitur dengan Forward Selection pada Estimasi Tingkat Obesitas menggunakan Random Forest," *SISTEMASI: Jurnal Sistem Informasi*, vol. 12, no. 3, pp. 860–873, 2023, doi: 10.32520/stmsi.v12i3.3125.

[22] S. L. M. Sitio, *Machine Learning Berbasis Pohon Keputusan: Implementasi Decision Tree dan Random Forest di Google Colab*. Cilacap: MEDIA PUSTAKA INDO, 2026. [Online]. Available: www.mediapustakaindo.com

[23] A. R. Husaini, "Klasifikasi Berbasis Machine Learning untuk Prediksi Kondisi Kesehatan Berdasarkan Gejala Umum COVID-19," *EJECTS: Journal Computer, Technology and Informations System*, vol. 5, no. 1, pp. 13–18, 2025, Accessed: Jun. 25, 2026. [Online]. Available: https://jurnal.unda.ac.id/index.php/ejects/article/download/557/369

[24] W. A. Pamungkas, "Penerapan Klasifikasi Data Mining Untuk Prediksi Dengan Metode Algoritma Decission Tree," in *Seminar Nasional Sains dan Teknologi "SainTek" Seri IV*, Tangerang Selatan, 2025, pp. 655–678. Accessed: Jun. 25, 2026. [Online]. Available: https://conference.ut.ac.id/index.php/saintek/article/view/6737

[25] A. Khikam, N. Martyan Anggadimas, and M. Udin, "Implementasi Decision Tree Untuk Klasikasi Obesitas," 2025.

[26] M. Rizky, D. Nasien, Johan, and G. Tendra, "Penerapan Decision Tree Menggunakan Algoritma Cart Dalam Prediksi Indeks Massa Tubuh Berbasis Web," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 7, no. 2, pp. 168–172, 2025, doi: 10.35145/jmapteksi.v7i2.4986.

[27] N. Rosidah, R. Prathivi, and Susanto, "Optimasi Metode Random Forest Untuk Klasifikasi Risiko Obesitas Berdasarkan Pola Makan," *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 1, pp. 56–62, 2025, Accessed: Jun. 26, 2026. [Online]. Available: https://www.academia.edu/download/121400118/Optimasi_Metode_Random_Forest_Untuk_Klasifikasi_Risiko_Obesitas_Berdasarkan_Pola_Makan.pdf

[28] 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.

[29] R. V. Suryadinata and D. A. Sukarno, "The Effect Of Physical Activity On The Risk Of Obesity In Adulthood," *The Indonesian Journal of Public Health*, vol. 14, no. 1, pp. 106–116, 2019, doi: 10.20473/ijph.vl14il.2019.106-116.

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R Second Edition," 2021.