IMPLEMENTASI ALGORITMA MACHINE LEARNING MENGGUNAKAN MODEL RANDOM FOREST UNTUK KLASIFIKASI OBESITAS

Diki Fakhrizal, Aditia Ananda Sutrisno, Nazhat Afza Zain, Galih Angga Mukti, Arif Setiawan
Teknik Informatika, Universitas Muria Kudus
Jl. Lkr. Utara, Kayuapu Kulon, Gondangmanis, Kec. Bae, Kabupaten Kudus, Jawa Tengah 59327
202251101@std.umk.ac.id

ABSTRAK

Obesitas merupakan ancaman kesehatan global yang sering disepelekan, banyak metode yang digunakan mendeteksi obesitas salah satunya dengan Indeks Massa Tubuh (BMI) yang masih memiliki keterbatasan dalam mengenali lemak tubuh atau massa otot. Penelitian ini bertujuan mengembangkan klasifikasi obesitas menggunakan model Random Forest, model ini mengklasifikasikan obesitas berdasarkan beberapa fitur seperti usia, berat badan, tinggi badan, aktivitas fisik, dan riwayat keluarga. Dataset yang digunakan didapat dari Kaggle, yang memiliki 2.111 baris data dan 17 fitur, proses yang dilalui meliputi analisis eksplorasi, penanganan ketidakseimbangan data, penghapusan duplikasi, serta konversi data kategorik menjadi numerik. Selanjutnya untuk melatih dan menguji model Random Forest dilakukan pembagian data menjadi 80% data pelatihan dan 20% data pengujian. Evaluasi model meliputi metrik akurasi, *Confusion Matrix*, *precision*, *recall*, F1-score. Hasil yang didapat dari penelitian menunjukan model dapat mengklasifikasikan obesitas sesuai kelas dengan akurasi tinggi mencapai 96,65%. Pada *Confusion Matrix*, model menunjukan kinerja yang baik di berbagai kategori obesitas dengan nilai *Macro avg* dan *Weighted avg* sebesar 0.96 dan 0.97. Dengan hasil tersebut terbukti bahwa model mampu mengidentifikasi obesitas berdasarkan tingkatnya menggunakan fitur perilaku dan fisiologis pengguna, dan dapat membantu pada sistem pendukung keputusan medis dalam pendeteksian obesitas sejak dini.

Kata kunci : *Klasifikasi, Obesitas, Machine Learning, Random Forest*

1. PENDAHULUAN

Obesitas merupakan salah satu tantangan global yang paling urgent saat ini. Menurut data yang diambil dari World Health Organization (WHO), secara umum obesitas telah meningkat secara signifikan dalam beberapa dekade terakhir. Sejak tahun 1975, tingkat obesitas global meningkat tiga kali lipat [1]. Diperkirakan pada tahun 2030, orang yang memiliki kelebihan berat badan akan menyentuh angka 2,16 miliar, dan yang menderita obesitas diperkirakan akan mencapai 1,12 miliar di seluruh dunia. Data yang diperoleh dari Survei Kesehatan dan Morbiditas Nasional menunjukkan bahwa wanita lebih mungkin obesitas sebesar 29,6% dibandingkan dengan pria yang hanya mencapai 25% [2].

Fenomena ini menjadikan obesitas sebagai faktor risiko yang signifikan untuk sejumlah penyakit jangka panjang. Seperti penyakit jantung, diabetes tipe 2, dan juga dapat menyebabkan timbulnya berbagai jenis kanker [3]. Meningkatnya obesitas akan memberikan dampak bagi sistem kesehatan masyarakat dan ekonomi global [4]. Perubahan signifikan pada pola makan yang dapat menyebabkan peningkatan konsumsi makanan olahan yang tinggi kalori, serta kurangnya aktivitas fisik. Di Indonesia, obesitas menjadi masalah kesehatan yang terus meningkat [5].

Berdasarkan data yang didapatkan dari Riset Kesehatan Dasar (Riskesdas) menyatakan bahwa obesitas pada orang dewasa terus meningkat [6]. Pada tahun 2013, 11,5% penduduk Indonesia dengan usia lebih dari 18 tahun mengalami over weight. Sementara itu, secara umum obesitas juga meningkat dari 10,5% pada tahun 2007. Sedangkan pada tahun 2013 mengalami peningkatan yaitu di angka 14,8% dan pada tahun 2018 sebesar 21,8% [7]. Salah satu faktor utama dari obesitas adalah perubahan lifestyle yang cenderung lebih sering untuk memakan fast food yang dibarengi dengan minimnya aktivitas fisik. Lebih jauh, obesitas pada anak-anak juga menunjukkan angka yang mengkhawatirkan, yang dapat menimbulkan masalah kesehatan jangka panjang [8].

Salah satu pendekatan tradisional yang paling sering digunakan untuk menentukan status obesitas adalah melakukan perhitungan *Body Mass Index* (BMI), yang diperoleh dari membagi berat badan ke dalam kilogram dengan kuadrat tinggi badan dalam meter (kg/m2). Akan tetapi, metode ini memiliki keterbatasan yaitu tidak akurat mencerminkan distribusi lemak tubuh atau massa otot. BMI tidak dapat membedakan apakah seseorang yang kelebihan badan disebabkan oleh lemak subkutan, lemak visceral, atau massa otot. Oleh karena itu masih diperlukan pengukuran tambahan seperti lingkar pinggang karena memberikan lebih banyak wawasan tentang akumulasi lemak perut. Lemak visceral yang berlebih dianggap lebih berbahaya daripada lemak yang terdistribusi secara merata, karena sangat terkait dengan peningkatan risiko penyakit kardiovaskular, diabetes tipe 2, dan gangguan metabolik lainnya [9].

Di era digital ini, kemajuan teknologi informasi telah mengubah bebearpa sektor termasuk sektor kesehatan dengan menuntut adanya pendekatan yang lebih canggih dalam proses analisis data dan pengambilan keputusan. Untuk memenuhi kebutuhan yang terus berkembang, *Machine Learning* dibuat

sebagai alat yang powerful untuk menganalisis sejumlah besar data kesehatan dengan cara mengidentifikasi pola kompleks yang mungkin diabaikan oleh metode tradisional. Algoritma ML pada dasarnya dirancang untuk meniru kecerdasan manusia melalui pembelajaran yang berkelanjutan dari data sehingga sangat cocok untuk prediksi klasifikasi terkait kesehatan.

Diantara banyaknya algoritma ML yang tersedia, Random Forest merupakan salah satu teknik yang paling cocok dan digunakan untuk banyak tugas klasifikasi. Sebagai metode dalam kategori *ensemble learning*, metode ini menggabungkan prediksi dari beberapa decision tree untuk menghasilkan hasil yang lebih akurat. Random Forest tidak hanya unggul dalam menangani data berdimensi tinggi dan non linier, akan tetapi menunjukkan ketahanan yang lebih besar terhadap *overfitting* dibandingkan dengan *decision tree*. Keunggulan ini menjadikan metode yang ideal untuk mengembangkan model klasifikasi obesitas berdasarkan beberapa parameter yang telah ada.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi obesitas menggunakan algoritma Random Forest, berdasarkan fitur-fitur seperti berat badan, usia, gender, aktivitas fisik, hingga riwayat keluarga. Untuk melakukan evaluasi kinerja model secara komprehensif, metrik standar seperti akurasi, presisi, recall, f1-skor juga akan digunakan. Hasil penelitian ini diharapkan dapat memberikan kontribusi yang berarti bagi pengembangan sistem pendukung keputusan medis untuk mendeteksi obesitas secara dini berdasarkan fitur-fitur yang digunakan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Berdasarkan penelitian yang telah dilakukan oleh Ratih Nuur Azizah, dkk. Pada tahun 2024 dengan judul "Optimasi Metode Support Vector Machine Linear Dengan Algoritma Pada Klasifikasi Tingkat Obesitas". Penelitian ini bertujuan untuk mengklasifikasikan tingkat obesitas menggunakan *Support Vector Machine* (SVM) yang dioptimasi dengan *Genetic Algorithm* (GA). Optimasi menggunakan *Genetic Algorithm* mampu meningkatkan akurasi model SVM dalam mengklasifikan obesitas secara signifikan. Akurasi tertinggi yang dicapai adalah 97,9% dengan rasio split *data training* 80% dan *data testing* 20% [10].

Sedangkan penelitian yang dilakukan oleh Satferisnan Qoris Firman, dkk. Pada tahun 2025 yang berjudul "Klasifikasi Tingkat Obesitas Berdasarkan Pola Hidup dan Kebiasaan Konsumsi Makanan Menggunakan Metode K-Nearest Neighbor". Penelitian ini bertujuan untuk mengklasifikasikan tingkat obesitas berdasarkan pola hidup dan kebiasaan makan menggunakan metode KNN yang terdiri atas 1610 baris data, dengan pembagian 80% *data training* dan 20% *data testing*. Proses *preprocessing data* dilakukan dengan menggunakan SMOTE untuk mengatasi ketidak seimbangan kelas dan *Min-Max Scaling* untuk menormalisasi data. Penelitian ini juga menggunakan *K-Fold Cross Validation* untuk mengevaluasi model. Akurasi maksimum yang didapatkan adalah 79,96% menggunakan K=1[11].

2.2. *Machine Learning*

Machine Learning merupakan ranah dari *artificial intelligence* yang berperan penting dalam pengolahan data valid guna menghasilkan proses pembelajaran yang adaptif. Kemampuan sistem ini terletak pada kemampuannya dalam menganalisis data secara otomatis, sehingga memungkinkan pengambilan keputusan yang akurat serta responsif terhadap berbagai permasalahan kompleks tanpa perlu intervensi manusia secara berulang. Pembelajaran dalam machine learning didasarkan pada pengalaman dari data historis yang terus berkembang, menjadikan sistem mampu melakukan generalisasi terhadap pola-pola tertentu. Berbagai pendekatan dalam *machine learning* mencakup metode terawasi (*supervised learning*), tanpa pengawasan (*unsupervised learning*), pembelajaran penguatan (*reinforcement learning*), semi-supervised learning, online learning, serta active learning. Efektivitas model machine learning sangat dipengaruhi oleh integrasi tiga elemen utama, yakni daya komputasi perangkat keras, algoritma matematis yang digunakan, serta kualitas dan kuantitas data yang dianalisis [12].

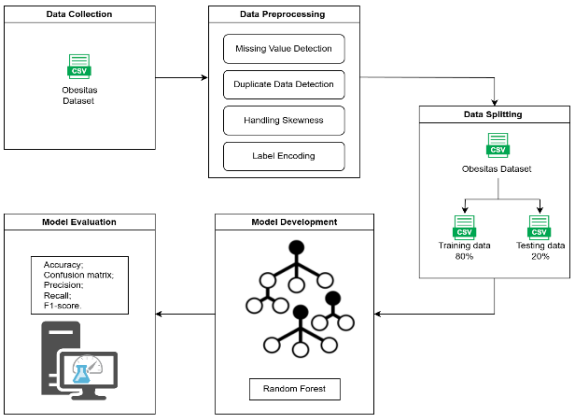
2.3. Algoritma Random Forest

Metode Random Forest merupakan bentuk pengembangan dari algoritma Decision Tree yang menggabungkan sejumlah pohon keputusan yang telah dilatih secara independen. Setiap pohon dalam model ini dibangun menggunakan sampel acak dari data latih, di mana pemilihan atribut dilakukan dari subset fitur yang dipilih secara acak untuk meningkatkan keragaman antar pohon. Keunggulan utama Random Forest terletak pada kemampuannya dalam meningkatkan akurasi prediksi, khususnya dalam kondisi data yang tidak lengkap atau mengandung nilai pencilan (*outliers*). Selain itu, metode ini efisien dalam penggunaan memori dan penyimpanan data. Salah satu fitur penting dari Random Forest adalah kemampuan dalam melakukan seleksi atribut secara otomatis, yang memungkinkan identifikasi fitur paling relevan sehingga dapat memperbaiki performa model klasifikasi secara keseluruhan. Dengan karakteristik tersebut, Random Forest terbukti efektif dalam menangani big data dan kompleksitas parameter model yang tinggi.

Random Forest (RF) merupakan pengembangan dari metode bagging yang dirancang untuk menurunkan varians model statistik serta merepresentasikan variasi data melalui pengambilan sampel *bootstrap* secara acak dari data pelatihan. Model kemudian membentuk prediksi agregat terhadap observasi baru berdasarkan hasil dari beberapa model individu. RF terbukti mampu

meningkatkan akurasi prediksi dari berbagai algoritma pembelajaran terawasi, khususnya *decision tree*. Sebagai salah satu algoritma ensemble learning berbasis pohon yang paling luas digunakan, Random Forest mengintegrasikan sejumlah pohon keputusan dengan menerapkan berbagai strategi untuk mengurangi korelasi antar pohon dalam satu himpunan model. Proses *bootstrap sampling* berperan penting dalam mengurangi bias model pohon keputusan serta memungkinkan proses pemisahan fitur secara acak di setiap simpul pohon, sehingga menghasilkan struktur pohon yang lebih beragam dan generalis [13].

3. METODE PENELITIAN



Gambar 1. Diagram Alur dalam Pengembangan Model Klasifikasi Obesitas Menggunakan Random Forest

3.1. *Data Collection*

Dataset yang digunakan diperoleh dari sumber sekunder yaitu platform Kaggle yang terdiri dari 2.111 baris data dan 17 kolom fitur. Dataset ini berisi informasi yang terkait dengan informasi umum seperti jenis kelamin, usia, tinggi badan, berat badan hingga informasi mengenai gaya hidup seperti aktivitas, konsumsi makanan, kondisi kesehatan yang dapat digunakan untuk menganalisis status obesitas. Rincian kolom fitur dari dataset yang digunakan adalah sebagai berikut.

Tabel 1. Fitur dan Tipe Data pada Dataset Obesitas

Nama Fitur	**Tipe Data**
Gender	Object
Age	Float
Height	Float
Weight	float
Family_history_with_overweight	Object
FAVC	Object
FCVC	Float
NCP	Float
CAEC	Object
SMOKE	Object
CH2O	Float
SCC	Object
FAF	Float
TUE	Float
CALC	Object
MTRANS	Object
NObeyesdad	Object

3.2. *Data Preprocessing*

Pada tahap *data preprocessing*, dilakukan beberapa tahapan untuk memastikan kualitas data sebelum memasuki proses pemodelan. Pertama, dilakukan visualisasi dan *Exploratory Data Analysis* (EDA) untuk memahami distribusi variable, hubungan antar fitur dan mendeteksi anomali atau pola yang tidak biasa. Selanjutnya adalah tahap pengecekan *missing value*, kolom akan di hapus atau diisikan secara otomatis apabila terdapat *missing value*, kemudian dilanjutkan dengan penanganan distribusi data yang tidak seimbang atau *skewness*. Kemudian dilakukan proses penghapusan data duplikat yang dapat mengganggu analisis atau pelatihan model. Tahap terakhir yaitu melakukan pemberian label pada data kategorikal sehingga data non numerik dapat diubah menjadi format numerik yang dapat diproses menggunakan *machine learning.*

3.3. Pengembangan Model Random Forest

Setelah melalui tahap *data preprocessing*, data dibagi menjadi 80% data latih dan 20% data uji. Tahap selanjutnya adalah pembuatan model Random forest menggunakan *library python* yang telah disediakan yaitu Random Forest Classifier. Langkah pertama pada tahap ini adalah membuat objek model yang didefinisikan dengan menggunakan variabel rf menggunakan RandomForestClassifier() yang merupakan algoritma *ensemble* dari beberapa pohon keputusan. Model kemudian dilatih menggunakan *data training melalui* rf.fit(x_train, y_train). Dimana x_train merupakan fitur dan y_train merupakan data target. Setelah model selesai dilatih, dilakukan prediksi terhadap data test x_test dengan perintah y_pred = rf.predict(x_test). Hasil tersebut kemudian dievaluasi dengan menghitung akurasi menggunakan accuracy_score(y_test, y_pred) yang mengukur proporsi prediksi yang benar dibandingkan dengan total data test. Nilai akurasi ditampilkan menggunakan print(“akurasi: ”, accuracy).

3.4. Evaluasi Model

Untuk mengevaluasi model yang telah dibuat, digunakan beberapa metrik evaluasi yang umum digunakan dalam klasifikasi. *Accuracy* digunakan untuk mengatur proporsi prediksi yang benar terhadap model prediksi. Selain itu, *confussion matrix* juga digunakan untuk melihat lebih detail jumlah prediksi yang benar dan salah di setiap kelas, yang membantu mengidentifikasi kesalahan spesifik dalam klasifikasi. Laporan klasifikasi yang mencakup *precision*, *recall*, dan *f1-score* digunakan untuk memberikan gambaran mengenai kinerja model.

4. HASIL DAN PEMBAHASAN

Dalam hasil dan pembahasan penulis menyajikan temuan, hasil analisis dan melakukan interpretasi dari hasil yang telah ada. Seluruh hasil disajikan dalam bentuk narasi dan visualisasi guna memberikan

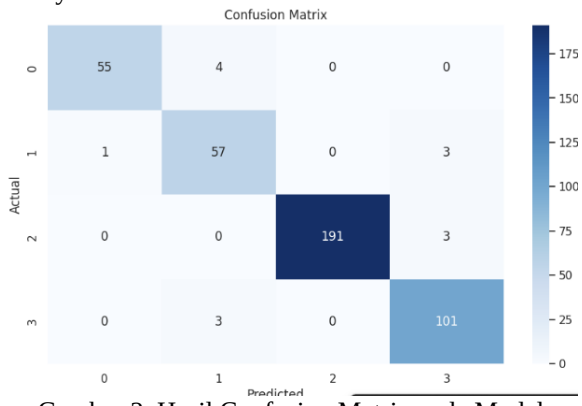
pemahaman yang mendalam terhadap topik yang sedang diteliti.

4.1. Performa Klasifikasi

Proses klasifikasi obesitas dalam penelitian ini dilakukan menggunakan model Random Forest dengan memanfaatkan dataset yang telah tersedia. Data dibagi dengan skenario pembagian 80:20, di mana 80% digunakan untuk proses pelatihan (*training*) dan 20% untuk pengujian (*testing*). Selain itu, nilai random *seed* ditetapkan pada angka 42 untuk memastikan reprodusibilitas hasil. Berdasarkan skema tersebut, model berhasil mencapai tingkat akurasi sebesar 96,65% dalam memprediksi klasifikasi obesitas.

4.2. Analisis *Confusion Matrix*

Berdasarkan Gambar 2 yang menunjukkan hasil *Confusion Matrix*, model Random Forest mampu memprediksi dengan benar sebanyak 55 data untuk kelas 0, 57 data untuk kelas 1, dan menunjukkan kinerja yang cukup signifikan pada kelas 2 dengan jumlah prediksi benar sebanyak 191 data. Sementara itu, untuk kelas 3, model berhasil mengklasifikasikan 101 data secara akurat. Secara keseluruhan, performa model dapat dikategorikan baik, khususnya pada prediksi kelas 2 yang memiliki tingkat kesalahan relatif rendah dibandingkan dengan kelas-kelas lainnya.



Gambar 2. Hasil Confusion Matrix pada Model Klasifikasi Obesitas Menggunakan Random Forest

4.3. *Classification Report*

	precision	recall	f1-score	support
Insufficient	0.98	0.93	0.96	59
Normal	0.89	0.95	0.92	61
Obesity	1.00	0.99	0.99	194
Overweight	0.96	0.97	0.97	104
accuracy			0.97	418
macro avg	0.96	0.96	0.96	418
weighted avg	0.97	0.97	0.97	418

Gambar 3. Hasil Confusion Matrix pada Model Klasifikasi Obesitas Menggunakan Random Forest

Berdasarkan Gambar 3 diatas, performa model pada masing-masing kelas menunjukkan hasil yang bervariasi. Pada kelas *Insufficient Weight*, model memperoleh nilai *precision* sebesar 0,98 dan *recall* sebesar 0,93, yang mengindikasikan bahwa model memiliki tingkat akurasi yang sangat tinggi dalam mengklasifikasikan kelas tersebut. Sementara itu, untuk kelas *Normal Weight*, terjadi sedikit penurunan performa dengan nilai *precision* sebesar 0,89 dan *recall* 0,93, menunjukkan bahwa model mengalami lebih banyak kesulitan dalam memprediksi kelas ini dibandingkan kelas lainnya. Pada kelas *Obesity*, model menunjukkan performa terbaik dengan nilai *precision* sebesar 1,00, *recall* 0,99, dan F1-score 0,99, yang mencerminkan tingkat keandalan yang sangat tinggi dalam klasifikasi kelas tersebut. Untuk kelas *Overweight*, model tetap menunjukkan kinerja yang baik dengan *precision* sebesar 0,96 dan *recall* sebesar 0,97.

Secara keseluruhan, akurasi model mencapai 97%, yang berarti bahwa dari 418 data yang tersedia, model mampu melakukan klasifikasi dengan tingkat ketepatan sebesar 97%. Selain itu, nilai *macro average*, yang merepresentasikan rata-rata performa model tanpa mempertimbangkan proporsi data pada setiap kelas, menunjukkan nilai stabil sebesar 0,96. Adapun nilai *weighted average*, yang memperhitungkan distribusi data pada tiap kelas, juga menunjukkan hasil yang memuaskan sebesar 0,97, mengindikasikan bahwa model bekerja secara konsisten dan efektif meskipun terdapat ketidakseimbangan jumlah data antar kelas.

4.4. Pembahasan

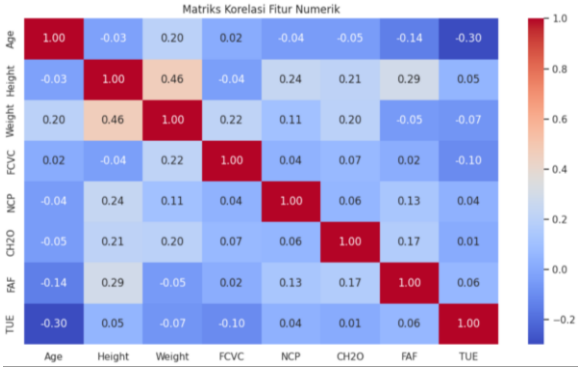
Model Random Forest menunjukkan kinerja yang sangat baik dalam mengklasifikasikan tingkat obesitas, sehingga membuktikan bahwa algoritma ini sangat cocok digunakan untuk data terstruktur, khususnya data bertipe tabular seperti pada dataset yang digunakan dalam penelitian ini. Sebagai algoritma berbasis ansambel, Random Forest memiliki kemampuan yang kuat dalam menangani distribusi kelas yang tidak seimbang secara efektif. Kesalahan klasifikasi yang terjadi kemungkinan disebabkan oleh adanya tumpang tindih antar kelas, khususnya pada fitur-fitur seperti indeks massa tubuh (BMI), tingkat aktivitas fisik, dan pola konsumsi makanan.

Jika dibandingkan dengan algoritma klasifikasi lainnya seperti *Naive Bayes* atau *Decision Tree* tunggal, Random Forest menawarkan generalisasi yang lebih baik dan ketahanan terhadap permasalahan *overfitting*. Hal ini disebabkan oleh mekanisme agregasi prediksi dari sejumlah pohon keputusan independen yang dilatih secara terpisah, sehingga mengurangi korelasi antar pohon dan meningkatkan stabilitas serta akurasi model secara keseluruhan.

4.5. Analisis Korelasi Fitur Numerik terhadap Obesitas

Analisis korelasi antar fitur numerik merupakan salah satu tahap penting dalam memahami hubungan antar variabel prediktor yang digunakan dalam pemodelan klasifikasi obesitas. Korelasi dihitung menggunakan metode Pearson dan divisualisasikan

dalam bentuk *heatmap* sebagaimana ditampilkan pada Gambar 4. Analisis ini bertujuan untuk mengevaluasi adanya potensi multikolinearitas serta menggali makna interpretatif dari keterkaitan antar fitur terhadap perilaku dan kondisi fisiologis yang dapat memengaruhi status obesitas.



Gambar 4. Matriks Korelasi Antar Variabel Numerik dalam Dataset Obesitas

Hasil analisis menunjukkan bahwa korelasi tertinggi terjadi antara variabel *Weight* (berat badan) dan *Height* (tinggi badan) dengan nilai koefisien sebesar 0.46, yang merepresentasikan hubungan positif moderat. Korelasi ini mencerminkan fakta fisiologis bahwa individu dengan tinggi badan yang lebih besar cenderung memiliki berat badan yang lebih tinggi pula. Meskipun variabel *Body Mass Index* (BMI) tidak secara eksplisit digunakan dalam penelitian ini, keterkaitan antara *Height* dan *Weight* menjadi relevan karena keduanya merupakan komponen utama dalam perhitungan BMI yang umum digunakan untuk mengidentifikasi obesitas. Dengan demikian, keberadaan korelasi ini mendukung keabsahan pemilihan fitur dalam model klasifikasi.

Selanjutnya, ditemukan pula korelasi negatif moderat antara variabel *Age* (usia) dan TUE (*Time Using Technology*) dengan nilai sebesar -0.30. Hubungan ini mengindikasikan bahwa individu dengan usia yang lebih tinggi cenderung menghabiskan waktu yang lebih sedikit untuk menggunakan perangkat digital. Fenomena ini dapat diinterpretasikan sebagai cerminan dari perbedaan pola aktivitas harian antar kelompok usia, di mana individu yang lebih muda cenderung memiliki paparan teknologi yang lebih tinggi dan berpotensi menjalani gaya hidup sedentari yang lebih intens. Dalam konteks prediksi obesitas, variabel TUE dapat berfungsi sebagai indikator tidak langsung dari tingkat aktivitas fisik, yang merupakan determinan penting dalam akumulasi lemak tubuh.

Sebagian besar korelasi antar fitur numerik lainnya menunjukkan nilai yang relatif rendah, berada di bawah ambang ±0.30. Sebagai contoh, hubungan antara FAF (frekuensi aktivitas fisik) dan CH_2O (konsumsi air harian) memiliki nilai korelasi sebesar 0.17, sedangkan antara NCP (jumlah konsumsi makanan per hari) dan FCVC (frekuensi konsumsi sayuran) hanya sebesar 0.04. Rendahnya korelasi antar fitur ini menunjukkan bahwa tidak terjadi tumpang tindih informasi yang signifikan di antara variabel, yang berarti setiap fitur memberikan kontribusi unik terhadap proses klasifikasi. Karakteristik ini sangat menguntungkan dalam penerapan algoritma Random Forest, yang mengandalkan keberagaman pohon keputusan yang dilatih pada subset fitur yang berbeda.

Lebih lanjut, hasil korelasi ini menunjukkan bahwa tidak terdapat indikasi kuat adanya multikolinearitas, karena tidak ditemukan pasangan variabel dengan koefisien korelasi melebihi 0.80. Meskipun Random Forest secara teoritis cukup tangguh dalam menangani fitur-fitur yang berkorelasi tinggi, keberadaan fitur yang relatif independen tetap memberikan keuntungan, baik dari sisi efisiensi komputasi, interpretabilitas model, maupun akurasi prediksi.

Temuan ini mengindikasikan bahwa fitur numerik yang digunakan dalam penelitian memiliki keragaman dan nilai informasi yang tinggi, yang dapat memperkuat kapabilitas generalisasi model. Keberagaman tersebut juga memberikan justifikasi terhadap performa tinggi yang dicapai model Random Forest dalam mengklasifikasikan tingkat obesitas. Ke depan, disarankan untuk melengkapi analisis ini dengan evaluasi korelasi antara masing-masing fitur terhadap variabel target klasifikasi obesitas. Hal ini dapat dilakukan melalui pendekatan statistik seperti ANOVA atau uji Kruskal-Wallis, atau melalui ekstraksi feature importance dari model, guna memperoleh pemahaman yang lebih mendalam terkait kontribusi relatif tiap fitur dalam proses pengambilan keputusan model.

Dengan demikian, analisis korelasi ini tidak hanya berfungsi sebagai dasar eksploratif awal, tetapi juga memberikan kontribusi nyata dalam validasi struktur data, interpretasi klinis, serta penguatan arsitektur pemodelan prediktif dalam studi klasifikasi obesitas berbasis *machine learning*.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma Random Forest menunjukkan performa yang sangat baik dalam mengklasifikasikan tingkat obesitas individu berdasarkan kombinasi fitur fisiologis dan kebiasaan hidup, dengan akurasi mencapai 97% serta nilai *precision* dan *recall* yang tinggi pada sebagian besar kelas. Proses pengolahan data yang mencakup penanganan nilai hilang, deteksi *outlier*, transformasi distribusi, normalisasi, dan *encoding* kategori turut berkontribusi terhadap kinerja model. Visualisasi melalui *confusion matrix* memperkuat temuan bahwa model mampu melakukan prediksi secara akurat di berbagai kategori tingkat obesitas. Untuk pengembangan lebih lanjut, Penulis sarankan agar model diperluas dengan data yang lebih beragam dan representatif, serta ditambah dengan fitur relevan lainnya seperti pola tidur dan konsumsi makanan guna meningkatkan kompleksitas dan

ketajaman prediksi. Selain itu, penerapan model alternatif seperti *neural network* atau *deep learning* juga direkomendasikan dalam penelitian selanjutnya untuk mengeksplorasi potensi arsitektur pembelajaran yang lebih kompleks dan adaptif terhadap variabilitas data, sehingga hasilnya dapat memberikan kontribusi yang lebih luas bagi masyarakat dalam konteks pencegahan dan deteksi dini obesitas.

DAFTAR PUSTAKA

[1] E. Safariyani, F. Ainun Nisah, and N. Rahmawati, "Upaya Hidup Sehat Melalui Penyuluhan Bahaya Obesitas Pada Balita," vol. 6, 2022.

[2] H. T. Santoso, F. A. Felmidi, A. N. Fadhila, A. Ristyawan, and E. Daniati, "Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC," Online, 2024.

[3] A. H. Santoso, M. S. Kartolo, T. P. Alifia, K. F. Kusuma, F. C. Gunaidi, and J. Kurniawan, "Pelayanan Skrining Obesitas dan Obesitas Sentral pada Populasi Lanjut Usia melalui Pengukuran Indeks Massa Tubuh Dan Lingkar Pinggang," *Karunia: Jurnal Hasil Pengabdian Masyarakat Indonesia*, vol. 3, no. 2, pp. 62–68, Jun. 2024, doi: 10.58192/karunia.v3i2.2150.

[4] D. Amelia and R. Kurniawan, "Penerapan Algoritma Random Forest Untuk Prediksi Pejualan Dan Persedian Produk Pada Toko Frozen Food Anisa," *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 2, pp. 843–848, 2025.

[5] I. Yustisia, L. B. Kurniawan, T. Esa, Syahrijuita, and S. A. Thamrin, "The relationship between age, obesity indices, and cardiometabolic risk factors in Women: Findings from a point-of-care health screening in South Sulawesi, Indonesia," *Clin Epidemiol Glob Health*, vol. 33, May 2025, doi: 10.1016/j.cegh.2025.102048.

[6] D. Vidiawati *et al.*, "Excess body weight and its associated factors among first-year health sciences university students in Indonesia," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0322773.

[7] I. A. Liberty *et al.*, "The characteristics and risk of obesity central and concomitant impaired fasting glucose: Findings from a cross-sectional study," *PLoS One*, vol. 19, no. 6 June, Jun. 2024, doi: 10.1371/journal.pone.0305604.

[8] E. Widhiastuti, M. H. Huda, H. Susanto, M. D. Kurniasari, and H. E. Putra, "The Synergistic Effect Of High Bmi And Low Physical Activity On Gout Arthritis Risk: A Case -Control Study In West Sumatera Indonesia," *Menara Medika*, vol. 7, no. 2, pp. 422–435, Mar. 2025, doi: 10.31869/mm.v7i2.6555.

[9] T. M. Pratamawati *et al.*, "Clinical profile of hypertension patients in primary health Care in Cirebon Regency, Indonesia," *Endocrine and Metabolic Science*, vol. 18, Jun. 2025, doi: 10.1016/j.endmts.2025.100232.

[10] R. N. Azizah, B. Rahmat, and H. Maulana, "Optimasi Metode Support Vector Machine Linear Dengan Algoritma Pada Klasifikasi Tingkat Obesitas," *Jurnal Ilmiah Wahana Pendidikan, Desember*, vol. 2024, no. 24, pp. 469–477.

[11] F. Reynaldi Valerian, M. Syarief, D. Abdul Fatah, J. Raya Telang, K. Kamal, and J. Timur, "Klasifikasi Tingkat Obesitas Menggunakan Metode Gbm Dan Confusion Matrix," 2025.

[12] A. Rachman Andhika, "Machine Learning Dalam Pengembangan Perangkat Lunak," *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 1, p. 1211, 2025.

[13] I. Kurniawan, D. Cahya Putri Buani, W. Apriliah, R. Amegia Saputra, and P. Korespondensi, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin Implementation Of Random Forest Algorithm For Determining Recipients Of Raskin," vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.202396225.