Modem: Jurnal Informatika dan Sains Teknologi
Volume 4, Nomor 3, Juli 2026
e-ISSN: 3046-7209; p-ISSN: 3046-7217, Hal. 93-105
DOI: https://doi.org/10.62951/modem.v4i3.955
Tersedia: https://journal.aptii.or.id/index.php/Modem

Komparasi Algoritma *Machine Learning* dengan *Hyperparameter Tuning* pada Prediksi Level Obesitas

Dwi Utami[1*], Fathoni Dwiatmoko[2], Alinsa Novsagita[3]
[1,3]Sistem Informasi, Universitas Muhammadiyah Kalianda, Indonesia
[2]Informatika, Universitas Indonesia Mandiri, Indonesia
**Penulis Korespondensi: utami21.du@gmail.com*

Abstract*. Obesity has become a growing global health concern due to its association with various chronic diseases, including diabetes mellitus, hypertension, and cardiovascular disorders. This study investigates the performance of Support Vector Machine (SVM), Random Forest, and Gradient Boosting algorithms enhanced through hyperparameter tuning for obesity level prediction. The study utilized the Obesity Level dataset, which contains 2,111 instances, 16 input features, and one target class. Following the data preprocessing stage, 2,087 instances were included in the analysis. Correlation analysis was conducted to identify the nine most relevant features used for model development. Experimental results demonstrate that the Random Forest model achieved the best predictive performance compared to SVM and Gradient Boosting, with an accuracy of 90.91%, precision of 91.19%, recall of 90.91%, and an F1-score of 90.94%. The dataset was partitioned into training and testing subsets using an 80:20 split. Furthermore, the optimal model was deployed in a Streamlit-based web application to facilitate obesity level prediction. These findings suggest that Random Forest, when optimized through hyperparameter tuning, provides a reliable and effective approach for multiclass obesity classification*.

Keywords*: Gradient Boosting; Obesity Level; Optimization; Random Forest; SVM.*

Abstrak. Obesitas telah menjadi masalah kesehatan global yang kian meningkat karena kaitannya dengan berbagai penyakit kronis, termasuk diabetes melitus, hipertensi, dan gangguan kardiovaskular. Penelitian ini mengkaji kinerja algoritma *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* yang ditingkatkan melalui penyesuaian hiperparameter (*hyperparameter tuning*) untuk prediksi tingkat obesitas. Penelitian ini menggunakan dataset Obesity Level yang mencakup 2.111 instans, 16 fitur masukan, dan satu kelas target. Setelah tahap pra-pemrosesan data, sebanyak 2.087 instans disertakan dalam analisis. Analisis korelasi dilakukan untuk mengidentifikasi sembilan fitur paling relevan yang digunakan dalam pengembangan model. Hasil eksperimen menunjukkan bahwa model *Random Forest* mencapai kinerja prediksi terbaik dibandingkan dengan SVM dan *Gradient Boosting*, dengan akurasi 90,91%, presisi 91,19%, recall 90,91%, dan skor F1 90,94%. Dataset dibagi menjadi subset pelatihan dan pengujian dengan rasio 80:20. Selanjutnya, model optimal tersebut diimplementasikan dalam aplikasi web berbasis Streamlit untuk memfasilitasi prediksi tingkat obesitas. Temuan ini menunjukkan bahwa *Random Forest*, jika dioptimalkan melalui penyesuaian hiperparameter, memberikan pendekatan yang andal dan efektif untuk klasifikasi obesitas multikelas.

Kata Kunci: *Gradient Boosting*; Level Obesitas; Optimasi; *Random Forest*; SVM.

1. LATAR BELAKANG

Penumpukan lemak tubuh yang berlebihan sebagai hasil dari ketidakseimbangan energi yang dikonsumsi dan dikeluarkan dalam jangka waktu yang lama dikenal sebagai obesitas. Karena prevalensinya yang terus meningkat dan efeknya terhadap kualitas hidup, kondisi ini telah menjadi salah satu masalah kesehatan global. Obesitas, yang ditandai dengan peningkatan berat badan yang tidak normal, juga merupakan faktor risiko utama diabetes melitus, hipertensi, dan penyakit kardiovaskular. (Wie & Siddik, 2022). Jumlah kasus obesitas terus meningkat, menjadikannya salah satu masalah utama dalam bidang kesehatan masyarakat. Obesitas meningkatkan risiko berbagai penyakit kronis dan dapat dialami oleh orang dari berbagai kelompok usia. Oleh karena itu, deteksi dini dan edukasi yang mudah diakses diperlukan.

Naskah Masuk: 04 Mei 2026; Revisi: 02 Juni 2026; Diterima: 02 Juli 2026; Terbit: 21 Juli 2026

Pengembangan sistem prediksi berbasis pembelajaran mesin adalah salah satu pendekatan yang dapat digunakan. Sistem ini akan menggunakan algoritma klasifikasi untuk menemukan tingkat obesitas berdasarkan data yang tersedia. Diharapkan sistem ini akan membantu masyarakat melakukan pemeriksaan awal sebelum mendapatkan perawatan lebih lanjut dari tenaga medis. (Sibi et al., 2022). Dalam pengembangan sistem prediksi, penggunaan algoritma pembelajaran mesin adalah salah satu teknik yang banyak digunakan. Algoritma klasifikasi *Support Vector Machine* (SVM) mencari hyperplane yang ideal untuk memisahkan data ke dalam kelas yang berbeda. Untuk melakukan proses ini, jarak (margin) antara hyperplane dan data terdekat dari masing-masing kelas dimaksimalkan, yang menghasilkan klasifikasi yang lebih akurat. Karena kemampuan untuk menangani data yang besar dan pola data yang kompleks, SVM menjadi salah satu algoritma yang paling banyak digunakan dalam berbagai penelitian klasifikasi (Setiawan et al., 2025).

Gradient Boosting adalah algoritma lain yang banyak digunakan untuk tugas klasifikasi selain SVM. GB adalah metode pembelajaran kelompok yang membangun model prediksi dengan menggabungkan berbagai pohon keputusan (decision tree) secara bertahap. Setiap pohon baru dibuat untuk memperbaiki kesalahan prediksi dari pohon sebelumnya, meningkatkan kinerja model. Proses pembelajaran difokuskan pada pengurangan fungsi kerugian (loss function). GB berfungsi dengan baik untuk berbagai masalah klasifikasi, seperti prediksi tingkat obesitas, karena kemampuan untuk menangani hubungan yang kompleks antar fitur (Valerian et al., 2025). *Random Forest*, bersama dengan *Gradient Boosting*, adalah salah satu algoritma ensemble learning yang paling banyak digunakan untuk tugas klasifikasi. Berbeda dengan *Gradient Boosting*, yang secara bertahap membangun banyak pohon keputusan untuk memperbaiki kesalahan model sebelumnya. *Random Forest* membangun banyak pohon keputusan secara independen menggunakan teknik aggregating (bagging) bootstrap dan pemilihan fitur secara acak.

Untuk menentukan kelas akhir, mekanisme majority voting digunakan untuk menggabungkan hasil prediksi dari setiap pohon. Metode ini mengurangi risiko overfitting dengan memberi *Random Forest* kemampuan untuk menangkap hubungan nonlinier. (Syahputra et al., 2025). Penelitian ini bertujuan untuk membandingkan kinerja tiga algoritma pembelajaran mesin: *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting*, dalam memprediksi tingkat obesitas. *GridSearchCV* digunakan untuk mengoptimalkan setiap algoritma untuk menghasilkan kombinasi hyperparameter yang ideal, yang dapat meningkatkan kinerja model. Selanjutnya, metrik evaluasi yang digunakan digunakan untuk menganalisis dampak penyesuaian hyperparameter terhadap hasil klasifikasi. Selanjutnya,

model dengan kinerja terbaik digunakan dengan Streamlit. Hasil penelitian ini diharapkan dapat membantu proses prediksi level obesitas secara lebih dini dan membantu proses deteksi awal untuk pengambilan keputusan.

2. KAJIAN TEORITIS

Obesitas

Kelebihan lemak tubuh disebut obesitas, yang dapat menurunkan kualitas hidup dan meningkatkan risiko penyakit kronis seperti diabetes melitus tipe 2, penyakit kardiovaskular, dan beberapa kanker. Selain itu, obesitas juga dapat mengganggu kesehatan tulang, sistem reproduksi, dan kualitas tidur. (Khikam et al., 2025). World Health Organization (WHO) menyatakan bahwa status berat badan biasanya diukur dengan menggunakan Indeks Massa Tubuh (IMT) atau Indeks Massa Tubuh (BMI), yang merupakan perbandingan antara berat badan (kg) dan kuadrat tinggi badan (m2). Berdasarkan nilai BMI, status berat badan dikategorikan ke dalam kategori normal, kelebihan berat badan, atau obesitas tingkat I, II, dan III. Obesitas adalah salah satu penyakit jangka panjang yang semakin umum dan dipengaruhi oleh banyak faktor, termasuk ketidakseimbangan antara asupan dan pengeluaran energi, faktor genetik, gaya hidup, dan kondisi psikologis (Ibrahim & Atim, 2024).

Salah satu cara penting untuk mencegah dan menangani obesitas adalah melakukan aktivitas fisik.(Saputra et al., 2024). Aktivitas fisik yang dilakukan secara teratur dapat meningkatkan pengeluaran energi, memperbaiki metabolisme tubuh, dan menjaga keseimbangan antara asupan dan penggunaan energi, yang membantu mempertahankan berat badan ideal dan mencegah penumpukan lemak tubuh secara berlebihan. Sebaliknya, kurangnya aktivitas fisik dapat mengurangi pengeluaran energi, yang meningkatkan risiko kelebihan berat badan.

Machine Learning

Dengan menganalisis data untuk menemukan hubungan atau pola yang dapat digunakan dalam proses prediksi dan klasifikasi, *Machine Learning* adalah bagian dari Artificial Intelligence (AI) yang memungkinkan sistem komputer tanpa diprogram secara eksplisit untuk mempelajari pola data dan membuat prediksi atau keputusan secara otomatis untuk setiap situasi (Wijoyo et al., 2024). Dataset biasanya dibagi menjadi data pengujian (testing data) dan data pelatihan (training data). Data pengujian digunakan untuk mengevaluasi kemampuan model untuk menggeneralisasi dan memprediksi data baru yang belum pernah dipelajari sebelumnya, sedangkan data pelatihan digunakan untuk membuat model (Retnoningsih & Pramudita, 2020).

SVM

Salah satu algoritma pembelajaran mesin terawasi yang paling umum digunakan untuk tugas regresi dan klasifikasi adalah *Support Vector Machine* (SVM). Menentukan hyperplane yang ideal memungkinkan algoritma ini untuk memisahkan data ke dalam berbagai kelas dengan margin maksimum. Agar dapat membangun model klasifikasi terbaik, SVM memerlukan data berlabel selama proses pelatihan karena merupakan metode pembelajaran terawasi. SVM telah lama dikenal memiliki kemampuan untuk menangani data besar dan membuat prediksi yang akurat, terutama untuk masalah klasifikasi yang kompleks (Utami et al., 2025).

Gradient Boosting

Salah satu algoritma kelompok pembelajaran yang digunakan untuk tugas regresi dan klasifikasi adalah *Gradient Boosting*. Dengan menambahkan beberapa pohon keputusan secara bertahap, algoritma ini membangun model secara bertahap. Setiap pohon yang ditambahkan bertujuan untuk memperbaiki kesalahan (residual) yang dihasilkan oleh model sebelumnya (Nababan et al., 2025). Proses pembelajaran dilakukan dengan mengoptimalkan fungsi kerugian (loss function) menggunakan pendekatan turun gradient. Ini memungkinkan model untuk secara bertahap meningkatkan akurasi prediksinya. Pengajaran rasio mengontrol kontribusi setiap pohon. Proses iterasi berlanjut hingga mencapai jumlah iterasi yang ditetapkan atau kondisi konvergen. *Gradient Boosting* dikenal sebagai salah satu algoritma yang sangat baik untuk menyelesaikan berbagai masalah pembelajaran mesin karena kemampuan untuk menangani hubungan yang kompleks antara berbagai fitur dan karakteristik data (Valerian et al., 2025).

Random Forest

Salah satu algoritma pembelajaran mesin yang menggunakan kumpulan pohon keputusan untuk menyelesaikan tugas regresi dan klasifikasi adalah *Random Forest*. Algoritma ini menghasilkan sejumlah pohon keputusan dengan menggunakan teknik aggregating bootstrap, juga dikenal sebagai bagging, dan menggunakan pemilihan fitur secara acak. Kemudian, mekanisme suara mayoritas pada klasifikasi atau rata-rata pada regresi digunakan untuk menggabungkan hasil prediksi dari setiap pohon. *Random Forest* banyak digunakan pada berbagai penelitian klasifikasi, termasuk di bidang kesehatan, karena kemampuan metode ini untuk meningkatkan akurasi, mengurangi risiko overfitting, dan menangani hubungan nonlinier antar fitur (Syahputra et al., 2025).

Hyperparameter tuning

Tuning hyperparameter adalah proses optimasi yang dilakukan untuk menemukan kombinasi hyperparameter terbaik untuk meningkatkan kinerja model pembelajaran mesin. Proses ini dilakukan dengan menilai berbagai kombinasi parameter untuk menemukan konfigurasi yang memberikan kinerja terbaik (Rusman et al., 2023). Penelitian ini menggunakan *GridSearchCV* dari pustaka Scikit-learn untuk melakukan optimasi. *GridSearchCV* menguji setiap kombinasi hyperparameter yang telah ditentukan. Metode cross-validation digunakan untuk mengevaluasi setiap kombinasi. Hasil evaluasi terbaik digabungkan untuk memilih konfigurasi model yang ideal (Muhamad & Matin, 2023).

3. METODE PENELITIAN

Pengumpulan Data

Penelitian menggunakan Dataset Level Obesity, yang diperoleh dari Kaggle: https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels, terdiri dari 2.111 data, 16 atribut fitur, dan 1 atribut target (NObeyesdad). Usia (usia), jenis kelamin (jenis kelamin), tinggi badan (tinggi), berat badan, riwayat keluarga dengan kelebihan berat badan (riwayat keluarga dengan kelebihan berat badan), konsumsi makanan berkalori tinggi (FAVC), frekuensi konsumsi sayuran (FCVC), jumlah makan utama per hari (NCP), kebiasaan mengonsumsi makanan di antara waktu makan (CAEC), kebiasaan merokok (SMOKE), konsumsi air putih (CH2O), pemantauan konsumsi kalori (SCC), dan aktivitas fisik. Sebelum digunakan dalam proses pemodelan, dataset harus melalui tahap preprocessing. Ini dilakukan untuk meningkatkan kualitas data dan membantu proses klasifikasi dengan lebih efisien.

Preprocessing Data

Sebelum pemodelan dilakukan, proses untuk meningkatkan kualitas data termasuk dalam tahap preprocessing penelitian ini. Data bersih adalah tahap pertama. Ini termasuk membersihkan data duplikat, menangani nilai yang hilang, dan memperbaiki data yang tidak konsisten. Selanjutnya, label encoding digunakan untuk mengubah atribut kategorikal menjadi format numerik sehingga algoritma pembelajaran mesin dapat memprosesnya. Selanjutnya, analisis data eksploratif (EDA) dilakukan untuk memahami karakteristik data, menganalisis distribusi setiap atribut, menemukan outlier, dan menemukan kemungkinan masalah data. Selanjutnya, seleksi fitur dilakukan dengan menggunakan analisis korelasi. Tujuan dari proses ini adalah untuk menemukan atribut dengan hubungan terkuat dengan variabel target, sehingga hanya atribut yang relevan digunakan dalam proses pemodelan. Menurut hasil analisis korelasi, sembilan atribut khusus diidentifikasi untuk digunakan sebagai variabel tambahan dalam

model klasifikasi. Selanjutnya, skala fitur dilakukan dengan menggunakan StandardScaler. Proses ini bertujuan untuk menyamakan skala antar fitur, terutama untuk meningkatkan kinerja algoritma *Support Vector Machine* (SVM) yang sensitif terhadap perbedaan skala. Setelah itu, model klasifikasi dilatih dan diuji dengan data yang telah melewati seluruh tahapan preprocessing, dengan nilai rata-rata (*mean*) sebesar 0 dan standar deviasi sebesar 1.

Pembagian Data

Setelah seluruh tahapan preprocessing selesai, set data dibagi menjadi dua bagian: set pelatihan (*training set*) dan set pengujian. Tujuan dari kedua bagian ini adalah agar model dapat dilatih dengan sebagian besar data, kemudian dievaluasi dengan data yang tidak terlibat dalam proses pelatihan, sehingga dapat diukur secara objektif tingkat generalisasi model. Untuk membagi data, penelitian ini menggunakan metode split train-test dengan rasio 80:20. Dari 2.087 data yang dikumpulkan setelah proses pembersihan data, 1.669 digunakan sebagai data pelatihan, dan 418 digunakan sebagai data pengujian. Selama setiap proses pengujian, agar hasil pembagian data tetap konsisten, dengan menggunakan parameter random state = 42.

Pemodelan

Penelitian ini membandingkan kinerja tiga algoritma pembelajaran mesin: *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting*, dalam memprediksi tingkat obesitas pada tahap pemodelan. Ketiga algoritma tersebut dipilih karena masing-masing memiliki karakteristik dan keunggulan yang berbeda yang dapat digunakan untuk menyelesaikan masalah klasifikasi. Semua orang tahu bahwa SVM sangat bagus untuk menangani data besar dan menemukan batas pemisah (hyperplane) yang ideal. *Random Forest* dipilih karena metode pembelajaran ensemble berbasis baggingnya dapat meningkatkan akurasi dan mengurangi risiko overfitting. Di sisi lain, Graded Boosting memungkinkan performa klasifikasi yang tinggi karena membangun model secara bertahap dengan memperbaiki kesalahan prediksi pada setiap iterasi. Perbandingan ketiga algoritma dilakukan untuk menentukan model yang memiliki kinerja terbaik berdasarkan metrik evaluasi yang digunakan.

Hyperparameter Tuning

Setelah pemodelan selesai, *GridSearchCV* digunakan untuk mengoptimalkan setiap algoritma untuk mendapatkan kombinasi hyperparameter yang optimal. Metode ini bekerja dengan menguji secara menyeluruh seluruh kombinasi parameter yang telah ditentukan, kemudian menggunakan teknik cross-validation untuk mengevaluasi kinerja masing-masing kombinasi. Kombinasi hyperparameter yang menghasilkan nilai evaluasi terbaik dipilih sebagai konfigurasi optimal untuk setiap algoritma. Sebelum tahap evaluasi dan perbandingan

hasil dimulai, metode penyesuaian hyperparameter diharapkan dapat meningkatkan kemampuan generalisasi, akurasi, dan performa klasifikasi dari algoritma *Random Forest*, *Gradient Boosting*, dan *Support Vector Machine* (SVM).

Evaluasi Model

Tahap evaluasi model dilakukan untuk mengukur dan membandingkan kinerja *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* algoritma dalam memprediksi tingkat obesitas. Pengaruh proses optimasi terhadap kinerja masing-masing algoritma dipelajari secara menyeluruh melalui evaluasi yang dilakukan pada dua situasi berbeda, yaitu sebelum penerapan *hyperparameter tuning* dan setelah penerapan. Ini dilakukan dengan menggunakan *GridSearchCV*. Untuk mengevaluasi kinerja model, metrik klasifikasi seperti ketepatan, ketepatan, recall, dan skor F1 digunakan. Hasil evaluasi dari ketiga algoritma kemudian dibandingkan untuk menentukan model dengan kinerja terbaik.

Implementasi dengan Streamlit

Setelah proses evaluasi selesai, model dengan performa terbaik dipilih untuk diimplementasikan menggunakan Streamlit. Streamlit merupakan framework yang memungkinkan pengembangan aplikasi web secara sederhana untuk mengintegrasikan model *Machine Learning*. Framework ini menyediakan antarmuka pengguna yang interaktif dan mudah digunakan, sehingga pengguna dapat memasukkan data, menjalankan proses prediksi, serta melihat hasil prediksi secara cepat dan informatif (Syafarina & Zaenuddin, 2023). Implementasi ini bertujuan untuk membangun aplikasi prediksi level obesitas yang mampu menerima data masukan dari pengguna dan menghasilkan prediksi secara cepat berdasarkan model hasil pelatihan. Aplikasi tersebut diharapkan dapat mempermudah proses penggunaan model klasifikasi serta mendukung identifikasi awal tingkat obesitas sebagai alat bantu dalam pengambilan keputusan.

4. HASIL DAN PEMBAHASAN

Evaluasi Model Tanpa *Hyperparameter tuning*

Pada awal penelitian, pengujian dilakukan terhadap tiga algoritma pengajaran mesin: *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting*, yang semuanya menggunakan parameter bawaan. Sebelum memulai proses optimasi hyperparameter, tujuan pengujian ini adalah untuk mengetahui seberapa baik masing-masing algoritma dalam memprediksi level obesitas pada awalnya. Menilai ketepatan, ketepatan, recall, dan skor F1 digunakan untuk menilai model. Tabel 1 menunjukkan hasil pengujian.

Tabel 1.Hasil Evaluasi Model Tanpa *Hyperparameter tuning*.

Algoritma	Accuracy	Precision	Recall	F1-Score
SVM	80,38%	80,72%	80,38%	80,40%
Random Forest	91,63%	91,82%	91,63%	91,63%
Gradient Boosting	89,47%	89,76%	89,47%	89,53%

Berdasarkan Tabel 1, terlihat bahwa *Random Forest* memberikan performa terbaik dibandingkan algoritma lainnya. Hasil menunjukkan bahwa *Random Forest* sangat baik dalam mengklasifikasikan tingkat obesitas pada dataset yang digunakan. Algoritma ini memperoleh nilai accuracy sebesar 91,63%, precission sebesar 91,82%, recall sebesar 91,63%, dan skor F1 sebesar 91,63%. Performa *Random Forest* yang lebih tinggi dibandingkan algoritma lainnya dipengaruhi oleh karakteristiknya sebagai algoritma ensemble learning yang menggabungkan banyak decision tree melalui teknik bootstrap aggregating (bagging). Pendekatan ini mampu mengurangi variansi model, meningkatkan kemampuan generalisasi, serta mengurangi risiko overfitting. Selain itu, proses pemilihan subset data dan fitur secara acak pada setiap pohon keputusan membuat *Random Forest* lebih efektif dalam menangkap hubungan nonlinier antar fitur yang terdapat pada data obesitas.

Algoritma *Gradient Boosting* menempati urutan kedua dengan accuracy sebesar 89,47%, precision sebesar 89,76%, recall sebesar 89,47%, dan F1-score sebesar 89,53%. Hasil tersebut menunjukkan bahwa *Gradient Boosting* juga memiliki kemampuan klasifikasi yang baik. Sebagai algoritma boosting, *Gradient Boosting* membangun model secara bertahap dengan memperbaiki kesalahan prediksi pada iterasi sebelumnya. Pendekatan ini menghasilkan model yang akurat, meskipun memerlukan proses pelatihan yang lebih kompleks dibandingkan *Random Forest*. Sementara itu, *Support Vector Machine* (SVM) memperoleh performa paling rendah di antara ketiga algoritma dengan accuracy sebesar 80,38%, precision sebesar 80,72%, recall sebesar 80,38%, dan F1-score sebesar 80,40%. Meskipun SVM dikenal efektif dalam menangani data berdimensi tinggi, performanya pada penelitian ini masih berada di bawah *Random Forest* dan *Gradient Boosting*. Hal ini menunjukkan bahwa pemisahan kelas pada dataset obesitas memiliki pola yang cukup kompleks sehingga pendekatan berbasis hyperplane belum mampu memberikan hasil yang optimal menggunakan parameter bawaan.

Hasil Evaluasi Model dengan *Hyperparameter tuning*

Untuk mengetahui pengaruh optimasi parameter terhadap performa model, ketiga algoritma *Support Vector Machine* (SVM), *Random Forest*, dan *Gradient Boosting* dievaluasi kembali setelah proses pengaturan hyperparameter menggunakan *GridSearchCV* selesai. Menilai akurasi, ketepatan, recall, dan skor F1 digunakan. Tabel 2 menunjukkan hasil pengujian yang dilakukan setelah proses pengaturan hyperparameter.

Tabel 2. Hasil Evaluasi Model dengan *Hyperparameter tuning.*

Algoritma	Accuracy	Precision	Recall	F1-Score
SVM	83,01%	83,31%	83,01%	82,78%
Random Forest	90,91%	91,19%	90,91%	90,94%
Gradient Boosting	89,71%	89,96%	89,71%	89,77%

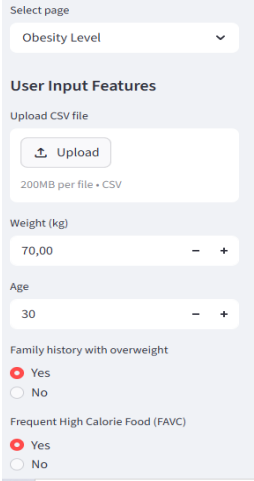
Tabel 2 menunjukkan bahwa algoritma *Random Forest* tetap memiliki kinerja terbaik setelah penyesuaian hyperparameter. Dengan nilai akurasi 90,91%, precission 91,19%, recall 90,91%, dan skor F1 90,94%, model ini masih menjadi algoritma prediksi obesitas dengan klasifikasi terbaik. Hasil menunjukkan bahwa *Random Forest* masih dapat mengklasifikasikan data multikelas dengan baik meskipun telah dilakukan proses optimasi parameter. Algoritma *Gradient Boosting* menempati posisi kedua dengan accuracy sebesar 89,71%, precision sebesar 89,96%, recall sebesar 89,71%, dan F1-score sebesar 89,77%. Algoritma ini meningkatkan seluruh metrik evaluasi dibandingkan dengan hasil sebelum penyesuaian hyperparameter. Hal ini menunjukkan bahwa proses optimasi parameter memiliki kemampuan untuk meningkatkan kemampuan Boosting Gradient untuk memperbaiki kesalahan prediksi pada setiap iterasi, yang menghasilkan model yang lebih akurat..

Algoritma *Support Vector Machine* (SVM) memperoleh accuracy sebesar 83,01%, precision sebesar 83,31%, recall sebesar 83,01%, dan F1-score sebesar 82,78%. Nilai akurasi dan presisi meningkat dibandingkan dengan model sebelum penyesuaian hyperparameter; ini menunjukkan bahwa SVM dapat lebih baik dalam membedakan kelas-kelas dalam dataset obesitas dengan memilih kombinasi parameter yang lebih baik. Hasil penelitian menunjukkan bahwa penyesuaian hyperparameter mengurangi kinerja *Random Forest*. Sebelum optimasi, *Random Forest* memperoleh accuracy sebesar 91,63%, sedangkan setelah optimasi nilainya menjadi 90,91%. Penurunan yang relatif kecil ini menunjukkan bahwa parameter bawaan *Random Forest*, atau parameter default, telah mampu bekerja dengan sangat baik pada dataset yang digunakan. Karena efektivitas optimasi sangat bergantung pada karakteristik dataset, ruang pencarian parameter (*search space*), dan konfigurasi parameter yang diuji, proses penyesuaian hyperparameter tidak selalu menghasilkan peningkatan kinerja. Secara keseluruhan, hasil evaluasi setelah *hyperparameter tuning* menunjukkan bahwa proses optimasi memberikan dampak yang berbeda pada setiap algoritma. SVM dan *Gradient Boosting* mengalami peningkatan performa dibandingkan model awal, sedangkan *Random Forest* mengalami sedikit penurunan meskipun tetap menjadi algoritma dengan performa terbaik. Temuan ini menunjukkan bahwa penggunaan *GridSearchCV* efektif dalam mengoptimalkan model tertentu, namun tidak selalu mampu menghasilkan model yang lebih baik dibandingkan parameter bawaan pada semua algoritma. Berdasarkan hasil perbandingan

sebelum dan sesudah *hyperparameter tuning*, *Random Forest* tetap dipilih sebagai model terbaik untuk diimplementasikan pada aplikasi prediksi level obesitas menggunakan Streamlit, karena menghasilkan nilai akurasi, precision, recall, dan F1-score tertinggi dibandingkan algoritma lainnya.

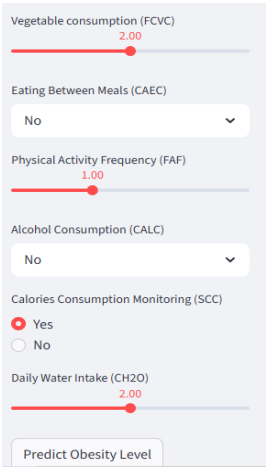
Implementasi dengan Streamlit

Model dengan performa terbaik, yaitu *Random Forest*, selanjutnya diimplementasikan menggunakan Streamlit sebagai aplikasi prediksi level obesitas. Implementasi ini bertujuan untuk memudahkan pengguna dalam melakukan prediksi berdasarkan data yang dimasukkan melalui antarmuka aplikasi.



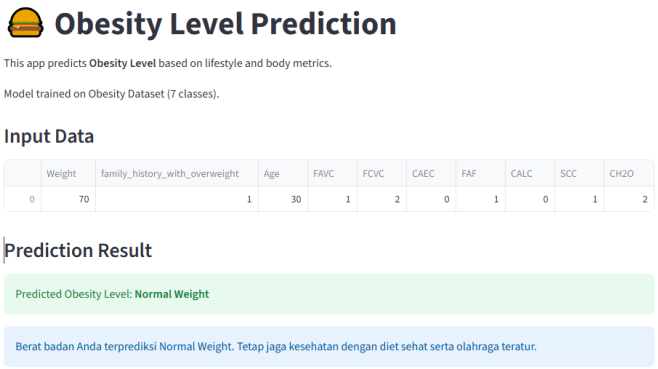
Gambar 1. Input Parameter Oleh Pengguna.

Berdasarkan Gambar 1, aplikasi menyediakan beberapa komponen masukan yang merepresentasikan atribut-atribut yang digunakan dalam proses klasifikasi, seperti Weight, Age, Family History with Overweight, Frequent High Calorie Food (FAVC), Setiap atribut disajikan dalam bentuk komponen antarmuka yang sesuai, seperti radio button, sehingga pengguna dapat mengisi data dengan mudah.



Gambar 2. Input Parameter Oleh Pengguna.

Pada Gambar 2, aplikasi menyediakan beberapa komponen masukan yang merepresentasikan faktor gaya hidup dan kebiasaan pengguna yang berpengaruh terhadap prediksi level obesitas. Atribut yang digunakan meliputi Vegetable Consumption (FCVC), Eating Between Meals (CAEC), Physical Activity Frequency (FAF), Alcohol Consumption (CALC), Calories Consumption Monitoring (SCC), dan Daily Water Intake (CH2O). Setiap atribut disajikan menggunakan komponen antarmuka yang sesuai dengan jenis datanya, seperti slider untuk data numerik, drop-down list untuk data kategorikal, dan radio button untuk data biner.



Gambar 3. Hasil Prediksi Level Obesitas.

Berdasarkan Gambar 3, setelah pengguna mengisi seluruh atribut yang diperlukan dan menekan tombol Predict Obesity Level, aplikasi akan menampilkan data masukan yang telah diproses oleh model dalam bentuk tabel. Penyajian data ini bertujuan untuk memudahkan pengguna melakukan verifikasi terhadap nilai-nilai yang telah diinput sebelum melihat hasil prediksi.

Selanjutnya, aplikasi menampilkan hasil prediksi level obesitas berdasarkan model *Random Forest* yang telah dipilih sebagai model terbaik. Pada contoh pengujian tersebut, aplikasi memprediksi bahwa pengguna berada pada kategori Normal Weight. Selain menampilkan kategori hasil klasifikasi, aplikasi juga memberikan rekomendasi singkat kepada pengguna. Penyampaian informasi tersebut diharapkan dapat membantu pengguna memahami hasil prediksi sekaligus meningkatkan kesadaran terhadap pentingnya menjaga gaya hidup sehat.

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memprediksi level obesitas dengan performa terbaik dibandingkan dengan *Support Vector Machine* (SVM) dan *Gradient Boosting*. Evaluasi model menunjukkan bahwa *Random Forest* memperoleh nilai

akurasi sebesar 90,91%, precision 91,19%, recall 90,91%, dan F1-score 90,94%, sehingga menjadi model terbaik pada penelitian ini. Hasil penelitian juga menunjukkan bahwa penerapan *hyperparameter tuning* dengan *GridSearchCV* berdampak pada masing-masing algoritma secara berbeda. Proses optimasi dapat meningkatkan kinerja beberapa algoritma, tetapi tidak untuk *Random Forest*, yang justru performanya menurun daripada model dengan parameter default. Hasil penelitian ini menunjukkan bahwa pengaturan hyperparameter bergantung pada fitur algoritma dan dataset yang digunakan. Model terbaik, selanjutnya diimplementasikan dengan menggunakan Streamlit sebagai aplikasi prediksi tingkat obesitas. Aplikasi ini dapat digunakan untuk membantu mengidentifikasi tingkat obesitas dengan cepat dan mendukung deteksi dini obesitas.

DAFTAR REFERENSI

Ibrahim, M. Y. I. I., & Atim, S. B. (2024). Klasifikasi Level Obesitas Menggunakan Decision Tree C45 Dalam Menentukan Akurasi Pada KriteriaInformation Gain, Gain Ratio, Gini Index. *JIKI (Jurnal Ilmu Komputer Dan Informatika)*, *5*(2).

Khikam, A., Anggadimas, N. M., Udin, M., Informatika, T., Pasuruan, U. M., & Pasuruan, K. (2025). Implementasi Decson Tree Untuk Klasifikasi Obesitas. *JATI (Jurnal Mahasiiswa Teknik Infrmatika*, *9*(3), 3946–3952.

Muhamad, I., & Matin, M. (2023). *Hyperparameter tuning* menggunakan *GridSearchCV* pada *Random Forest* untuk Deteksi Malware. *JURNAL MULTINETICS*, *9*(1), 43–50.

Nababan, J. M., Arto, I. M., Satria, P., Sumanto, Budiawan, I., & Pakpahan, R. (2025). Penerapan dan Perbandingan Algoritma SVM , Naive Bayes , dan *Gradient Boosting* dalam Prediksi Stroke. *Jurnal Teknik Informatika dan Teknologi Informasi*, *5*(November), 261–272.

Retnoningsih, E., & Pramudita, R. (2020). Mengenal *Machine Learning* Dengan Teknik Supervised dan Unsupervised Learning Menggunakan Python. *BINA INSANI ICT JOURNAL*, *7*(2), 156–165.

Rusman, J., Haryati, B. Z., & Michael, A. (2023). Optimasi Hiperparameter Tuning Pada Metde Supprt Vectr Machine Untuk Klasifikasi Tingkat Kematangan Buah Kopi. *J-Icon : Jurnal Informatika dan Komputer*, *11*(2), 195–202. https://doi.org/10.35508/jicon.v11i2.12571.

Saputra, D., Damayanti, I., & Sultoni, K. (2024). Prediksi BMI Berdasarkan Level Aktivitas Fisik dengan Metode Analisis *Machine Learning*. *Jurnal Pendidikan Kesehatan Rekreasi*, *10*(1), 165–175. https://doi.org/https://doi.org/10.59672/jpkr.v10i1.3499 P-ISSN.

Setiawan, A., Yanti, R., Ali, E., & Yenni, H. (2025). Optimasi Klasifikasi Tingkat Obesitas Pada Remaja Berdasarkan Pola Hidup Menggunakan SVM Dengan Teknik Smote. *Jurnal Algoritma*, *22*(2), 224–235. https://doi.org/10.33364/algoritma/v.22-2.2509.

Sibi, S. Y., Widiarti, A. R., & Dharma, U. S. (2022). Klasifikasi Tingkat Obesitas Mempergunakan Algoritma KNN. *SEMINAR NASIONAL CORISINDO*, 370–375.

Syafarina, G. A., & Zaenuddin. (2023). Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi. *METIK JURNAL*, *7*(2), 121–125. https://doi.org/10.47002/metik.v7i2.680.

Syahputra, A. R., Hidayat, R., Rismansyah, F., Sumanto, Budiawan, I., & Pakpahan, R. (2025). Komparasi Algoritma *Machine Learning* (SVM , *Random Forest* , dan Regresi Logistik) untuk Prediksi Tingkat Obesitas. *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, *5*(November).

Utami, D., Dwiatmoko, F., & Sivi, N. A. (2025). Analisis Pengaruh Bayesian Optimization Terhadap Kinerja SVM Dalam Prediksi Penyakit Diabetes. *Infotek : Jurnal Informatika dan Teknologi*, *8*(1), 140–150.

Valerian, F. R., Syarief, M., Fatah, D. A., Informasi, S., Madura, U. T., Kamal, K., & Timur, J. (2025). Klasifikasi Tingkat Obesitas Menggunakan Metode GBM dan Confusion Matrix. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(2), 2242–2249.

Wie, J. V., & Siddik, M. (2022). Penerapan Metode Naive Bayes Dalam Mengklasifikasi Tingkat Obesitas Pada Pria. *JOISIE Journal Of Information System And Informatics Engineering*, *6*(2), 69–77.

Wijoyo, A., Saputra, A. Y., Ristanti, S., Rafly, S., Ban, S., Komputer, F. I., Informatika, T., Pamulang, U., Raya, J., & No, P. (2024). Pembelajaran *Machine Learning*. *OKTAL: Jurnal Ilmu Komputer dan Science*, *3*(2), 375–380.