Jurnal Ilmiah Teknik Informatika dan Komunikasi
Volume 5. Nomor. 3 November 2025
E-ISSN: 2827-7945 ; P-ISSN: 2827-8127, Hal 660-673
DOI: https://doi.org/10.55606/juitik.v5i3.1716
Tersedia: https://journal.sinov.id/index.php/juitik

Komparasi Algoritma Machine Learning (SVM, Random Forest, dan Regresi Logistik) untuk Prediksi Tingkat Obesitas

Achmad Rivai Syahputra [1*], Rian Hidayat [2], Fathur Rismansyah [3], Sumanto [4], Imam Budiawan [5], Roida Pakpahan [6]
[1-6] Teknik Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika, Indonesia
*Penulis Korespondensi: achmadrivaisyahputra@gmail.com [1]

Abstract. *Obesity is a global health issue with a continuously increasing prevalence. Early prediction of obesity levels is crucial for designing more effective intervention strategies. This study aims to apply and analyze the performance of three machine learning classification methods: Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR), for predicting obesity levels. The research methodology utilizes a public dataset, ObesityLevels, downloaded from the Kaggle platform, which consists of 2111 medical and lifestyle records. The process includes data preprocessing to convert categorical features into numerical ones, splitting the data into training and testing sets with a 70:30 ratio, model training, and evaluation using accuracy, precision, recall, and F1-score metrics. The results indicate that the Random Forest (RF) algorithm achieved the highest performance, with an accuracy of 90.3%, precision of 90.3%, recall of 90.3%, and an F1-score of 90.3%. Based on these findings, it is concluded that the Random Forest model is the most effective choice for an obesity level prediction system based on the dataset used.*

Keywords: *Logistic Regression; Machine Learning; Obesity Classification; Random Forest; SVM.*

Abstrak. Obesitas merupakan masalah kesehatan global yang prevalensinya terus meningkat. Prediksi dini tingkat obesitas dapat membantu dalam merancang strategi intervensi yang lebih efektif. Penelitian ini bertujuan untuk menerapkan dan menganalisis kinerja tiga metode klasifikasi machine learning, yaitu Support Vector Machine (SVM), Random Forest (RF), dan Regresi Logistik (LR), untuk memprediksi tingkat obesitas. Metodologi penelitian menggunakan dataset publik ObesityLevels yang diunduh dari platform Kaggle, yang terdiri dari 2111 data rekam medis dan gaya hidup. Tahapan meliputi preprocessing data untuk mengubah fitur kategorikal menjadi numerik, pembagian data latih dan uji dengan proporsi 70:30, pelatihan model, dan evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa algoritma Random Forest (RF) mencapai kinerja tertinggi dengan nilai akurasi sebesar 90.3%, presisi 90.3%, recall 90.3%, dan F1-score 90.3%. Berdasarkan hasil tersebut, disimpulkan bahwa model Random Forest merupakan pilihan yang paling efektif untuk sistem prediksi tingkat obesitas berdasarkan dataset yang digunakan.

Kata kunci: Klasifikasi Obesitas; Hutan Acak; Pembelajaran Mesin; Regresi Logistik; SVM.

1. LATAR BELAKANG

Obesitas telah menjadi salah satu masalah kesehatan global yang paling mendesak, dengan prevalensi yang terus meningkat dari tahun ke tahun. Kondisi ini tidak lagi menjadi masalah di negara-negara maju saja, tetapi juga menjadi ancaman serius di negara berkembang, termasuk Indonesia. Pola makan yang telah bergeser ke arah konsumsi makanan jadi dan olahan menjadi salah satu pemicu utama meningkatnya angka obesitas di Indonesia (Ramadhani & Djamaluddin, 2024). Masalah ini semakin kompleks karena banyak faktor yang saling terkait, seperti usia, jenis kelamin, dan status sosial, yang berkontribusi terhadap kejadian obesitas di berbagai kelompok usia, termasuk remaja (Suha Ghina Raniya & Rosyada Amrina, 2022). Kondisi ini bukan sekadar masalah berat badan berlebih, tetapi merupakan pintu gerbang menuju berbagai penyakit kronis degeneratif. Obesitas telah terbukti menjadi

Naskah Masuk: 19 September 2025; Revisi: 21 Oktober 2025; Diterima: 08 November 2025; Terbit: 10 November 2025

faktor risiko utama untuk penurunan aktivitas fisik(Pengabdian et al., 2024). Selain itu, obesitas juga menjadi faktor risiko utama untuk diabetes mellitus tipe 2, penyakit kardiovaskular, dan hipertensi (Hidayah et al., 2022). Dampak kesehatan, ekonomi, dan sosial yang ditimbulkannya sangat besar, sehingga menimbulkan beban bagi individu dan sistem kesehatan. Oleh karena itu, diperlukan sebuah sistem prediksi yang dapat membantu mengidentifikasi risiko obesitas secara dini.

Di era digitalisasi, machine learning (ML) telah menjadi alat yang revolusioner dalam bidang kedokteran dan kesehatan masyarakat, khususnya untuk prediksi dini berbagai penyakit. Kemampuan ML untuk mengidentifikasi pola-pola kompleks dan tersembunyi dari data volume besar menjadikannya unggul dibandingkan metode statistik konvensional. Dalam konteks prediksi obesitas, pendekatan ML menjadi sangat relevan karena dapat memanfaatkan berbagai faktor risiko yang telah diidentifikasi dalam berbagai studi. Sebagai contoh, penelitian telah menunjukkan bahwa tingkat aktivitas fisik memiliki korelasi yang signifikan dengan status obesitas (Ar Rafiq et al., 2021). Faktor lain yang berpengaruh adalah pola makan yang tidak seimbang dan konsumsi makanan cepat saji secara berlebihan (Fatmala et al., 2022). Selain itu, penelitian lain juga menunjukkan bahwa konsumsi minuman berpemanis memiliki hubungan kuat dengan peningkatan risiko obesitas pada remaja (Emiliana & Setiarini, 2024). Berdasarkan temuan-temuan tersebut, pendekatan berbasis data menggunakan algoritma pembelajaran mesin berpotensi untuk mengidentifikasi pola obesitas secara lebih akurat.

Beberapa penelitian terdahulu telah menunjukkan hasil yang menjanjikan dalam konteks penerapan ML. Wie dan Siddik (2023) berhasil menerapkan algoritma Naïve Bayes untuk mengklasifikasikan tingkat obesitas pada pria, menunjukkan bahwa pendekatan berbasis ML dapat digunakan untuk mengidentifikasi tingkat obesitas secara efektif (Veron & Muhammad, 2022). Hal ini menegaskan bahwa metode klasifikasi memiliki potensi besar dalam menganalisis data kesehatan dan memberikan wawasan prediktif yang berharga (Kurnia Saraswati et al., 2020).

Meskipun telah banyak penelitian yang berhasil menerapkan berbagai algoritma ML, masih terdapat celah untuk melakukan komparasi langsung dan sistematis terhadap beberapa algoritma klasifikasi fundamental yang paling populer. Algoritma seperti Support Vector Machine (SVM), Random Forest (RF), dan Regresi Logistik (LR) masing-masing memiliki keunggulan dan kelemahan unik. SVM dikenal sangat efektif dalam menangani data berdimensi tinggi (Kurnia Saraswati et al., 2020). Random Forest unggul dalam menangkap hubungan non-linier dan mengurangi overfitting (Kurnia Saraswati et al., 2020). Sementara itu, Regresi Logistik menawarkan interpretabilitas yang tinggi (Ar Rafiq et al., 2021). Namun,

performa optimal dari masing-masing algoritma sangat bergantung pada karakteristik spesifik dataset yang digunakan. Oleh karena itu, penelitian ini bertujuan untuk (1) menerapkan metode klasifikasi SVM, Random Forest, dan Regresi Logistik pada dataset obesitas, dan (2) menganalisis serta membandingkan kinerja ketiga metode tersebut untuk menemukan model prediksi terbaik yang paling efektif untuk kasus ini.

Penelitian ini secara spesifik menggunakan dataset publik ObesityDataSet yang berisi data gaya hidup dan kondisi fisik individu dari beberapa negara Amerika Latin(Kurnia Saraswati et al., 2020). Dataset ini memuat berbagai variabel seperti kebiasaan makan, aktivitas fisik, serta faktor sosial dan demografis yang relevan untuk prediksi obesitas. Dengan membandingkan ketiga model fundamental ini secara head-to-head menggunakan metrik evaluasi yang sama, diharapkan dapat memberikan rekomendasi yang jelas dan berbasis bukti bagi praktisi kesehatan atau peneliti lain dalam mengembangkan sistem prediksi risiko obesitas yang lebih akurat dan andal. Hasil dari penelitian ini diharapkan dapat berkontribusi pada upaya pencegahan obesitas di tingkat individu dan populasi melalui pendekatan berbasis data yang lebih presisi.

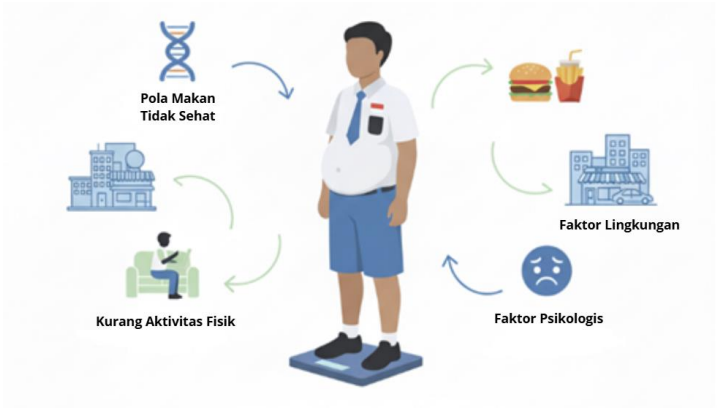
2. KAJIAN TEORITIS

Obesitas

Obesitas didefinisikan sebagai akumulasi lemak abnormal atau berlebihan yang dapat mengganggu kesehatan (Hanum, 2023). Faktor-faktor penyebab obesitas sangat kompleks dan meliputi faktor genetik, perilaku, dan lingkungan (Sumarni & Bangkele, 2023). Penelitian terkini menunjukkan bahwa pergeseran pola konsumsi ke arah makanan jadi dan olahan telah menjadi pemicu utama meningkatnya angka obesitas di Indonesia (Fatmala et al., 2022). Faktor risiko lain yang signifikan adalah konsumsi minuman berpemanis yang tinggi, yang terbukti berkontribusi terhadap kejadian obesitas terutama pada kelompok anak dan remaja (Fatmala et al., 2022).

Berbagai faktor lain juga berperan, seperti yang diidentifikasi dalam studi pada populasi remaja, mencakup pola makan, aktivitas fisik, dan faktor psikososial (Banjarnahor et al., 2022). Hubungan antara obesitas dan aktivitas fisik bersifat timbal balik; obesitas dapat menjadi faktor risiko penurunan aktivitas fisik, yang pada gilirannya memperburuk kondisi obesitas itu sendiri (Banjarnahor et al., 2022). Analisis data nasional seperti Riskesdas 2018 bahkan menunjukkan bahwa faktor seperti usia dan jenis kelamin memiliki hubungan signifikan dengan kejadian obesitas pada remaja di Indonesia (Putri et al., 2022).

Dalam dataset yang digunakan dalam penelitian ini, tingkat obesitas diklasifikasikan ke dalam tujuh kelas berdasarkan Indeks Massa Tubuh (IMT). Beberapa faktor risiko yang diamati dalam dataset seperti frekuensi aktivitas fisik (FAF), waktu penggunaan perangkat elektronik (TUE), dan riwayat obesitas dalam keluarga (family_history_with_overweight) mencerminkan kompleksitas faktor-faktor penyebab yang diidentifikasi dalam berbagai studi (Fauzan et al., 2023)(Zahari et al., 2022).

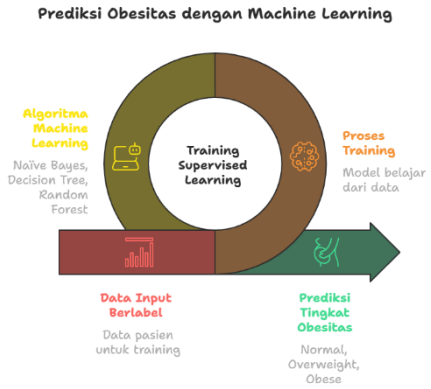


Gambar 1. Ilustrasi Obesitas.

Machine Learning

Machine Learning (ML) adalah bidang ilmu komputer yang memungkinkan sistem komputer untuk belajar dari data tanpa harus diprogram secara eksplisit (Ariyanto et al., 2023). Salah satu tugas utama dalam machine learning adalah supervised classification, di mana model dilatih menggunakan data berlabel untuk memprediksi kelas dari data baru yang tidak berlabel (Ariyanto et al., 2023). Dalam konteks prediksi penyakit, termasuk obesitas, ML telah menjadi alat yang revolusioner.

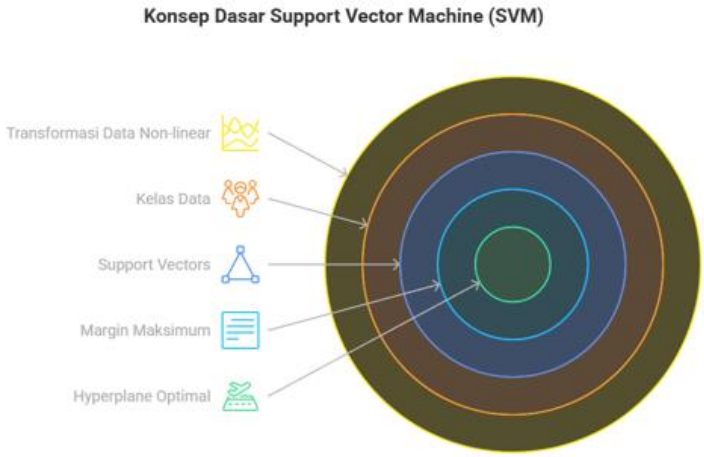
Penelitian terdahulu telah menunjukkan penerapan berbagai algoritma klasifikasi untuk memprediksi tingkat obesitas, seperti penggunaan algoritma Naïve Bayes yang berhasil mengklasifikasikan tingkat obesitas pada pria (Veron & Muhammad, 2022). Dalam penelitian ini, model dilatih untuk memetakan sekumpulan fitur input (misalnya usia, jenis kelamin, kebiasaan makan) ke output berupa kelas tingkat obesitas (Veron & Muhammad, 2022).



Gambar 2. Konsep Machine Learning.

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran yang kuat untuk masalah klasifikasi dan regresi (Mailoa, 2021). Inti dari SVM adalah menemukan hyperplane optimal yang memisahkan data dari kelas yang berbeda dalam ruang fitur. SVM bekerja dengan memaksimalkan margin, yaitu jarak terdekat antara hyperplane dan titik data dari setiap kelas (disebut support vectors) (Mailoa, 2021). Untuk data yang tidak dapat dipisahkan secara linier, SVM dapat menggunakan fungsi kernel seperti RBF atau polinomial (Mailoa, 2021). Algoritma ini merupakan salah satu metode klasifikasi fundamental yang sering digunakan dalam studi prediksi kesehatan (Banjarnahor et al., 2022).



Gambar 3. Ilustrasi SVM.

Random Forest

Random Forest adalah algoritma ensemble learning yang terdiri dari banyak decision trees (Arifani & Setiyaningrum, 2021). Algoritma ini bekerja dengan membangun beberapa pohon keputusan pada berbagai sub-sampel dari dataset dan menggunakan rata-rata (untuk regresi) atau voting (untuk klasifikasi) untuk meningkatkan akurasi prediksi dan mengontrol overfitting (Arifani & Setiyaningrum, 2021). Metode bagging (bootstrap aggregating)

digunakan untuk membuat setiap pohon keputusan menggunakan sampel data acak dengan pengembalian. Sifatnya yang robust dan mampu menangani hubungan non-linier membuat Random Forest menjadi pilihan populer dalam berbagai penelitian klasifikasi, termasuk di bidang kesehatan (Arifani & Setiyaningrum, 2021).

Gambar 4. Ilustrasi Random Forest.

Regresi Logistik

Regresi Logistik adalah algoritma klasifikasi statistik yang digunakan untuk memprediksi probabilitas dari sebuah kelas kategorikal (Ariyanto et al., 2023). Dalam kasus klasifikasi multi-kelas seperti prediksi obesitas, digunakan pendekatan one-vs-rest atau multinomial logistic regression(Veron & Muhammad, 2022). Model ini menggunakan fungsi sigmoid untuk mengubah output linier menjadi nilai probabilitas antara 0 dan 1 (Hita, 2020).

Meskipun namanya mengandung "regresi", ini adalah algoritma klasifikasi yang sederhana, cepat, dan mudah diinterpretasikan, yang sering dijadikan sebagai baseline dalam perbandingan model prediksi (Hita, 2020). Selain itu, regresi logistik menawarkan keunggulan dalam hal interpretabilitas koefisien, yang memungkinkan peneliti memahami pengaruh masing-masing variabel prediktor terhadap probabilitas terjadinya kelas tertentu. Kemampuan ini menjadikan regresi logistik sangat berguna dalam penelitian kesehatan masyarakat, termasuk analisis faktor risiko obesitas, karena hasil model dapat dijadikan dasar pengambilan kebijakan atau intervensi Kesehatan (Dikko et al., 24 C.E.)

Gambar 5. Ilustrasi Regresi Logistik.

Metrik Evaluasi Kinerja

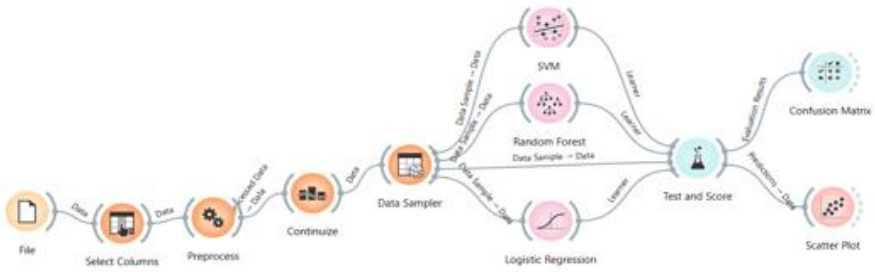
Untuk mengevaluasi performa model klasifikasi, digunakan beberapa metrik yang dihitung berdasarkan Confusion Matrix (Ariyanto et al., 2023). Komponennya adalah True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). Dari komponen ini, dapat dihitung:

Gambar 6. Metrik Evaluasi.

Metrik-metrik ini adalah standar dalam menilai kinerja model dan selalu dilaporkan dalam penelitian penerapan machine learning untuk prediksi penyakit (Veron & Muhammad, 2022).

3. METODE PENELITIAN

Penelitian ini menggunakan perangkat lunak Orange untuk membangun dan mengevaluasi model prediksi tingkat obesitas. Workflow analisis dirancang untuk mengolah data mentah.

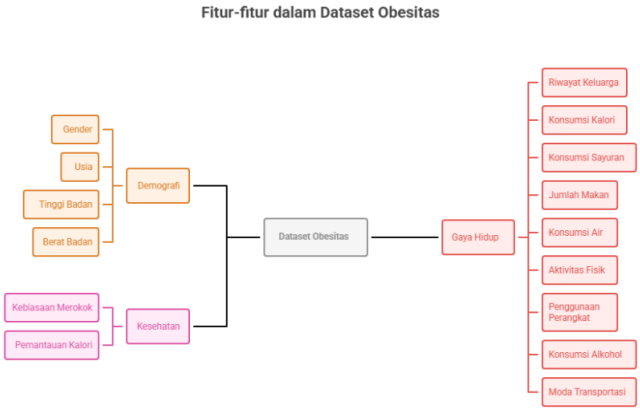


Gambar 7. Workflow Penelitian untuk Prediksi Tingkat Obesitas.

Menjadi model klasifikasi yang dapat diuji performansinya. Proses ini mencakup pengambilan data, praproses, pembagian data, pemodelan, dan evaluasi. Gambar 1 menunjukkan workflow penelitian yang lengkap, menggambarkan alur data dari sumbernya hingga tahap visualisasi hasil evaluasi.

Sumber Data

Penelitian ini menggunakan dataset publik yang diperoleh dari platform Kaggle dengan nama ObesityDataSet_raw_and_data_sinthetic.csv. Dataset ini berisi data demografis, kebiasaan makan, dan kondisi fisik individu dengan rentang usia 14 hingga 61 tahun. Dataset terdiri dari sekitar 2111 rekam data dengan 17 fitur (atribut) dan satu variabel target. Variabel target dalam penelitian ini adalah NObeyesdad, yang mengklasifikasikan tingkat obesitas menjadi tujuh kategori, yaitu Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II, dan Obesity Type III(Fatemeh Mehrparvar, 2024). Beberapa fitur penting dalam dataset antara lain:



Gambar 8. Fitur Penting Dalam Data.

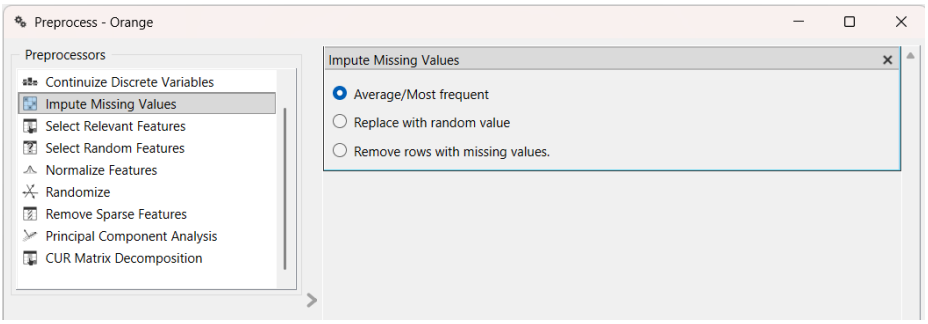
Preprocessing Data

Tahap praproses data bertujuan untuk membersihkan dan menyiapkan data agar dapat digunakan oleh algoritma machine learning. Dua node utama yang digunakan dalam tahap ini adalah Select Columns dan Continuize.

Select Columns: Pada tahap ini, dilakukan pemilihan fitur yang akan digunakan dalam pemodelan. Berdasarkan analisis, semua 17 fitur pada dataset dianggap relevan untuk memprediksi tingkat obesitas. Oleh karena itu, tidak ada fitur yang dibuang pada node Select Columns. Semua fitur digunakan dengan variabel NObeyesdad ditetapkan sebagai target.

Continuize: Langkah ini krusial untuk mengubah fitur-fitur kategorikal menjadi bentuk numerik. "Fitur-fitur kategorikal seperti Gender, CALC, CAEC, dan MTRANS tidak dapat langsung diproses oleh algoritma SVM dan Regresi Logistik. Oleh karena itu, fitur-fitur ini

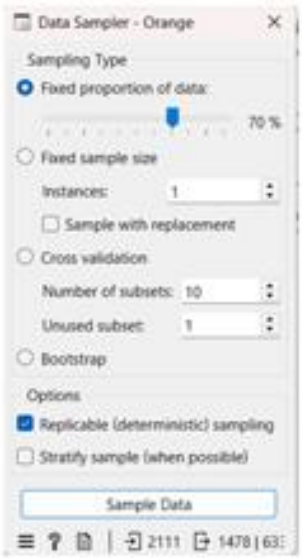
diubah menjadi representasi numerik menggunakan metode one-hot encoding melalui node Continuize." Proses ini mengubah setiap kategori dalam fitur kategorikal menjadi kolom biner baru (0 atau 1), yang memungkinkan algoritma untuk melakukan perhitungan matematis. Sementara itu, fitur numerik seperti Age, Height, dan Weight dibiarkan tetap seperti adanya.



Gambar 9. Tahap Preprocessing Data.

Pembagian Data

Setelah data diproses, langkah selanjutnya adalah membagi dataset menjadi data latih (training set) dan data uji (testing set). Pembagian ini dilakukan menggunakan node Data Sampler. "Dataset dibagi menjadi dua bagian: data latih (training set) sebesar 70% dan data uji (testing set) sebesar 30% menggunakan stratified sampling untuk menjaga proporsi kelas." Metode stratified sampling memastikan bahwa distribusi kelas pada variabel target (NObeyesdad) di data latih dan data uji seimbang, sehingga evaluasi model menjadi lebih valid dan tidak bias terhadap kelas minoritas.



Gambar 10. Pembagian Data Sampler.

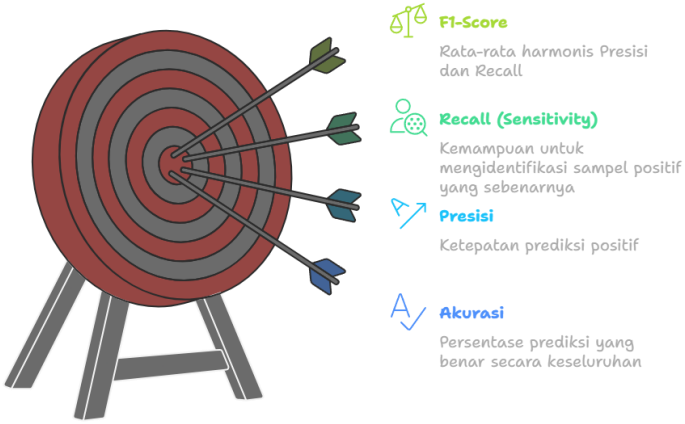
Pemodelan Klasifikasi

Pada tahap ini, dibangun tiga model klasifikasi yang berbeda untuk memprediksi tingkat obesitas. Ketiga model dilatih menggunakan data latih yang telah dihasilkan dari node Data Sampler. Model-model tersebut adalah:

Gambar 11. Pemodelan Klasifikasi.

Evaluasi Model

Evaluasi model merupakan tahap penting untuk mengukur seberapa baik performa ketiga model dalam memprediksi data yang belum pernah dilihat sebelumnya. Evaluasi dilakukan menggunakan node Test and Score. Pada node ini, ketiga model yang telah dilatih diuji menggunakan data uji yang sama (30% dari dataset) untuk memastikan perbandingan yang adil.



Gambar 12. Evaluasi Model.

Hasil evaluasi dari node Test and Score kemudian dianalisis lebih lanjut menggunakan node Confusion Matrix. Matriks ini memberikan visualisasi yang detail tentang performa model dengan menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas. Hal ini memungkinkan identifikasi jenis kesalahan yang paling sering dibuat oleh model, seperti apakah model cenderung salah mengklasifikasikan kelas tertentu ke kelas lainnya.

4. HASIL DAN PEMBAHASAN

Hasil Evaluasi Kinerja Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan metode *Cross Validation* dengan 10 *folds* serta pendekatan *Stratified* untuk memastikan distribusi kelas yang seimbang pada setiap *fold*. Metrik evaluasi yang digunakan meliputi AUC (*Area Under the Curve*), Akurasi (CA – *Classification Accuracy*), Presisi (*Precision*), Recall, F1-Score, dan MCC (*Matthews Correlation Coefficient*).Hasil perbandingan kinerja ketiga model disajikan pada tabel berikut:

Tabel 1. Evaluasi Kinerja Model.

Model	AUC	CA	F1	Prec	Recall	MCC
SVM	0.990	0.894	0.895	0.896	0.894	0.877
Random Forest	0.995	0.931	0.932	0.935	0.931	0.920
Logistic Regression	0.982	0.867	0.865	0.866	0.867	0.845

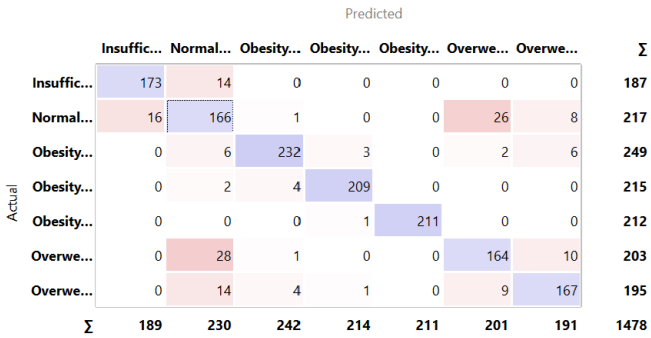
Berdasarkan Tabel 1, model Random Forest menunjukkan performa terbaik secara konsisten pada seluruh metrik evaluasi. Model ini memiliki nilai akurasi tertinggi sebesar 0.931, diikuti oleh SVM (0.894) dan Regresi Logistik (0.867).

Dari sisi F1-Score, yang merupakan gabungan antara precision dan recall, Random Forest juga unggul dengan nilai 0.932, menandakan keseimbangan yang sangat baik antara kemampuan model mengenali kelas positif dan meminimalkan kesalahan prediksi.

Selain itu, AUC sebesar 0.995 menunjukkan bahwa Random Forest memiliki kemampuan yang hampir sempurna dalam membedakan antara kelas positif dan negatif. Nilai MCC sebesar 0.920 memperkuat hasil tersebut, menandakan bahwa model memiliki korelasi prediksi yang sangat kuat terhadap kelas aktual, bahkan pada dataset yang berpotensi tidak seimbang.

Sementara itu, SVM menempati posisi kedua dengan hasil yang cukup kompetitif. Nilai AUC sebesar 0.990 dan MCC sebesar 0.877 menunjukkan bahwa model ini juga mampu melakukan klasifikasi dengan baik, meskipun sedikit lebih rendah dibandingkan Random Forest. Adapun Regresi Logistik menghasilkan performa paling rendah di antara ketiganya dengan AUC 0.982 dan MCC 0.845, yang menunjukkan bahwa asumsi linearitas model ini belum sepenuhnya cocok untuk pola data yang kompleks.

Analisis Confusion Matrix



Gambar 13. Hasil Confusion Matrix.

Berdasarkan Gambar 13, kesalahan klasifikasi yang paling sering terjadi pada model Random Forest adalah prediksi kelas Overweight_Level_I sebagai Normal_Weight, dengan jumlah sekitar 23 sampel. Kesalahan ini menunjukkan adanya kemiripan karakteristik fitur antara kedua kelas, seperti berat dan tinggi badan yang berada pada rentang batas di antara keduanya.

Kesalahan tertinggi kedua terjadi pada prediksi Obesity_Type_I yang diklasifikasikan sebagai Overweight_Level_II sebanyak 18 sampel. Pola ini menandakan adanya overlap antar kategori berat badan yang berurutan, yang wajar terjadi karena transisi antara level berat badan tersebut bersifat kontinu. Selain itu, distribusi kelas yang tidak seimbang (misalnya jumlah sampel Normal_Weight dan Overweight_Level_I yang lebih dominan) turut memengaruhi akurasi klasifikasi pada beberapa kategori minoritas.

Pembahasan Komprehensif

Secara keseluruhan, hasil penelitian menunjukkan bahwa Random Forest adalah model dengan performa paling unggul dibandingkan SVM dan Regresi Logistik. Keunggulan ini disebabkan oleh kemampuan Random Forest sebagai algoritma ensemble yang menggabungkan hasil dari banyak pohon keputusan untuk mengurangi overfitting dan menangani hubungan non-linear antar fitur seperti Age, Height, Weight, serta variabel gaya hidup (FAF, CH2O, CALC, dan MTRANS).

Sebaliknya, SVM sangat bergantung pada pemilihan parameter dan jenis kernel yang digunakan (dalam hal ini RBF Kernel), sedangkan Regresi Logistik mengasumsikan hubungan linear antar variabel, yang mungkin tidak mampu menangkap kompleksitas data obesitas secara menyeluruh.

Hasil ini juga sejalan dengan penelitian Palechor & de la Hoz (Palechor & de la Hoz Manotas, 2019) yang melaporkan bahwa algoritma berbasis pohon keputusan seperti Random Forest mampu mencapai akurasi lebih dari 90% pada klasifikasi data obesitas. Dengan

demikian, hasil penelitian ini memperkuat bukti empiris bahwa Random Forest merupakan model yang paling sesuai dan andal untuk prediksi tingkat obesitas, terutama pada dataset dengan kombinasi variabel numerik dan kategorikal.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menunjukkan keunggulan algoritma Random Forest, keterbatasan utamanya terletak pada sifat data yang statis dan tidak mampu menangkap evolusi perilaku individu seiring waktu. Dengan keyakinan, arah penelitian lanjutan yang paling menjanjikan dan jelas adalah beralih sepenuhnya pada penerapan arsitektur Long Short-Term Memory (LSTM). Metode ini secara spesifik dirancang untuk menganalisis data *time-series*, yang akan memungkinkan integrasi data longitudinal dari *wearable devices* seperti tingkat aktivitas fisik, pola tidur, dan variabilitas denyut jantung. Dengan memanfaatkan LSTM, penelitian selanjutnya tidak hanya akan meningkatkan akurasi, tetapi secara fundamental akan mengubah model prediksi dari yang bersifat statis menjadi sebuah sistem yang dinamis dan personal, yang mampu memberikan peringatan dini dan intervensi yang proaktif dalam manajemen obesitas.

DAFTAR REFERENSI

Ar Rafiq, A., Sutono, & Lukman Wicaksana, A. (2021). Pengaruh aktivitas fisik terhadap penurunan berat badan dan tingkat kolesterol pada orang dengan obesitas: Literature review. *Jurnal Keperawatan Klinis Dan Komunitas, 5*(3), 167-178. https://doi.org/10.22146/jkkk.60362

Arifani, S., & Setiyaningrum, Z. (2021). Faktor perilaku berisiko yang berhubungan dengan kejadian obesitas pada usia dewasa di Provinsi Banten tahun 2018. *Jurnal Kesehatan, 14*(2), 160-168. https://doi.org/10.23917/jk.v14i2.13738

Ariyanto, Fatmawati, T. Y., Chandra, F., & Efni, N. (2023). Perilaku mahasiswa dalam pencegahan obesitas di STIKes Baiturrahim Jambi. *Jurnal Akademika Baiturrahim Jambi (JABJ), 12*(1), 201-206. https://doi.org/10.36565/jab.v12i1.696

Banjarnahor, R. O., Banurea, F. F., Panjaitan, J. O., Pasaribu, R. S. P., & Hafni, I. (2022). Faktor-faktor risiko penyebab kelebihan berat badan dan obesitas pada anak dan remaja: Studi literatur. *Tropical Public Health Journal, 2*(1), 35-45. https://doi.org/10.32734/trophico.v2i1.8657

Dikko, M. U., Hussaini, U., Alkali, Z. A., Bandiya, M. A. M., & Abdullahi, M. (24 C.E.). The moderating effect of corporate governance in the relationship women owned enterprises: A proposed conceptual framework. *Fudma Journal of Management Sciences, 6*(2), 167-186.

Emiliana, N., & Setiarini, A. (2024). Hubungan konsumsi minuman berpemanis dengan kejadian obesitas pada anak dan remaja: A systematic literature review. https://repository.unar.ac.id/jspui/handle/123456789/10983

Fatemeh Mehrparvar. (2024). Obesity levels. In *Kaggle*. Kaggle. https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels/data

Fatmala, T., Rohmah, M., Maulidia Septimar, Z., & Tangerang, S. Y. (2022). The relationship of sweet beverage consumption with obesity in adolescents. *Nusantara Hasana Journal*, *2*(1), 1-Page.

Fauzan, M. R., Sarman, Rumaf, F., Darmin, Tutu, C. G., & Alkhair. (2023). Upaya pencegahan obesitas pada remaja menggunakan media komunikasi. *Jurnal Pengabdian Kepada Masyarakat MAPALUS*, *1*(2), 29-34. https://e-journal.stikesgunungmaria.ac.id/index.php/jpmm/article/view/39

Hanum, A. M. (2023). Faktor-faktor penyebab terjadinya obesitas pada remaja. *Healthy Tadulako Journal (Jurnal Kesehatan Tadulako)*, *9*(2), 137-147. https://doi.org/10.22487/htj.v9i2.539

Hidayah, N. A., Stikes, K., Cipta, B., & Purwokerto, H. (2022). Khasanah, N. A. H. (2022). Hubungan usia, jenis kelamin dan status obesitas dengan kejadian hipertensi di wilayah Puskesmas Sumbang II Kabupaten Banyumas. *Jurnal Bina Cipta Husada*, *18*(1), 43-55.

Hita, I. P. A. D. (2020). Efektivitas metode latihan aerobik dan anaerobik untuk menurunkan tingkat overweight dan obesitas. *Jurnal Penjakora*, *7*(2), 135. https://doi.org/10.23887/penjakora.v7i2.27375

Kurnia Saraswati, S., Dhista Rahmaningrum, F., Naufal Zidane Pahsya, M., Paramitha, N., Wulansari, A., Rossa Ristantya, A., Magdalena Sinabutar, B., Estetika Pakpahan, V., & Nandini, N. (2020). Media kesehatan masyarakat Indonesia literature review: Faktor risiko penyebab obesitas. *Media Kesehatan Masyarakat Indonesia*, *70-74*. https://ejournal.undip.ac.id/index.php/mkmi https://doi.org/10.14710/mkmi.20.1.70-74

Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di Indonesia. *Journal of Information Systems for Public Health*, *6*(1), 44. https://doi.org/10.22146/jisph.44455

Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data in Brief*, *25*, 104344. https://doi.org/10.1016/j.dib.2019.104344

Pengabdian, J., Jupe, M., Di, O., & Kendal, K. (2024). Program pengabdian masyarakat untuk mengatasi. *1*, 21-23.

Putri, R. N., Nugraheni, S. A., & Pradigdo, S. F. (2022). Faktor-faktor yang berhubungan dengan kejadian obesitas sentral pada remaja usia 15-18 tahun di Provinsi DKI Jakarta (Analisis Riskesdas 2018). *Media Kesehatan Masyarakat Indonesia*, *21*(3), 169-177. https://doi.org/10.14710/mkmi.21.3.169-177

Ramadhani, I. C., & Djamaluddin, S. (2024). Pengaruh konsumsi jadi dan olahan terhadap obesitas di Indonesia. *Jurnal Informatika Ekonomi Bisnis*. https://www.infeb.org/index.php/infeb/article/view/974

Suha Ghina Raniya, & Rosyada Amrina. (2022). Faktor-faktor yang berhubungan dengan kejadian obesitas pada remaja umur 13-15 tahun di Indonesia (analisis lanjut data Riskesdas 2018). *Ilmu Gizi Indonesia*, *06*(01), 43-56. https://doi.org/10.35842/ilgi.v6i1.339

Sumarni, S., & Bangkele, E. Y. (2023). Persepsi orang tua, guru dan tenaga kesehatan tentang obesitas pada anak dan remaja. *Healthy Tadulako Journal (Jurnal Kesehatan Tadulako), 9*(1), 58-64. https://doi.org/10.22487/htj.v9i1.658

Veron, W. J., & Muhammad, S. (2022). Penerapan metode naive bayes dalam mengklasifikasi tingkat obesitas pada pria. *JOISIE Journal Of Information System And Informatics Engineering*, *6*(2), 69-77.

Zahari, Q. F., Prashanti, N. A. S., Salsabella, S., Jumiatmoko, J., Hafidah, R., & Nurjannah, N. E. (2022). Kemampuan fisik motorik anak usia dini dengan masalah obesitas. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, *6*(4), 2844-2851. https://doi.org/10.31004/obsesi.v6i4.1570