JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 5, NO. 4 (2026)
DOI: https://doi.org/10.55537/cosie.v5i4.1855

e-ISSN 2827-8356
p-ISSN 2827-9999

Klasifikasi Tingkat Obesitas Menggunakan Gaussian Naïve Bayes dan *10-Fold Cross Validation*

Obesity Level Classification Using the Gaussian Naive Bayes and 10-Fold Cross-Validation

Rivaldi Prima Nanda[*1], Muhammad Siddik Hasibuan[2], Mhd. Ikhsan Rifki[3]
[1,2,3]Universitas Islam Negeri Sumatera Utara Medan
[1]rivaldiprimanandaxtkj2@gmail.com, [2]muhammadsiddik@uinsu.ac.id, [3]rifki.mhdikhsan@uinsu.ac.id

Received: August 1, 2026 | Revised August 4, 2026 | Accepted: August 29, 2026 | Published: September 13, 2026

Abstrak

Klasifikasi obesitas merupakan tugas penting dalam analisis data kesehatan untuk mengidentifikasi tingkat obesitas berdasarkan karakteristik individu. Penelitian ini bertujuan untuk mengimplementasikan algoritma Gaussian Naïve Bayes guna mengklasifikasikan tingkat obesitas serta mengevaluasi perbedaan estimasi kinerja yang diperoleh melalui metode pembagian data latih-uji (*train-test split*) 70:30 dan 10-Fold Cross Validation. Dataset yang digunakan terdiri dari 1.000 catatan observasi yang memuat lima variabel prediktor: usia, tinggi badan, berat badan, indeks massa tubuh (IMT), dan tingkat aktivitas fisik. Tingkat obesitas dikategorikan ke dalam empat kelas: berat badan kurang, Normal, kelebihan berat badan, dan obesitas. Model diimplementasikan menggunakan Python dan dievaluasi melalui dua skenario validasi. Hasil eksperimen menunjukkan bahwa model Gaussian Naïve Bayes mencapai akurasi sebesar 95,33% dengan pendekatan pembagian data latih-uji. Evaluasi menggunakan 10-Fold Cross Validation menghasilkan akurasi rata-rata 96,30%, presisi 96,66%, recall 96,45%, dan skor F1 sebesar 96,48%. Temuan ini mengindikasikan bahwa strategi validasi yang berbeda dapat menghasilkan estimasi kinerja yang berbeda pula. Pada dataset yang digunakan dalam penelitian ini, Gaussian Naïve Bayes menunjukkan kinerja klasifikasi yang tinggi, sementara 10-Fold Cross Validation memberikan estimasi kinerja di berbagai partisi data, sehingga menawarkan penilaian kinerja model yang lebih komprehensif. Penelitian ini memberikan bukti empiris mengenai penggunaan strategi validasi yang berbeda untuk klasifikasi tingkat obesitas.

Kata Kunci: Klasifikasi Obesitas; Gaussian Naïve Bayes; Machine Learning; K-Fold Cross Validation; Data Mining.

Abstract

Obesity classification is an important task in health-related data analysis for identifying obesity levels based on individual characteristics. This study aims to implement the Gaussian Naïve Bayes algorithm for obesity level classification and evaluate differences in performance estimates obtained using a 70:30 train-test split and 10-Fold Cross Validation. The dataset consisted of 1,000 observational records containing five predictor variables: age, height, weight, body mass index (BMI), and physical activity level. Obesity levels were categorized into four classes: Underweight, Normal, Overweight, and Obese. The model was implemented using Python and evaluated through two validation scenarios. Experimental results showed that the Gaussian Naïve Bayes model achieved an accuracy of 95.33% using the train-test split approach. Evaluation using 10-Fold Cross Validation produced an average accuracy of 96.30%, precision of 96.66%, recall of 96.45%, and F1-score of 96.48%. These findings indicate that different validation strategies can produce different performance estimates. On the dataset used in this study, Gaussian Naïve Bayes demonstrated high classification performance, while 10-Fold Cross Validation provided performance estimates across multiple data partitions, offering a broader assessment of model performance. This study provides empirical evidence regarding the use of different validation strategies for obesity level classification.

Keywords: Obesity Classification; Gaussian Naïve Bayes; Machine Learning; K-Fold Cross Validation; Data Mining

DOI: https://doi.org/10.55537/cosie.v5i4.1855

1. PENDAHULUAN

Peningkatan prevalensi obesitas dalam beberapa tahun terakhir menjadikan kondisi ini sebagai salah satu tantangan utama kesehatan masyarakat karena berkaitan dengan berbagai penyakit kronis. Pada tahun 2022 terdapat sekitar 2,5 miliar orang dewasa mengalami kelebihan berat badan dan lebih dari 890 juta di antaranya tergolong obesitas. Kondisi ini berkontribusi terhadap meningkatnya risiko diabetes melitus, penyakit kardiovaskular, hipertensi, gangguan metabolik, serta penurunan kualitas hidup masyarakat [1]. Selain faktor genetik, obesitas dipengaruhi oleh berbagai faktor seperti usia, tinggi badan, berat badan, indeks massa tubuh (IMT), pola makan, dan aktivitas fisik [2] [3] [4]. Oleh karena itu, identifikasi tingkat obesitas secara cepat dan akurat menjadi langkah penting dalam mendukung upaya pencegahan serta pengambilan keputusan di bidang kesehatan. Kemajuan teknologi informasi telah mendorong pemanfaatan *machine learning* secara luas dalam berbagai bidang, termasuk pengolahan dan analisis data kesehatan. Sebagai salah satu cabang kecerdasan buatan, *machine learning* memungkinkan sistem untuk mempelajari pola dan hubungan yang terdapat dalam data sehingga dapat menghasilkan prediksi maupun klasifikasi secara otomatis [5], [6]. Dalam bidang obesitas, berbagai pendekatan *machine learning* telah digunakan untuk memprediksi maupun mengklasifikasikan tingkat obesitas berdasarkan karakteristik fisik, perilaku, dan gaya hidup individu. Berbagai penelitian menunjukkan bahwa pendekatan tersebut mampu memberikan performa klasifikasi yang baik pada data kesehatan, sehingga semakin banyak dimanfaatkan sebagai alat bantu analisis data kesehatan berbasis komputasi. Berdasarkan metode pembelajarannya, *machine learning* umumnya dikelompokkan menjadi dua pendekatan utama, yaitu *supervised learning* dan *unsupervised learning*. Pada penelitian yang berfokus pada klasifikasi tingkat obesitas, pendekatan supervised learning dinilai lebih tepat karena proses pembelajaran model dilakukan menggunakan data yang telah memiliki label kelas sebagai acuan dalam membangun pola klasifikas [7] [8], [9]

Salah satu teknik *supervised learning* yang sering digunakan dalam *data mining* yaitu algoritma Naïve Bayes. Algoritma ini termasuk metode klasifikasi probabilistik yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antaratribut. Keunggulan Naïve Bayes terletak pada proses komputasi yang sederhana, waktu pelatihan yang cepat, serta kemampuannya menghasilkan performa yang baik pada berbagai kasus klasifikasi [10][11]. Untuk data numerik kontinu, digunakan varian Gaussian Naïve Bayes yang memodelkan distribusi data menggunakan distribusi normal sehingga sesuai diterapkan pada atribut kesehatan seperti usia, berat badan, tinggi badan, dan indeks massa tubuh. Penelitian sebelumnya menunjukkan bahwa Gaussian Naïve Bayes mampu menghasilkan performa klasifikasi yang baik pada data kesehatan. Penelitian oleh [12] menerapkan Gaussian Naïve Bayes pada klasifikasi diagnosis penyakit jantung dan memperoleh tingkat akurasi yang tinggi. Hasil serupa juga dilaporkan oleh [13] pada klasifikasi pasien gagal jantung untuk klasifikasi status gizi balita menggunakan Naïve Bayes dengan teknik *K-Fold Cross Validation*. Selain pada bidang kesehatan, algoritma Naïve Bayes juga menunjukkan performa kompetitif pada analisis sentimen dan klasifikasi teks dibandingkan beberapa algoritma klasifikasi lainnya [14] [15] [16]. Pada kasus klasifikasi obesitas [17], beberapa penelitian telah membandingkan berbagai algoritma machine learning untuk memperoleh model terbaik. [15] membandingkan K-Nearest Neighbor, Naïve Bayes, dan Random Forest dalam prediksi obesitas dan menunjukkan bahwa metode *machine learning* mampu menghasilkan tingkat klasifikasi yang tinggi. Penelitian oleh [18] mengevaluasi beberapa algoritma klasifikasi tingkat obesitas menggunakan *K-Fold Cross Validation* dan *Area Under Curve* (AUC), sedangkan penelitian oleh [19] menekankan pentingnya metode evaluasi yang tepat untuk memperoleh model prediksi yang lebih andal. Temuan tersebut meberitahu bahwa suatu keberhasilan model klasifikasi tidak hanya dipengaruhi oleh pemilihan algoritma, tetapi juga oleh metode validasi yang digunakan.

Meskipun berbagai penelitian telah berhasil menerapkan *machine learning* pada klasifikasi kesehatan, sebagian besar penelitian masih menggunakan pendekatan *train-test split* tunggal dalam proses evaluasi model [20]. Pendekatan ini menghasilkan estimasi performa berdasarkan satu pembagian data tertentu sehingga hasil evaluasi yang diperoleh dapat berbeda apabila digunakan partisi data yang berbeda. Oleh karena itu, diperlukan strategi evaluasi yang dapat memberikan gambaran performa model pada beberapa pembagian data, terutama ketika ukuran dataset relatif terbatas [14] [21]. Salah satu strategi evaluasi yang banyak digunakan untuk memperoleh estimasi performa model pada berbagai partisi data adalah *K-Fold Cross Validation.* Teknik ini membagi dataset menjadi beberapa subset dan melakukan proses pelatihan serta pengujian secara berulang sehingga setiap data memiliki kesempatan yang sama untuk menjadi data uji maupun data latih [22]. Beberapa penelitian menunjukkan *bahwa K-Fold Cross Validation* memungkinkan evaluasi model dilakukan pada beberapa pembagian data yang berbeda sehingga menghasilkan estimasi performa yang lebih komprehensif dibandingkan penggunaan satu partisi data [16] [14] [23].

DOI: https://doi.org/10.55537/cosie.v5i4.1855

Berdasarkan kajian literatur yang telah dilakukan, penelitian terkait klasifikasi tingkat obesitas umumnya berfokus pada perbandingan performa antaralgoritma *machine learning*, sedangkan kajian yang membandingkan hasil evaluasi model berdasarkan strategi validasi yang berbeda masih relatif terbatas. Beberapa penelitian telah menerapkan *K-Fold Cross Validation* sebagai metode evaluasi, namun belum banyak penelitian yang secara khusus mengevaluasi perbedaan estimasi performa Gaussian Naïve Bayes yang diperoleh melalui pendekatan train-test split dan *10-Fold Cross Validation* pada klasifikasi tingkat obesitas. Dengan demikian, masih terdapat kebutuhan untuk memahami bagaimana penggunaan strategi validasi yang berbeda dapat memengaruhi hasil evaluasi model pada dataset obesitas. Oleh karena itu, penelitian ini menerapkan algoritma Gaussian Naïve Bayes untuk mengklasifikasikan tingkat obesitas berdasarkan lima atribut utama, yaitu umur, tinggi badan, berat badan, indeks massa tubuh (IMT), dan aktivitas fisik. Selain itu, penelitian ini membandingkan hasil evaluasi model yang diperoleh menggunakan pendekatan train-test split 70:30 dan *10-Fold Cross Validation* untuk menganalisis perbedaan estimasi performa yang dihasilkan oleh kedua strategi validasi tersebut.

Kontribusi penelitian ini terletak pada evaluasi empiris terhadap perbedaan estimasi performa Gaussian Naïve Bayes yang dihasilkan oleh dua strategi validasi, yaitu train-test split 70:30 dan *10-Fold Cross Validation,* pada klasifikasi tingkat obesitas. Penelitian ini tidak berfokus pada pengembangan algoritma baru, melainkan pada analisis bagaimana strategi validasi yang berbeda menghasilkan estimasi performa yang berbeda pada dataset yang sama. Hasil penelitian diharapkan dapat menjadi referensi dalam pemilihan metode validasi pada penelitian klasifikasi kesehatan berbasis machine learning, khususnya pada kasus klasifikasi tingkat obesitas.

2. BAHAN DAN METODE

2.1. Desain Penelitian

Penelitian ini menerapkan metode kuantitatif dengan desain eksperimen untuk mengukur kemampuan algoritma Gaussian Naïve Bayes dalam mengidentifikasi kategori tingkat obesitas. Proses penelitian diawali dengan pengumpulan dataset, kemudian dilanjutkan dengan tahap pra-pemrosesan untuk menyiapkan data sebelum digunakan dalam pemodelan. Setelah itu dilakukan pembangunan model klasifikasi, proses pelatihan model, serta pengujian performa menggunakan beberapa metrik evaluasi. Untuk memperoleh hasil penilaian yang lebih reliabel, penelitian ini membandingkan dua pendekatan validasi, yaitu pembagian data menggunakan metode *train-test split* dan teknik *10-Fold Cross Validation*. Perbandingan kedua metode tersebut dilakukan untuk mengetahui konsistensi serta tingkat kestabilan performa model dalam mengklasifikasikan data obesitas. Perbandingan kedua metode dilakukan untuk mengevaluasi perbedaan estimasi performa yang dihasilkan oleh masing-masing strategi validasi serta mengamati konsistensi hasil evaluasi model.

2.2. Dataset

Penelitian ini menggunakan *dataset* sekunder yang diperoleh dari repositori publik Kaggle dengan nama *Obesity Prediction Dataset*. *Dataset* terdiri dari 1.000 data observasi yang merepresentasikan karakteristik fisik individu dan tingkat aktivitas fisik. Variabel target yang digunakan adalah kategori obesitas *(Obesity Category),* sedangkan variabel prediktor meliputi umur *(Age)*, tinggi badan *(Height)*, berat badan *(Weight)*, indeks massa tubuh (*Body Mass Index*/BMI), dan tingkat aktivitas fisik (*Physical Activity Level*). Kategori obesitas pada dataset terdiri atas empat kelas, yaitu *Underweight*, Normal, *Overweight*, dan *Obese*. Kategori obesitas terdiri atas empat kelas, yaitu *Underweight* sebanyak 143 data, *Normal* sebanyak 371 data, *Overweight* sebanyak 200 data, dan *Obese* sebanyak 286 data. Seluruh atribut prediktor berupa data numerik kontinu sehingga sesuai untuk dimodelkan menggunakan algoritma Gaussian Naïve Bayes. Dataset yang digunakan merupakan dataset publik yang telah memiliki label kategori obesitas dari penyedia dataset. Berdasarkan dokumentasi dataset, kategori obesitas ditentukan berdasarkan karakteristik individu yang tersedia dalam dataset dan digunakan sebagai label klasifikasi pada penelitian ini. Contoh data mentah (*raw dataset*) ditunjukkan pada Tabel 1.

Tabel 1. Sampel *Raw Dataset*

Age	Gender	Height	Weight	BMI	Physical Activity Level	Obesity Category
56	Male	173.575	71.982	23.892	4	Normal
69	Male	164.127	89.959	33.395	2	Obese
46	Female	168.072	72.931	25.818	4	Overweight

DOI: https://doi.org/10.55537/cosie.v5i4.1855

32	Male	168.460	84.887	29.912	3	Overweight
60	Male	183.569	69.039	20.488	3	Normal

Tabel 1 menunjukkan contoh data mentah sebelum dilakukan proses pra-pemrosesan. Dataset masih memuat seluruh atribut asli yang diperoleh dari sumber data, termasuk atribut *Gender* yang pada tahap selanjutnya tidak digunakan dalam proses pemodelan. Lalu Tabel 2 menunjukkan variabel yang digunakan dalam penelitian.

Tabel 2. Variabel Penelitian

Variabel	**Deskripsi**	**Tipe**
Age	Umur individu	Numerik
Height	Tinggi badan individu	Numerik
Weight	Berat badan individu	Numerik
BMI	Indeks Massa Tubuh	Numerik
Physical Activity Level	Tingkat aktivitas fisik	Numerik
Obesity Category	Kategori obesitas	Kategorikal

2.3. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas dan kesesuaian data sebelum digunakan dalam proses klasifikasi. Tahapan ini meliputi *data cleaning, label encoding*, dan *feature selection*. Pada tahap *data cleaning* dilakukan pemeriksaan terhadap *missing value*, *outlier*, dan data duplikat. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki *missing value* dan tidak ditemukan data duplikat. Deteksi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) dengan batas bawah sebesar Q1 − 1,5(IQR) dan batas atas sebesar Q3 + 1,5(IQR). Hasil pemeriksaan menunjukkan terdapat 18 data yang teridentifikasi sebagai outlier. Namun, data tersebut tetap dipertahankan karena masih merepresentasikan variasi alami karakteristik individu dalam dataset obesitas. Selanjutnya dilakukan *label encoding* pada variabel target dengan mengubah kategori obesitas ke dalam bentuk numerik, yaitu *Overweight* = 0, *Underweight* = 1, *Normal* = 2, dan *Obese* = 3. Tahap terakhir adalah *feature selection* dengan mempertahankan atribut yang digunakan dalam penelitian, yaitu *Age, Height, Weight*, BMI, dan *Physical Activity Level*, serta menghapus atribut yang tidak digunakan dalam pemodelan seperti *Gender*. Contoh data setelah melalui proses pra-pemrosesan ditunjukkan pada Tabel 3.

Tabel 3. Dataset Setelah Pra-pemrosesan

Age	**Height**	**Weight**	**BMI**	**Physical Activity Level**	**Obesity Category**
56	173.575	71.982	23.892	4	2
69	164.127	89.959	33.395	2	3
46	168.072	72.931	25.818	4	0
32	168.460	84.887	29.912	3	0
60	183.569	69.039	20.488	3	2

Tabel 3. menunjukkan contoh data setelah dilakukan tahap pra-pemrosesan. Pada tahap ini atribut *Gender* telah dihapus melalui proses *feature selection*, sedangkan variabel target *Obesity Category* telah ditransformasikan ke bentuk numerik menggunakan teknik *label encoding*. Dataset hasil pra-pemrosesan inilah yang digunakan pada tahap pemodelan Gaussian Naïve Bayes.

2.4. Model Gaussian Naïve Bayes

Model klasifikasi yang digunakan dalam penelitian ini adalah Gaussian Naïve Bayes. Algoritma ini merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen terhadap fitur lainnya. Gaussian Naïve Bayes dipilih karena mampu menangani data numerik kontinu dengan mengasumsikan bahwa setiap atribut mengikuti distribusi normal (Gaussian). Probabilitas posterior suatu kelas dihitung menggunakan Teorema Bayes yang ditunjukkan pada Persamaan (1).

$$P(C \mid X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (1)$$

DOI: https://doi.org/10.55537/cosie.v5i4.1855

dimana P(C|X) adalah probabilitas posterior kelas C terhadap data X, P(X|C) adalah probabilitas likelihood, P(C) adalah probabilitas prior, dan P(X) adalah *probabilitas evidence*. Kelas dengan nilai probabilitas posterior tertinggi akan dipilih sebagai hasil klasifikasi.

2.5. Skenario Pengujian

Evaluasi model dilakukan menggunakan dua skenario pengujian. Skenario pertama menggunakan metode train-test split dengan proporsi 70% data latih dan 30% data uji. Pembagian data dilakukan menggunakan parameter random state sebesar 42 untuk menjaga konsistensi hasil eksperimen. Skenario kedua menggunakan *Stratified 10-Fold Cross Validation*. Pada metode ini, dataset dibagi menjadi sepuluh subset dengan mempertahankan proporsi distribusi kelas pada setiap fold. Sembilan subset digunakan sebagai data latih dan satu subset digunakan sebagai data uji. Proses tersebut diulang sebanyak sepuluh kali hingga setiap subset memperoleh kesempatan menjadi data uji satu kali. Nilai performa akhir diperoleh dari rata-rata *(mean)* seluruh hasil pengujian pada sepuluh *fold*. Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka Pandas, NumPy, Matplotlib, Seaborn, dan Scikit-learn.

2.6. Metrik Evaluasi

Kinerja model dievaluasi menggunakan *Accuracy, Precision, Recall, F1-Score*, dan *Confusion Matrix. Accuracy* digunakan untuk mengukur proporsi prediksi yang benar terhadap keseluruhan data. Precision menunjukkan tingkat ketepatan prediksi pada suatu kelas, sedangkan Recall mengukur kemampuan model dalam mengidentifikasi seluruh data yang termasuk ke dalam kelas tertentu. F1-Score digunakan untuk memberikan keseimbangan antara *Precision* dan *Recall*. Karena penelitian ini melibatkan klasifikasi multikelas dengan distribusi data yang tidak sepenuhnya seimbang, nilai *Precision, Recall,* dan *F1-Score* dihitung menggunakan pendekatan *weighted average* sehingga setiap kelas memperoleh bobot yang sama dalam proses evaluasi. Selain metrik evaluasi, *confusion matrix* digunakan untuk menganalisis distribusi hasil klasifikasi dan mengidentifikasi pola kesalahan prediksi pada setiap kategori obesitas. Hasil *confusion matrix* disajikan dalam bentuk matriks aktual dan prediksi untuk memberikan informasi yang lebih rinci mengenai performa model pada masing-masing kelas. Persamaan yang digunakan untuk menghitung metrik evaluasi ditunjukkan pada Persamaan (2) hingga Persamaan (5). *Accuracy*

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

Precision

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

Recall

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

F1-*Score*

$$F1\text{-}Score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \tag{5}$$

3. HASIL

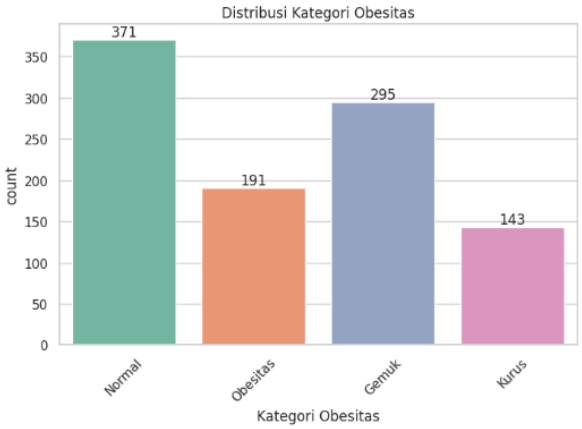
Dataset yang digunakan terdiri atas 1.000 data observasi dengan lima atribut prediktor, yaitu umur, tinggi badan, berat badan, indeks massa tubuh (IMT), dan tingkat aktivitas fisik. Variabel target berupa kategori obesitas yang terdiri dari empat kelas, yaitu Kurus, Normal, Gemuk, dan Obesitas. Pada tahap pra-pemrosesan dilakukan penyesuaian nama atribut, pemetaan label kelas ke dalam bahasa Indonesia, pengecekan *missing value*, serta transformasi label menggunakan teknik label *encoding*. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki data kosong sehingga seluruh data dapat digunakan pada proses pelatihan dan pengujian model. Selain pemeriksaan missing value, dilakukan pula pengecekan data duplikat dan *outlier*. Hasil pemeriksaan menunjukkan bahwa tidak ditemukan data duplikat pada dataset. Deteksi *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR) pada seluruh atribut numerik. Meskipun terdapat beberapa nilai yang berada di luar rentang IQR, nilai tersebut masih merepresentasikan kondisi

DOI: https://doi.org/10.55537/cosie.v5i4.1855

obesitas yang valid sehingga tidak dihapus dari dataset. Dengan demikian, seluruh 1.000 data observasi tetap digunakan dalam proses analisis.

Tabel 4. Hasil *Label Encoding*

Kategori Obesitas	Label
Gemuk	0
Kurus	1
Normal	2
Obesitas	3



Gambar 1. Distribusi Kelas Obesitas (Hasil EDA).

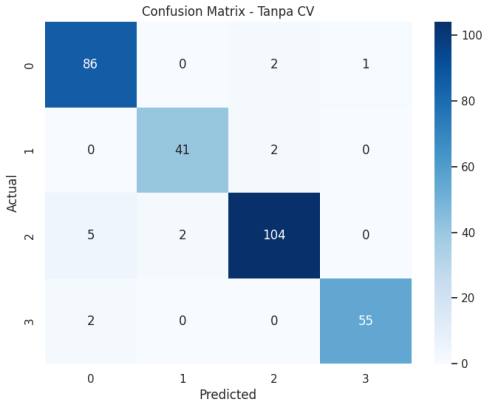
Berdasarkan Gambar 1 terlihat bahwa kategori Normal merupakan kelas mayoritas dengan 371 data, sedangkan kategori Kurus merupakan kelas minoritas dengan 143 data. Meskipun terdapat perbedaan jumlah data antar kelas, distribusi tersebut masih cukup representatif untuk proses klasifikasi dan evaluasi model. Distribusi lengkap kelas pada dataset terdiri atas 371 data kategori Normal, 191 data kategori Obesitas, 295 data kategori Gemuk, dan 143 data kategori Kurus. Distribusi ini menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) dalam tingkat ringan hingga sedang sehingga evaluasi model tidak hanya menggunakan akurasi, tetapi juga *precision, recall*, dan F1-*score*. Algoritma Gaussian Naïve Bayes digunakan dalam penelitian ini untuk mengklasifikasikan tingkat obesitas berdasarkan atribut umur, tinggi badan, berat badan, indeks massa tubuh (IMT), dan aktivitas fisik. Dataset yang digunakan terdiri dari 1.000 data observasi yang telah melalui tahap pra-pemrosesan dan dibagi menggunakan metode *train-test split* dengan rasio 70:30. Proses pembagian data dilakukan menggunakan parameter random_state=42 untuk menjamin reproduksibilitas hasil penelitian. Selain itu, digunakan pendekatan stratifikasi sehingga proporsi masing-masing kelas tetap terjaga pada data latih dan data uji. Hasil evaluasi model Gaussian Naïve Bayes ditunjukkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Gaussian Naïve Bayes

Metrik	Nilai
Accuracy	95,33%
Precision	95,58%
Recall	95,54%
F1-Score	95,54%

Perhitungan *precision, recall*, dan F1-*score* menggunakan metode *weighted average* untuk mempertimbangkan distribusi jumlah data pada masing-masing kelas. Berdasarkan Tabel 3, model Gaussian Naïve Bayes menghasilkan akurasi sebesar 95,33%. Nilai *precision, recall*, dan F1-*score* yang berada di atas 95% menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar pada data uji yang digunakan. Namun demikian, hasil tersebut merepresentasikan performa model pada satu pembagian data (*single train-test split*) sehingga masih bergantung pada komposisi data latih dan data uji yang terbentuk.

DOI: https://doi.org/10.55537/cosie.v5i4.1855



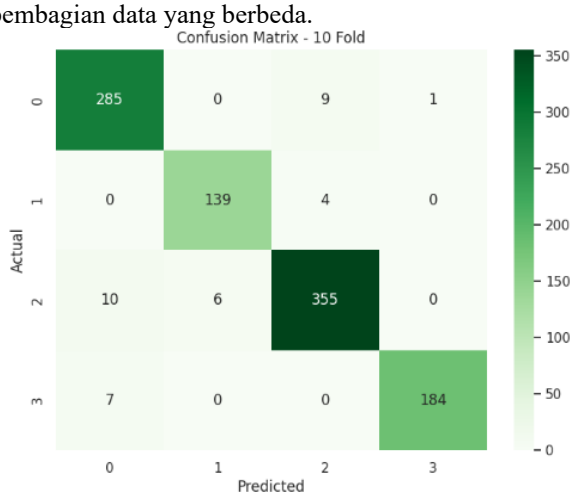
Gambar 2. *Confusion Matrix* Gaussian Naïve Bayes

Berdasarkan *confusion matrix*, sebanyak 286 dari 300 data uji berhasil diklasifikasikan dengan benar, sedangkan 14 data mengalami kesalahan klasifikasi. Dominasi nilai pada diagonal utama menunjukkan bahwa sebagian besar observasi berhasil diprediksi sesuai dengan kelas sebenarnya. Kesalahan klasifikasi masih ditemukan pada beberapa kelas yang memiliki karakteristik atribut yang saling berdekatan, terutama antara kategori Normal, Gemuk, dan Obesitas. Untuk memperoleh estimasi performa yang lebih komprehensif, model kemudian dievaluasi menggunakan teknik *10-Fold Cross Validation*. Pada skenario ini digunakan StratifiedKFold dengan nilai n_splits=10, shuffle=True, dan random_state=42 sehingga distribusi kelas tetap terjaga pada setiap fold. Hasil pengujian ditampilkan pada Tabel 6.

Tabel 6. Hasil Evaluasi Gaussian Naïve Bayes dengan *10-Fold Cross Validation*

Metrik	Nilai
Accuracy	96,30%
Precision	96,66%
Recall	96,45%
F1-Score	96,48%

Selain nilai rata-rata, evaluasi menggunakan *10-Fold Cross Validation* menghasilkan variasi performa antar-fold yang relatif kecil. Hal ini menunjukkan bahwa model memberikan hasil yang cukup konsisten pada berbagai pembagian data yang berbeda.



Gambar 3. *Confusion Matrix* Gaussian Naïve Bayes dengan *10-Fold Cross Validation*

Distribusi prediksi benar masih mendominasi diagonal utama matriks. Temuan ini menunjukkan bahwa performa model relatif konsisten pada berbagai subset data yang digunakan selama proses validasi silang. Selanjutnya dilakukan perbandingan performa antara Gaussian Naïve Bayes tanpa validasi silang dan Gaussian Naïve Bayes dengan *10-Fold Cross Validation*.

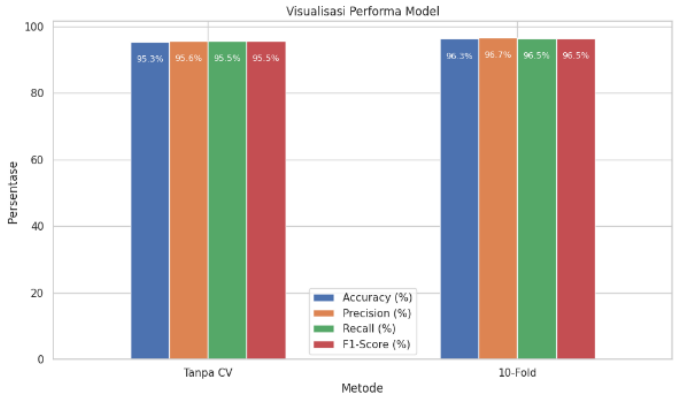
Tabel 7. Perbandingan Performa Model

Metode	Accuracy	Precision	Recall	F1-Score
Gaussian Naïve Bayes	95,33%	95,58%	95,54%	95,54%

DOI: https://doi.org/10.55537/cosie.v5i4.1855

Metode	Accuracy	Precision	Recall	F1-Score
Gaussian Naïve Bayes + 10-Fold Cross Validation	96,30%	96,66%	96,45%	96,48%

Terdapat perbedaan nilai performa antara kedua skenario evaluasi. Namun demikian, perbedaan tersebut tidak dapat langsung diinterpretasikan sebagai peningkatan kemampuan klasifikasi model karena kedua pendekatan menggunakan mekanisme evaluasi yang berbeda. *Train-test split* menghasilkan satu estimasi performa berdasarkan satu partisi data, sedangkan *10-Fold Cross Validation* menghasilkan rata-rata performa dari sepuluh kali pengujian. Oleh karena itu, hasil *cross validation* lebih tepat dipandang sebagai estimasi performa yang diperoleh dari beberapa pembagian data yang berbeda. Untuk memperjelas perbandingan performa kedua metode, hasil evaluasi divisualisasikan dalam bentuk grafik pada Gambar 4.



Gambar 4. Perbandingan Akurasi

Berdasarkan Tabel 7 terlihat bahwa rata-rata akurasi yang diperoleh melalui *10-Fold Cross Validation* sedikit lebih tinggi dibandingkan *train-test split*. Temuan ini menunjukkan bahwa estimasi performa model dapat berubah bergantung pada strategi evaluasi yang digunakan. Oleh karena itu, penggunaan 10-*Fold Cross Validation* memberikan gambaran yang lebih komprehensif mengenai performa Gaussian Naïve Bayes pada dataset yang digunakan dibandingkan evaluasi menggunakan satu pembagian data saja. Meskipun demikian, hasil penelitian ini masih memiliki beberapa keterbatasan. Dataset yang digunakan berasal dari satu sumber data sekunder sehingga belum dapat menggambarkan populasi yang lebih luas. Selain itu, penelitian ini belum melakukan validasi eksternal maupun perbandingan dengan algoritma klasifikasi lain yang lebih kompleks. Oleh karena itu, hasil yang diperoleh perlu diinterpretasikan sesuai dengan ruang lingkup dataset dan rancangan eksperimen yang digunakan.

4. PEMBAHASAN

Berdasarkan hasil evaluasi, model Gaussian Naïve Bayes mampu membedakan kategori obesitas dengan tingkat ketepatan yang tinggi pada seluruh metrik pengukuran. Pada skenario pembagian data menggunakan *train-test split* 70:30, model menghasilkan *accuracy* sebesar 95,33%, *precision* sebesar 95,58%, *recall* sebesar 95,54%, dan F1-*score* sebesar 95,54%. Tingginya nilai keempat metrik tersebut menunjukkan bahwa hubungan antara atribut umur, tinggi badan, berat badan, indeks massa tubuh (IMT), dan aktivitas fisik memiliki pola yang dapat dipelajari dengan baik oleh model Gaussian Naïve Bayes. Kondisi ini terjadi karena seluruh atribut yang digunakan merupakan data numerik kontinu yang sesuai dengan asumsi distribusi Gaussian yang menjadi dasar perhitungan probabilitas pada algoritma Gaussian Naïve Bayes. Namun, hasil tersebut hanya menggambarkan performa model pada satu pembagian data sehingga belum dapat digunakan untuk menyimpulkan kemampuan generalisasi model secara lebih luas. Selain itu, tingginya performa tidak secara langsung dapat dikaitkan dengan kesesuaian asumsi distribusi Gaussian karena penelitian ini belum melakukan analisis atau uji distribusi terhadap masing-masing atribut.

Hasil penelitian ini sejalan dengan penelitian [12] yang menerapkan Gaussian Naïve Bayes untuk klasifikasi diagnosis penyakit jantung dan memperoleh performa klasifikasi yang tinggi. Temuan serupa juga dilaporkan oleh [14] pada klasifikasi pasien gagal jantung, di mana Gaussian Naïve Bayes mampu memberikan hasil prediksi yang akurat melalui pendekatan probabilistik. Kesamaan hasil tersebut menunjukkan bahwa Gaussian Naïve Bayes merupakan algoritma yang efektif untuk menyelesaikan permasalahan klasifikasi pada bidang kesehatan karena memiliki proses komputasi yang sederhana, waktu pelatihan yang cepat, dan kemampuan generalisasi yang baik. Temuan tersebut menunjukkan bahwa Gaussian Naïve Bayes dapat diterapkan pada berbagai permasalahan klasifikasi di bidang kesehatan, terutama ketika atribut yang digunakan berupa data numerik. Namun, hasil dari penelitian-penelitian

tersebut tidak dapat secara langsung digunakan untuk menyimpulkan bahwa Gaussian Naïve Bayes memiliki kemampuan generalisasi yang sama pada dataset obesitas yang digunakan dalam penelitian ini.

Selain itu, penerapan *10-Fold Cross Validation* menghasilkan peningkatan performa model dengan nilai *accuracy* sebesar 96,30%, *precision* sebesar 96,66%, *recall* sebesar 96,45%, dan F1-*score* sebesar 96,48%. Nilai tersebut lebih tinggi dibandingkan hasil yang diperoleh melalui train-test split 70:30, tetapi perbedaan tersebut tidak dapat secara langsung diinterpretasikan sebagai peningkatan kemampuan model. Hal ini disebabkan oleh kedua pendekatan menggunakan prosedur evaluasi yang berbeda. Train-test split menghasilkan satu estimasi berdasarkan satu pembagian data, sedangkan 10-Fold Cross Validation menghasilkan estimasi rata-rata dari sepuluh proses validasi. Pada proses *10-Fold Cross Validation*, seluruh data memperoleh kesempatan untuk menjadi data latih dan data uji secara bergantian sehingga evaluasi tidak bergantung pada satu pembagian data tertentu. Dengan demikian, risiko bias evaluasi akibat pemilihan sampel data dapat diminimalkan [14], [21], [22].

Temuan penelitian ini juga mendukung hasil penelitian [13] yang menyatakan bahwa penerapan *K-Fold Cross Validation* dapat memberikan estimasi performa model berdasarkan beberapa pembagian data pada klasifikasi status gizi balita. Penelitian [16] juga menunjukkan bahwa penggunaan teknik *cross validation* menghasilkan pengukuran performa yang tidak hanya bergantung pada satu partisi data karena model diuji pada beberapa kombinasi data yang berbeda. Oleh karena itu, peningkatan performa yang diperoleh pada penelitian ini menunjukkan bahwa *10-Fold Cross Validation* tidak hanya berfungsi sebagai metode evaluasi, tetapi juga mampu memberikan gambaran yang lebih akurat mengenai kemampuan model ketika diterapkan pada data baru. Dengan demikian, pada penelitian ini, *10-Fold Cross Validation* lebih tepat dipandang sebagai strategi evaluasi untuk memperoleh estimasi performa yang lebih komprehensif, bukan sebagai metode yang meningkatkan kemampuan klasifikasi model itu sendiri.

Interpretasi terhadap hasil penelitian juga perlu mempertimbangkan kemungkinan hubungan yang sangat kuat antara variabel IMT dan kategori obesitas. IMT digunakan sebagai salah satu atribut prediktor dalam penelitian ini, sementara kategori obesitas pada dataset perlu dipastikan tidak dibentuk berdasarkan nilai IMT atau kriteria yang secara langsung berkaitan dengan IMT. Apabila pembentukan label menggunakan IMT, penggunaan IMT sebagai prediktor dapat menyebabkan target leakage atau hubungan yang hampir deterministik antara prediktor dan target sehingga dapat menggelembungkan nilai performa model. Oleh karena itu, sumber dataset perlu diverifikasi untuk memastikan mekanisme pembentukan label kategori obesitas sebelum performa model diinterpretasikan sebagai kemampuan prediktif yang bermakna.

Selain potensi target leakage, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan merupakan dataset sekunder dari satu sumber sehingga karakteristik data belum tentu mewakili populasi yang lebih luas. Penelitian ini juga belum menggunakan validasi eksternal pada dataset independen dan belum membandingkan Gaussian Naïve Bayes dengan model klasifikasi lain yang lebih kuat sebagai baseline. Oleh karena itu, hasil performa yang diperoleh perlu dipahami dalam konteks dataset dan rancangan evaluasi yang digunakan.

Secara keseluruhan, penelitian ini menunjukkan bahwa Gaussian Naïve Bayes dapat digunakan untuk mengklasifikasikan tingkat obesitas pada dataset yang digunakan dengan nilai performa yang tinggi. Penggunaan *10-Fold Cross Validation* menghasilkan estimasi performa rata-rata yang sedikit lebih tinggi dibandingkan train-test split 70:30 dan memungkinkan evaluasi dilakukan melalui beberapa pembagian data. Namun, perbedaan nilai tersebut tidak dapat dianggap sebagai bukti bahwa *10-Fold Cross Validation* meningkatkan kemampuan model. Validasi pada dataset eksternal, pemeriksaan terhadap kemungkinan target leakage, serta perbandingan dengan algoritma klasifikasi lainnya diperlukan untuk memperoleh kesimpulan yang lebih kuat mengenai kemampuan generalisasi model pada klasifikasi tingkat obesitas.

5. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Gaussian Naïve Bayes untuk mengklasifikasikan tingkat obesitas berdasarkan lima variabel prediktor, yaitu umur, tinggi badan, berat badan, indeks massa tubuh (IMT), dan tingkat aktivitas fisik. Hasil evaluasi menggunakan metode *train-test split* menghasilkan akurasi sebesar 95,33%, sedangkan evaluasi menggunakan 10-Fold Cross Validation menghasilkan rata-rata akurasi sebesar 96,30%, *precision* sebesar 96,66%, *recall* sebesar 96,45%, dan F1-*score* sebesar 96,48%. Hasil penelitian menunjukkan bahwa penggunaan strategi validasi yang berbeda memberikan estimasi performa yang berbeda terhadap model yang sama. Berdasarkan rata-rata hasil pengujian, *10-Fold Cross Validation* menghasilkan estimasi performa yang sedikit lebih tinggi dibandingkan *train-test split* pada dataset yang digunakan. Meskipun demikian, hasil penelitian ini masih terbatas pada satu dataset sekunder dan belum melibatkan validasi eksternal. Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset yang lebih beragam serta membandingkan Gaussian Naïve Bayes dengan algoritma klasifikasi lainnya untuk memperoleh evaluasi yang lebih komprehensif.

DOI: https://doi.org/10.55537/cosie.v5i4.1855

DAFTAR PUSTAKA

[1] W. Kurdanti *et al.*, "Jurnal Gizi Klinik Indonesia Faktor-Faktor yang Mempengaruhi Kejadian Obesitas pada Remaja," *Jurnal Gizi Klinik Indonesia*, vol. 11, no. 04, pp. 179–190, 2025.

[2] Juliatin and Hasniar, "Faktor-Faktor Penyebab Terjadinya Obesitas pada Remaja," *Student Research Journal (SRJ)*, vol. 2, no. 6, pp. 137–149, 2024, doi: 10.55606/srj-yappi.v2i6.1633.

[3] Nurzakiah, E. L. Achadi, and R. A. D. Sartika, "Faktor Risiko Obesitas pada Orang Dewasa Urban dan Rural," *Kesmas*, vol. 5, no. 1, pp. 29–34, 2023, doi: 10.21109/kesmas.v5i1.159.

[4] S. Aziz, Y. Pramana, and Sukarni, "Hubungan Aktivitas Fisik Dengan Kejadian Obesitas Pada Remaja," *Mahesa: Malahayati Health Student Journal*, vol. 3, pp. 1115–1124, 2023.

[5] J. M. Gorriz, R. Martin-Clemente, F. Segovia, J. Ramirez, A. Ortiz, and J. Suckling, "Is K-fold cross validation the best model selection method for machine learning?," *Information Fusion*, vol. 135, p. 104404, Nov. 2026, doi: 10.1016/j.inffus.2026.104404.

[6] M. N. Fahmi, "Implementasi Machine Learning menggunakan Python Library: Scikit-Learn," *Sains Data*, vol. 2, pp. 87–96, 2023.

[7] R. S. Nurhalizah, R. Ardianto, and Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *JIKI*, vol. 4, no. 1, pp. 61–72, 2024.

[8] M. S. Hasibuan, Y. R. Nasution, I. Komputer, U. Islam, and N. Sumatera, "OPTIMASI MODEL SEMI-SUPERVISED LEARNING DENGAN SVM," vol. 9, no. 2, pp. 231–239, 2024.

[9] M. S. Hasibuan and Suhardi, "Higher Education and Health Breakthroughs: Residual Network for Pneumonia Detection from Chest X-ray Images," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 15, no. 5, 2025, doi: 10.18517/ijaseit.15.5.20483.

[10] J. Oriana, W. Sinaga, M. Fathurahman, S. Wahyuningsih, and M. N. Hayati, "Evaluating Different K Values in K-Fold Cross Validation for Binary Logistic Regression to Classify Poverty," *Jurnal Gaussian*, vol. 8, no. 2, pp. 189–198, 2025.

[11] M. A. Faradeya and E. R. Subhiyakto, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes," *Jurnal Algoritma*, pp. 115–127, 2025, doi: 10.33364/algoritma/v.22-1.2178.

[12] I. Akil and I. Chaidir, "Classification of Heart Disease Diagnoses Using Gaussian Naive Bayes," *Komputasi*, vol. 21, pp. 31–36, 2024.

[13] Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naive Bayes Classifier dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *JOSYC*, vol. 4, no. 3, 2023, doi: 10.47065/josyc.v4i3.3414.

[14] H. Hafid, "Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor pada Data Kasus Covid-19 di Indonesia," *Journal of Mathematics, Computations, and Statistics*, vol. 6, no. 2, pp. 161–168, 2023.

[15] M. Andani, J. Triloka, S. Y. Irianto, and H. W. Nugroho, "Performance Comparison of K-Nearest Neighbor, Naive Bayes, and Random Forest Algorithms in Obesity Prediction," *Sinkron*, vol. 9, no. 1, pp. 502–510, 2025.

[16] S. Khoerunnisa, D. F. Shiddiq, and D. Nurhayati, "Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink," *MALCOM*, vol. 5, pp. 566–577, 2025.

[17] M. Jamali *et al.*, "Applications of traditional machine learning and deep learning algorithms in obesity prediction or classification: a systematic review of comparative performance," *NPJ Digit. Med.*, 2026, doi: 10.1038/s41746-026-02986-8.

[18] H. T. Santoso, F. A. Felmidi, A. Nur, A. Ristyawan, and E. Daniati, "Analisis Kinerja Algoritma Data Mining pada Klasifikasi Tingkat Obesitas dengan K-Fold Cross Validation dan AUC," *Inotek*, vol. 8, pp. 113–122, 2024.

[19] T. H. Pinem and Z. P. Putra, "Evaluasi Kinerja Algoritma Klasifikasi Deep Learning dalam Prediksi Diabetes," vol. 17, no. 1, pp. 17–28, 2025, doi: 10.22441/fifo.2025.v17i1.003.

[20] M. M. A. Zaid and A. A. Mohammed, "Diabetes Prediction Using Machine Learning: Methods, Challenges, and Insights from a Systematic Literature Review," *JKMC*, vol. 12, no. 2, pp. 21–27, 2025.

[21] Wijiyanto, A. I. Pradana, Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.

[22] J. O. W. Sinaga, M. Fathurahman, S. Wahyuningsih, and M. N. Hayati, "Evaluating Different K Values in K-Fold Cross Validation for Binary Logistic Regression to Classify Poverty," *Jurnal Varian*, vol. 8, no. 2, pp. 189–198, 2025, doi: 10.30812/varian.v8i2.4403.

[23] R. S. Pall, S. Yadav, S. Bhalerao, S. Sahu, R. Ahluwalia, and B. Awadhiya, “Comprehensive Evaluation of Machine Learning for Type 2 Diabetes Risk Prediction: Large-Scale External Validation and Fairness Analysis,” in *2026 International Conference on Intelligent Processing, Hardware, Electronics, and Radio Systems (CIPHER)*, IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/cipher70417.2026.11523789.