

KLASIFIKASI TINGKAT OBESITAS BERBASIS DATA ANTROPOMETRI MENGGUNAKAN ALGORITMA SVM DAN KNN

Dian Sela Anjelina^{1*}, Asriyanik², Prajoko³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi, Sukabumi, Indonesia

Keywords:

Obesitas;
Data Antropometri;
Support Vector Machine;
K-Nearest Neighbor;
SMOTE.

Correspondent Email:

disela252@ummi.ac.id

Abstrak. Penelitian ini membangun model klasifikasi tingkat obesitas berbasis data antropometri menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Data berasal dari pasien Penyakit Tidak Menular (PTM) pada Puskesmas Kota Sukabumi. Setelah pembersihan data, digunakan 2.662 data dengan variabel usia, jenis kelamin, berat badan, dan tinggi badan. Indeks Massa Tubuh (IMT) dihitung dari berat badan dan tinggi badan untuk membentuk label kelas berdasarkan klasifikasi IMT Asia-Pasifik, yaitu underweight, normal, overweight, obesitas I, dan obesitas II, tetapi tidak digunakan sebagai fitur input model. Tahapan penelitian mengikuti CRISP-DM dengan pipeline SMOTE, StandardScaler, SVM kernel RBF, dan KNN. Hasil evaluasi menunjukkan SVM memperoleh akurasi 97,75%, sedangkan KNN memperoleh 91,93%. Cross validation menghasilkan rata-rata akurasi 98,31% pada SVM dan 93,50% pada KNN. Penambahan usia dan jenis kelamin meningkatkan akurasi SVM dari 97,00% menjadi 97,75%. Hasil ini menunjukkan bahwa SVM lebih unggul untuk klasifikasi tingkat obesitas berbasis data antropometri.

Abstract. This study developed an obesity level classification model based on anthropometric data using Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). The dataset was obtained from non-communicable disease (NCD) patients at Sukabumi City Health Center. After data cleaning, 2,662 records containing age, sex, weight, and height were used. Body Mass Index (BMI) was calculated from weight and height to generate class labels based on the Asia-Pacific BMI classification, namely underweight, normal, overweight, obesity I, and obesity II, but it was not used as a model input feature. The research followed CRISP-DM with a pipeline consisting of SMOTE, StandardScaler, SVM with an RBF kernel, and KNN. The results show that SVM achieved 97.75% accuracy, while KNN achieved 91.93%. Cross validation produced 98.31% mean accuracy for SVM and 93.50% for KNN. Adding age and sex improved SVM accuracy from 97.00% to 97.75%. These results indicate that SVM is superior for obesity level classification based on anthropometric data.



Copyright © 2026 The Author(s),
Published by Jurnal Informatika
dan Teknik Elektro Terapan. This
is an open-access article
distributed under the terms of the
Creative Commons Attribution-
NonCommercial 4.0 International
License (CC BY-NC 4.0).

1. PENDAHULUAN

Obesitas menjadi masalah kesehatan yang meningkat dan berhubungan dengan risiko penyakit tidak menular, seperti diabetes melitus, hipertensi, dan penyakit kardiovaskular. World Health Organization melaporkan bahwa prevalensi obesitas global terus meningkat sejak 1990 [1]. Di Indonesia, klasifikasi status obesitas dewasa menggunakan Indeks Massa Tubuh (IMT) dengan standar Asia-Pasifik yang tercantum dalam pedoman Kementerian Kesehatan Republik Indonesia [2].

IMT mudah digunakan karena hanya membutuhkan berat badan dan tinggi badan. Namun, proses klasifikasi yang hanya berhenti pada perhitungan IMT belum memanfaatkan pola data antropometri lain. Dalam data kesehatan, usia dan jenis kelamin dapat memberi informasi tambahan terhadap variasi komposisi tubuh dan pola status gizi. Karena itu, pendekatan machine learning dapat digunakan untuk membangun klasifikasi tingkat obesitas yang lebih adaptif terhadap pola data.

Penelitian sebelumnya telah membandingkan beberapa algoritma untuk klasifikasi obesitas, seperti KNN, SVM, Decision Tree, Random Forest, dan Artificial Neural Network [3], [4]. Hasil penelitian tersebut menunjukkan bahwa SVM sering memberi performa kompetitif pada data numerik dan multikelas. Penelitian lain juga menegaskan bahwa data antropometri dapat digunakan sebagai dasar klasifikasi obesitas berbasis machine learning [5], sedangkan optimasi SVM dengan SMOTE dan tuning parameter dapat meningkatkan kinerja pada data yang tidak seimbang [6].

Meskipun demikian, sebagian penelitian masih berfokus pada klasifikasi umum dan belum menelaah secara spesifik pengaruh penambahan variabel usia dan jenis kelamin terhadap performa SVM. Selain itu, pengujian model perlu membandingkan SVM dengan baseline yang sederhana, seperti KNN, pada dataset yang sama. Studi berbasis JITET juga menunjukkan bahwa SVM relevan untuk klasifikasi data kesehatan multivariabel [7].

Penelitian ini bertujuan untuk membangun model klasifikasi tingkat obesitas berbasis data antropometri menggunakan SVM dan KNN, menganalisis pengaruh penambahan usia dan jenis kelamin terhadap kinerja SVM, serta membandingkan performa SVM dan KNN pada data pasien Penyakit Tidak Menular Puskesmas Kota Sukabumi. Kebaruan penelitian ini terletak pada penggunaan label IMT Asia-Pasifik sebagai ground truth, pemisahan IMT dari fitur input, perbandingan skenario fitur SVM, dan implementasi model dalam aplikasi web sederhana.

2. TINJAUAN PUSTAKA

2.1 Machine Learning untuk Klasifikasi Kesehatan

Machine learning merupakan pendekatan komputasi yang memungkinkan sistem mempelajari pola dari data untuk menghasilkan prediksi atau klasifikasi [5]. Dalam bidang kesehatan, machine learning dapat digunakan untuk mengolah data pasien, mengenali pola antarvariabel, dan membantu proses pengambilan keputusan berbasis data [8]. Penelitian ini menggunakan pendekatan supervised learning karena label kelas obesitas telah ditentukan berdasarkan klasifikasi IMT Asia-Pasifik [2]. Model dilatih menggunakan data antropometri, kemudian diuji untuk memprediksi kategori tingkat obesitas pada data baru [5].

2.2 Klasifikasi Obesitas Berbasis Antropometri

Data antropometri merupakan data pengukuran fisik tubuh, seperti berat badan, tinggi badan, usia, dan jenis kelamin. Dalam penelitian ini, berat badan dan tinggi badan digunakan untuk menghitung IMT, lalu IMT dipakai untuk membentuk label kelas. Rumus IMT dituliskan pada Persamaan (1).

$$IMT = \frac{\text{berat badan (kg)}}{(\text{tinggi badan (m)})^2} \quad (1)$$

Kelas target mengikuti klasifikasi IMT Asia-Pasifik. IMT hanya digunakan sebagai dasar pembentukan label. IMT tidak digunakan sebagai fitur input model agar model belajar dari variabel antropometri yang independen.

Tabel 1. Klasifikasi IMT Asia-Pasifik

No.	Kategori	Rentang IMT
1	Underweight	< 18,5
2	Normal	18,5 - 22,9
3	Overweight	23,0 - 24,9
4	Obesitas I	25,0 - 29,9
5	Obesitas II	≥ 30,0

2.3 Support Vector Machine dan K-Nearest Neighbor

Support Vector Machine merupakan algoritma supervised learning yang membentuk hyperplane untuk memisahkan kelas data. SVM dengan kernel Radial Basis Function dapat menangani pola data nonlinier dan cocok untuk data numerik dengan fitur terbatas [9].

K-Nearest Neighbor melakukan klasifikasi berdasarkan kedekatan jarak antara data baru dan data latih. KNN mudah diimplementasikan dan sering dipakai sebagai baseline, tetapi sensitif terhadap skala fitur dan pemilihan jumlah tetangga [10]. Karena itu, penelitian ini menggunakan StandardScaler pada kedua algoritma.

2.4 CRISP-DM dan Evaluasi Model

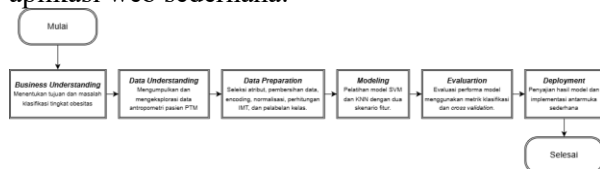
CRISP-DM menyediakan tahapan sistematis untuk penelitian data mining, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [11]. Kinerja model klasifikasi dapat dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score [8]. Cross validation digunakan untuk melihat stabilitas performa model [12]. ROC-AUC multikelas digunakan untuk melihat kemampuan model membedakan kelas [13]. Permutation importance digunakan untuk menilai kontribusi fitur terhadap prediksi [14]. Paired t-test digunakan untuk menguji signifikansi perbedaan performa dua model [15].

3. METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini menggunakan metode eksperimen kuantitatif berbasis machine learning untuk membangun model klasifikasi tingkat obesitas berdasarkan data antropometri. Alur penelitian

disusun menggunakan kerangka CRISP-DM yang terdiri atas business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Tahap business understanding digunakan untuk menetapkan tujuan klasifikasi tingkat obesitas. Tahap data understanding dilakukan untuk memahami struktur data, atribut, dan kualitas data. Tahap data preparation digunakan untuk menyiapkan dataset bersih, membentuk label kelas, membagi data, menyeimbangkan kelas, dan melakukan normalisasi fitur. Tahap modeling digunakan untuk membangun model SVM dan KNN. Tahap evaluation digunakan untuk mengukur performa model, sedangkan tahap deployment digunakan untuk mengimplementasikan model ke dalam aplikasi web sederhana.



Gambar 1. Alur Penelitian Berdasarkan CRISP-DM

3.2. Sumber Data dan Teknik Pengumpulan Data

Data yang digunakan merupakan data sekunder pasien Penyakit Tidak Menular pada Puskesmas Kota Sukabumi. Data diperoleh dari hasil pencatatan pelayanan kesehatan yang telah tersedia pada instansi, sehingga peneliti tidak melakukan pengukuran langsung terhadap pasien. Setelah proses seleksi dan pembersihan data, diperoleh 2.662 data yang digunakan dalam pemodelan.

Selain pengumpulan data sekunder, penelitian ini juga menggunakan studi pustaka untuk mendukung dasar teori, penentuan klasifikasi IMT Asia-Pasifik, pemilihan algoritma, serta teknik evaluasi model. Sumber pustaka yang digunakan berkaitan dengan klasifikasi obesitas, data antropometri, IMT, SVM, KNN, SMOTE, CRISP-DM, dan evaluasi model klasifikasi.

3.3. Variabel Penelitian dan Pembentukan Label

Variabel input yang digunakan dalam penelitian terdiri atas usia, jenis kelamin, berat badan (BB), dan tinggi badan (TB). Variabel nama pasien tidak digunakan karena tidak relevan terhadap proses klasifikasi dan berkaitan dengan kerahasiaan identitas pasien. IMT dihitung dari BB dan TB untuk membentuk label kelas berdasarkan klasifikasi IMT Asia-Pasifik. IMT tidak digunakan sebagai fitur input model.

Tabel 2. Variabel Penelitian

No.	Variabel	Tipe	Keterangan
1	Usia	Numerik	Umur pasien dalam tahun

No.	Variabel	Tipe	Keterangan
2	Jenis kelamin	Kategorikal	Perempuan atau laki-laki
3	Berat badan	Numerik	Berat badan dalam kg
4	Tinggi badan	Numerik	Tinggi badan dalam cm

3.4. Data Preparation

Data preparation dilakukan melalui beberapa tahap, yaitu penghapusan data duplikat, konversi usia, BB, dan TB ke bentuk numerik, penanganan nilai kosong, encoding jenis kelamin, perhitungan ulang IMT, pembentukan label kelas, pembagian data, penyeimbangan kelas, dan normalisasi fitur. Jenis kelamin dikodekan menjadi 0 untuk perempuan dan 1 untuk laki-laki. Kode tersebut hanya berfungsi sebagai representasi numerik agar dapat diproses oleh algoritma.

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratifikasi kelas. Dari 2.662 data, sebanyak 2.129 data digunakan sebagai data latih dan 533 data digunakan sebagai data uji. SMOTE diterapkan hanya pada data latih untuk menyeimbangkan distribusi kelas, sedangkan data uji tetap dipertahankan pada distribusi asli agar evaluasi model tetap objektif. StandardScaler digunakan dalam pipeline karena SVM dan KNN sensitif terhadap perbedaan skala fitur.

3.5. Modeling dan Teknik Evaluasi

Penelitian ini membangun tiga model. Model pertama adalah SVM multidimensi dengan fitur usia, jenis kelamin, BB, dan TB. Model kedua adalah KNN multidimensi dengan fitur yang sama sebagai model pembanding. Model ketiga adalah SVM sederhana dengan fitur BB dan TB untuk menguji pengaruh penambahan usia dan jenis kelamin terhadap performa SVM.

Tabel 3. Konfigurasi Model

No.	Model	Fitur	Konfigurasi
1	SVM multidimensi	Usia, JK, BB, TB	SMOTE, StandardScaler, SVC RBF
2	KNN multidimensi	Usia, JK, BB, TB	SMOTE, StandardScaler, KNN
3	SVM sederhana	BB, TB	StandardScaler, SVC RBF

Optimasi parameter dilakukan menggunakan GridSearchCV dengan Stratified K-Fold Cross Validation sebanyak lima lipatan. Teknik ini digunakan untuk mencari kombinasi parameter terbaik pada model SVM dan KNN. Parameter hasil optimasi kemudian digunakan pada tahap pengujian model terhadap data uji.

Evaluasi model dilakukan menggunakan accuracy, precision, recall, F1-score, confusion matrix, classification report, cross validation, ROC-AUC multiclass, validation curve, permutation importance, paired t-test, dan perbandingan skenario fitur. Evaluasi tersebut digunakan untuk mengetahui kinerja SVM, membandingkan performa SVM dan KNN, serta menilai pengaruh penambahan usia dan jenis kelamin terhadap performa model.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian dan pembahasan mengenai klasifikasi tingkat obesitas berbasis data antropometri menggunakan algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). Penyajian hasil disusun berdasarkan tahapan CRISP-DM, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [11]. SVM dan KNN digunakan karena keduanya relevan dalam penelitian klasifikasi obesitas dan data kesehatan berbasis machine learning [3]. Setiap tahapan menjelaskan proses yang dilakukan, mulai dari pemahaman masalah, pemeriksaan data, persiapan dataset, pembangunan model, evaluasi performa, hingga implementasi model ke dalam aplikasi web. Hasil evaluasi dianalisis untuk mengetahui kinerja SVM, membandingkan performa SVM dan KNN, serta menilai pengaruh penambahan variabel usia dan jenis kelamin terhadap kinerja model.

4.1. Business Understanding

Tahap business understanding menetapkan masalah utama penelitian, yaitu kebutuhan model klasifikasi tingkat obesitas berbasis data antropometri. Penilaian obesitas secara konvensional umumnya bertumpu pada perhitungan IMT. Pendekatan tersebut praktis, tetapi belum menunjukkan bagaimana pola usia, jenis kelamin, berat badan, dan tinggi badan dapat dipelajari oleh model klasifikasi.

Berdasarkan masalah tersebut, penelitian diarahkan untuk membangun model klasifikasi tingkat obesitas menggunakan SVM sebagai model utama dan KNN sebagai pembanding. Penelitian juga menguji pengaruh penambahan usia dan jenis kelamin terhadap SVM melalui perbandingan antara fitur sederhana, yaitu BB dan TB, dengan fitur multidimensi, yaitu usia, JK, BB, dan TB. Output akhirnya berupa model terbaik yang dapat diterapkan pada aplikasi web sederhana.

4.2. Data Understanding

Tahap data understanding dilakukan untuk memahami struktur dataset, atribut yang tersedia, dan kesiapan data untuk pemodelan. Data berasal dari pasien Penyakit Tidak Menular pada

Puskesmas Kota Sukabumi. Setelah seleksi dan pembersihan, dataset yang digunakan berjumlah 2.662 data. Atribut yang dipakai sebagai fitur input adalah usia, jenis kelamin, BB, dan TB. Kolom IMT dan kategori awal tidak digunakan sebagai fitur karena IMT dihitung ulang dari BB dan TB, lalu kategori dibentuk ulang berdasarkan standar IMT Asia-Pasifik.

Tabel 4. Distribusi Kelas Tingkat Obesitas

No.	Kategori	Jumlah	Persentase
1	Underweight	535	20,10%
2	Normal	535	20,10%
3	Overweight	535	20,10%
4	Obesitas I	596	22,39%
5	Obesitas II	461	17,32%
Total		2.662	100%

Berdasarkan Tabel 4, seluruh kategori tingkat obesitas memiliki jumlah data yang cukup untuk klasifikasi multikelas. Kategori Obesitas I menjadi kelas terbesar dengan 596 data atau 22,39%, sedangkan Obesitas II menjadi kelas terkecil dengan 461 data atau 17,32%. Kategori Underweight, Normal, dan Overweight masing-masing berjumlah 535 data atau 20,10%. Perbedaan jumlah antar kelas menunjukkan perlunya penyeimbangan data pada tahap data preparation agar model tidak cenderung mempelajari kelas dengan jumlah data lebih besar.

4.3. Data Preparation

Tahap data preparation dilakukan untuk menyiapkan dataset agar dapat digunakan dalam proses pemodelan. Proses ini mencakup penghapusan data duplikat, konversi usia, berat badan, dan tinggi badan ke bentuk numerik, penanganan nilai kosong, encoding jenis kelamin, perhitungan ulang IMT, pembentukan label kelas, pembagian data, penyeimbangan kelas, dan normalisasi fitur. Perempuan dikodekan sebagai 0 dan laki-laki sebagai 1. Kode tersebut hanya berfungsi sebagai representasi numerik agar dapat diproses oleh algoritma, bukan menunjukkan urutan tingkat.

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan stratifikasi kelas. Dari 2.662 data, sebanyak 2.129 data digunakan sebagai data latih dan 533 data digunakan sebagai data uji. Stratifikasi digunakan agar proporsi setiap kelas tetap terjaga pada data latih dan data uji.

Tabel 5. Distribusi Data Latih Sebelum dan Sesudah SMOTE

Kategori	Sebelum	Sesudah
Underweight	428	476
Normal	428	476
Overweight	428	476
Obesitas I	476	476
Obesitas II	369	476
Total	2.129	2.380

Berdasarkan Tabel 5, distribusi data latih sebelum SMOTE masih menunjukkan perbedaan jumlah antar kelas, terutama pada kelas Obesitas II yang memiliki 369 data. Setelah SMOTE diterapkan, seluruh kelas pada data latih menjadi seimbang dengan jumlah masing-masing 476 data, sehingga total data latih menjadi 2.380 data. SMOTE hanya diterapkan pada data latih, sedangkan data uji tetap dipertahankan pada distribusi asli agar evaluasi model tetap objektif.

Tahap akhir dilakukan normalisasi fitur menggunakan StandardScaler karena SVM dan KNN sensitif terhadap perbedaan skala data. Dengan tahapan ini, dataset telah memiliki struktur yang bersih, seimbang pada data latih, dan siap digunakan pada tahap modeling.

4.4. Modeling

Tahap modeling membangun tiga model. Model pertama adalah SVM multidimensi dengan fitur usia, JK, BB, dan TB. Model kedua adalah KNN multidimensi dengan fitur yang sama sebagai baseline pembanding. Model ketiga adalah SVM sederhana dengan fitur BB dan TB untuk menilai pengaruh penambahan usia dan jenis kelamin. Seluruh model menggunakan target kelas yang sama, yaitu Underweight, Normal, Overweight, Obesitas I, dan Obesitas II.

Berdasarkan hasil optimasi menggunakan GridSearchCV dengan Stratified K-Fold Cross Validation sebanyak lima lipatan, diperoleh parameter terbaik pada model SVM dan KNN sebagaimana ditunjukkan pada Tabel 6.

Tabel 6. Hasil Optimasi Parameter Model

No.	Model	Parameter Terbaik
1	SVM multidimensi	Kernel=RBF, C=100, Gamma=0,1
2	KNN multidimensi	metric=Manhattan, n_neighbors=9, weights=distance

Berdasarkan Tabel 6, model SVM multidimensi memperoleh parameter terbaik berupa kernel RBF, C=100, dan Gamma=0,1. Model KNN multidimensi memperoleh parameter terbaik berupa metric Manhattan, n_neighbors=9, dan weights = distance. Parameter tersebut digunakan dalam proses prediksi pada data uji dan menjadi dasar evaluasi performa model pada tahap berikutnya.

4.5. Evaluation

Tahap evaluation dilakukan untuk mengukur kinerja model SVM dan KNN pada data uji sebanyak 533 data. Evaluasi dilakukan menggunakan accuracy, precision, recall, F1-score, confusion matrix, cross validation, ROC-AUC multiclass, validation curve, permutation importance, paired t-test, perbandingan skenario fitur, dan validasi data baru. Kombinasi evaluasi

tersebut digunakan agar performa model tidak hanya dilihat dari akurasi keseluruhan, tetapi juga dari performa setiap kelas, stabilitas model, signifikansi perbedaan model, serta kontribusi fitur terhadap hasil prediksi.

Model SVM multidimensi memperoleh akurasi sebesar 97,75%. Dari 533 data uji, sebanyak 521 data berhasil diklasifikasikan dengan benar dan 12 data mengalami kesalahan klasifikasi. Nilai macro precision sebesar 97,80%, macro recall sebesar 97,75%, dan macro F1-score sebesar 97,75%. Hasil ini menunjukkan bahwa SVM mampu menjaga performa klasifikasi yang merata pada seluruh kelas tingkat obesitas.

Model KNN memperoleh akurasi sebesar 91,93%, dengan macro precision sebesar 92,35%, macro recall sebesar 91,93%, dan macro F1-score sebesar 91,88%. Nilai tersebut menunjukkan bahwa KNN masih mampu melakukan klasifikasi dengan baik, tetapi performanya lebih rendah dibandingkan SVM. Kesalahan KNN lebih banyak terjadi pada kelas yang memiliki rentang IMT berdekatan, khususnya Normal, Overweight, dan Obesitas I.

Tabel 7. Ringkasan Evaluasi Model SVM dan KNN

Model	Acc.	Macro P	Macro R	Macro F1
SVM	97,75%	97,80%	97,75%	97,75%
KNN	91,93%	92,35%	91,93%	91,88%

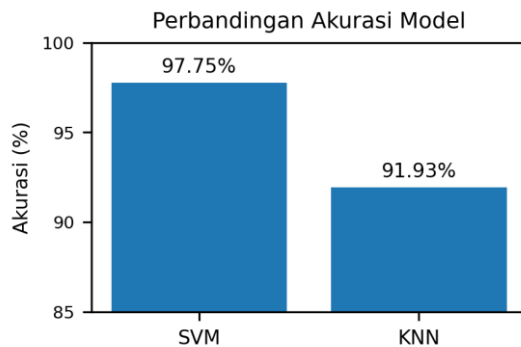
Hasil classification report menunjukkan bahwa SVM memiliki performa tinggi pada seluruh kategori. Kelas Underweight memperoleh F1-score 0,99, Normal 0,98, Overweight 0,97, Obesitas I 0,97, dan Obesitas II 0,99. Pada model KNN, nilai F1-score setiap kelas masih berada pada kategori baik, tetapi lebih rendah dibandingkan SVM, terutama pada kelas Normal dan Overweight.

Tabel 8. Classification Report Ringkas SVM dan KNN

Kategori	F1 SVM	F1 KNN	Support
Underweight	0,99	0,93	107
Normal	0,98	0,88	107
Overweight	0,97	0,90	107
Obesitas I	0,97	0,92	120
Obesitas II	0,99	0,97	92

Confusion matrix SVM menunjukkan bahwa sebagian besar hasil prediksi berada pada diagonal utama. Kesalahan klasifikasi SVM hanya terjadi pada sebagian kecil data, terutama pada kelas yang berdekatan secara nilai IMT. Pada model KNN, jumlah kesalahan lebih besar karena algoritma ini sangat bergantung pada kedekatan jarak antar data. Pola ini menjelaskan mengapa SVM lebih stabil

dalam memisahkan kelas yang batas antropometrinya berdekatan.



Gambar 2. Perbandingan akurasi model SVM dan KNN

Hasil cross validation memperkuat hasil pengujian pada data uji. SVM memperoleh rata-rata akurasi sebesar 98,31%, sedangkan KNN memperoleh rata-rata akurasi sebesar 93,50%. Variasi skor SVM juga lebih kecil dibandingkan KNN, sehingga SVM menunjukkan stabilitas performa yang lebih baik pada lima fold pengujian.

Tabel 9. Hasil Cross Validation dan Paired t-test

Model	Skor Fold	Mean
SVM	0,9831; 0,9775; 0,9868; 0,9887; 0,9793	98,31%
KNN	0,9512; 0,9250; 0,9417; 0,9248; 0,9323	93,50%

Paired t-test terhadap skor cross validation menghasilkan p-value sebesar 0,000442. Nilai tersebut lebih kecil dari 0,05, sehingga perbedaan performa antara SVM dan KNN signifikan secara statistik. Karena rata-rata akurasi SVM lebih tinggi dibandingkan KNN, maka SVM dipilih sebagai model utama dalam penelitian ini.

Pengujian ROC-AUC multiclass menggunakan pendekatan One-vs-Rest menghasilkan nilai AUC sebesar 1,00 pada setiap kelas dalam pembulatan dua desimal, dengan macro average sebesar 0,9998. Hasil ini menunjukkan bahwa SVM memiliki kemampuan pemisahan kelas yang sangat kuat. Validation curve juga menunjukkan bahwa SVM memiliki kemampuan pemisahan kelas yang sangat kuat. Validation curve juga menunjukkan bahwa parameter C = 100 menghasilkan skor validasi tertinggi dengan selisih kecil antara skor training dan validation, sehingga tidak terlihat indikasi overfitting yang kuat pada rentang parameter yang diuji.

Tabel 10. Ringkasan ROC-AUC dan Validation Curve SVM

Komponen	Hasil
ROC-AUC tiap kelas	1,00
ROC-AUC macro average	0,9998
Parameter C terbaik	100
Indikasi overfitting	Tidak kuat

Perbandingan skenario fitur dilakukan untuk mengetahui pengaruh penambahan usia dan jenis kelamin terhadap kinerja SVM. Model SVM sederhana yang hanya menggunakan BB dan TB memperoleh akurasi sebesar 97,00%, sedangkan SVM multidimensi yang menggunakan usia, JK, BB, dan TB memperoleh akurasi sebesar 97,75%. Dengan demikian, penambahan usia dan jenis kelamin meningkatkan akurasi sebesar 0,75 persentase poin.

Tabel 11. Skenario Fitur dan Permutation Importance

Komponen	Nilai
Akurasi SVM sederhana (BB, TB)	97,00%
Akurasi SVM multidimensi	97,75%
Peningkatan akurasi	0,75 poin
Importance BB	0,771107
Importance TB	0,215009
Importance usia	Rendah
Importance JK	Mendekati nol

Hasil permutation importance menunjukkan bahwa BB menjadi fitur paling dominan, diikuti oleh TB. Temuan ini logis karena label kelas dibentuk berdasarkan IMT yang dihitung dari BB dan TB. Usia dan jenis kelamin tetap memberikan kontribusi tambahan, tetapi pengaruhnya relatif kecil dibandingkan BB dan TB. Dengan demikian, penambahan fitur demografis dapat dipertahankan untuk memperkaya representasi data, tetapi interpretasi model harus tetap mempertimbangkan bahwa ground truth dibentuk dari IMT.

Tabel 12. Validasi Data Baru

Komponen	Nilai
Jenis kelamin	Laki-laki
Usia	61 tahun
Berat badan	80 kg
Tinggi badan	1,60 m
IMT	31,25
Kategori manual	Obesitas II
Prediksi sistem	Obesitas II

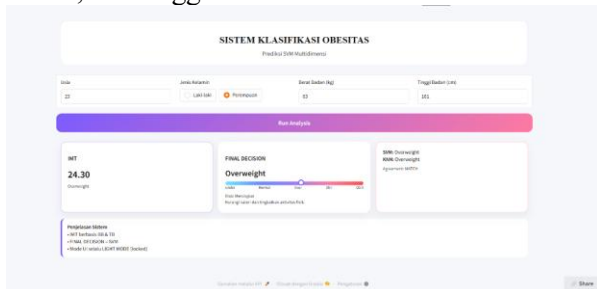
Validasi dunia nyata dilakukan menggunakan data pasien baru laki-laki berusia 61 tahun dengan berat badan 80 kg dan tinggi badan 1,60 m. Nilai IMT pasien tersebut adalah 31,25 dan termasuk kategori Obesitas II berdasarkan klasifikasi Asia-Pasifik. Sistem menghasilkan prediksi yang konsisten pada kategori Obesitas II. Hasil ini menunjukkan bahwa model dapat menerima data baru dan menghasilkan prediksi sesuai format fitur yang digunakan pada tahap pelatihan.

4.6. Deployment

Tahap deployment dilakukan dengan mengimplementasikan model klasifikasi tingkat obesitas ke dalam aplikasi web sederhana

menggunakan Gradio dan Hugging Face Spaces. Model yang telah dilatih disimpan menggunakan Joblib agar dapat dimuat kembali tanpa melakukan proses pelatihan ulang. Aplikasi menerima input berupa usia, jenis kelamin, berat badan, dan tinggi badan. Data tersebut kemudian diproses oleh model SVM dan KNN untuk menghasilkan prediksi kategori tingkat obesitas.

Sistem juga menghitung nilai IMT sebagai informasi pendukung untuk membantu interpretasi hasil. Namun, IMT tidak digunakan sebagai fitur input model, karena fitur utama yang diproses oleh model tetap terdiri atas usia, jenis kelamin, berat badan, dan tinggi badan.



Gambar 3. Tampilan Hasil Prediksi Aplikasi Klasifikasi Tingkat Obesitas.

Berdasarkan Gambar 3, aplikasi menampilkan hasil prediksi berupa nilai IMT, kategori prediksi SVM, kategori prediksi KNN, interpretasi kondisi tubuh, serta posisi kategori pada skala klasifikasi obesitas. Pengujian sistem menggunakan input usia 23 tahun, jenis kelamin perempuan, berat badan 63 kg, dan tinggi badan 161 cm menghasilkan IMT sebesar 24,30. Hasil prediksi SVM dan KNN menunjukkan kategori yang sama, yaitu Overweight.

Hasil tersebut menunjukkan bahwa model berhasil diimplementasikan ke dalam aplikasi web dan dapat menghasilkan prediksi secara real-time berdasarkan data antropometri pengguna. Sistem ini dapat digunakan sebagai alat bantu skrining awal tingkat obesitas, tetapi tidak menggantikan diagnosis atau pemeriksaan langsung oleh tenaga kesehatan.

4.7. Pembahasan Temuan

Temuan utama penelitian ini menunjukkan bahwa SVM lebih unggul dibandingkan KNN pada klasifikasi tingkat obesitas berbasis data antropometri. Keunggulan tersebut terlihat dari akurasi, macro F1-score, cross validation, dan hasil paired t-test. Secara metodologis, hasil ini dapat dijelaskan dari kemampuan SVM dengan kernel RBF dalam membentuk batas keputusan nonlinier, sehingga lebih stabil pada kelas yang memiliki rentang antropometri berdekatan. Sebaliknya, KNN lebih rentan mengalami kesalahan pada area transisi antar kelas karena proses klasifikasinya bergantung pada kedekatan jarak antar data.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa SVM memiliki performa kompetitif pada klasifikasi obesitas multikelas [3] dan klasifikasi kesehatan berbasis data multidimensi [7]. Perbedaan penting penelitian ini terletak pada penggunaan IMT sebagai dasar pembentukan label kelas, bukan sebagai fitur input model. Hasil permutation importance tetap menunjukkan bahwa BB dan TB menjadi fitur dominan karena keduanya merupakan komponen utama pembentuk IMT. Dengan demikian, hasil model perlu diinterpretasikan sebagai klasifikasi kategori obesitas berbasis label IMT, bukan sebagai diagnosis klinis obesitas secara menyeluruh.

Implikasi penelitian ini adalah bahwa SVM dapat digunakan sebagai dasar pengembangan sistem skrining awal tingkat obesitas berbasis data antropometri. Namun, penelitian ini masih terbatas pada variabel usia, jenis kelamin, BB, dan TB serta data dari satu wilayah layanan kesehatan. Penelitian selanjutnya perlu menambahkan variabel klinis dan perilaku, seperti lingkaran pinggang, pola makan, aktivitas fisik, riwayat penyakit, dan riwayat keluarga, serta melakukan pengujian pada data dari wilayah lain.

5. KESIMPULAN

- SVM menunjukkan kinerja sangat baik dalam klasifikasi tingkat obesitas dengan akurasi 97,75%, macro F1-score 97,75%, dan weighted F1-score 0,98.
- KNN memperoleh akurasi 91,93%. Nilai ini baik, tetapi masih di bawah SVM pada data uji dan cross validation.
- Cross validation menunjukkan SVM lebih stabil dengan rata-rata akurasi 98,31%, sedangkan KNN memperoleh 93,50%. Paired t-test menghasilkan p-value 0,000442, sehingga perbedaan performa signifikan.
- Penambahan usia dan jenis kelamin meningkatkan akurasi SVM dari 97,00% menjadi 97,75%. Kontribusinya kecil, sedangkan berat badan dan tinggi badan tetap menjadi fitur paling dominan.
- Model SVM berhasil diimplementasikan ke aplikasi web berbasis Gradio dan Hugging Face Spaces sehingga dapat digunakan untuk prediksi real-time berbasis data antropometri.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Puskesmas Kota Sukabumi yang telah menyediakan data penelitian, Program Studi Teknik Informatika Universitas Muhammadiyah Sukabumi, Fakultas Sains dan Teknologi, serta dosen pembimbing yang telah memberikan arahan dalam penyusunan penelitian ini.

DAFTAR PUSTAKA

- [1] "WHO," who.int, 8 Desember 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>. [Accessed 23 Mei 2026].
- [2] Kementerian Kesehatan Republik Indonesia, "Keputusan Menteri Kesehatan Republik Indonesia Nomor HK.01.07/MENKES/509/2025 tentang Pedoman Nasional Pelayanan Klinis Tata Laksana Obesitas Dewasa," Kementerian Kesehatan Republik Indonesia, Jakarta, 2025.
- [3] S. A. Utiahman and A. M. M. Pratama, "Analisis Perbandingan KNN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 3137-3146, 2024.
- [4] G. Sugisandhea and T. Handhayani, "PERBANDINGAN KINERJA KNN, SVM, DAN ANN UNTUK MEMPREDIKSI LEVEL OBESITAS," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 13, no. 2, pp. 1-9, 2025.
- [5] G. Airlangga, "Machine Learning-Based Obesity Classification: A Comparative Study Using Self-Reported Survey Data and Ensemble Learning Models," *Jurnal Teknologi Informatika dan Komputer MH. Thamrin*, vol. 11, no. 1, pp. 347-361, 2025.
- [6] K. N. A. Nur, A. Wanto and P. Poningsih, "OPTIMIZATION OF SVM ALGORITHM FOR OBESITY CLASSIFICATION WITH SMOTE TECHNIQUE AND HYPERPARAMETER TUNING," *JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)*, vol. 11, no. 2, pp. 291-303, 2025.
- [7] E. E. Pratiwi, A. R. Aisy, R. Rahmaddeni and N. Ananta, "KLASIFIKASI KESEHATAN MENTAL PADA USIA REMAJA MENGGUNAKAN METODE SVM," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 2, pp. 445-453, 2025.
- [8] L. Hakim, A. Sobri, L. Sunardi and D. Nurdiansyah, "Prediksi penyakit jantung berbasis mesin learning dengan menggunakan metode k-nn," *JURNAL DIGITAL TEKNOLOGI INFORMASI*, vol. 7, no. 2, pp. 14-20, 2024.
- [9] T. Hidayatulloh and L. Yusuf, "KLASIFIKASI TIPE BERAT TUBUH MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *INTI NUSA MANDIRI*, vol. 18, no. 1, pp. 71-77, 2023.
- [10] A. P. Siregar and S. I. Al Idrus, "Implementasi Algoritma K-Nearest Neighbors (KNN) dalam Deteksi Dini Hipertensi berdasarkan Analisis Tekanan Darah," *INCODING: Journal of Informatic and Computer Science Engineering*, vol. 5, no. 2, pp. 305-318, 2025.
- [11] A. Rianti, N. W. A. Majid and A. Fauzi, "CRISP-DM: Metodologi Proyek Data Science," in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, Surakarta, 2023.
- [12] W. A. I. Pradana, S. and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, pp. 239 - 248, 2024.
- [13] I. Maulana, A. M. Siregar, R. and A. Fauzi, "OPTIMIZATION OF MACHINE LEARNING MODEL ACCURACY FOR BRAIN TUMOR CLASSIFICATION WITH PRINCIPAL COMPONENT ANALYSIS," *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 3, pp. 903-915, 2024.
- [14] N. Cahyani and R. Irsyada, "Performance Comparison of SelectKBest and Permutation Importance in Feature Selection for Diabetes Prediction," *Brilliance*, vol. 5, no. 1, pp. 529-541, 2025.
- [15] A. D. K. Dewi and D. W. Utomo, "Optimasi Decision Tree menggunakan Pendekatan Ensemble Bagging untuk Klasifikasi Klinis Hipertensi," *EDUMATIC: Jurnal Pendidikan Informatika*, vol. 9, no. 3, pp. 728-747, 2025.
- [16] S. P. J. Simanjuntak, R. A. Aziz, I. A. Prasetyo, M. R. Al Ikhsan, A. Cio, S. Wibawa and M. Muharrom, "PERANCANGAN SISTEM PREDIKSI RISIKO DIABETES MENGGUNAKAN LOGISTIC REGRESSION DAN RANDOM FOREST," *INFOTECH journal*, vol. 11, no. 2, pp. 433-441, 2025.
- [17] A. R. Syahputra, R. Hidayat, F. Rismansyah, S. Sumanto, I. Budiawan and R. Pakpahan, "Komparasi Algoritma Machine Learning (SVM, Random Forest, dan Regresi Logistik) untuk Prediksi Tingkat Obesitas," *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, vol. 5, no. 3, pp. 660-673, 2025.