Jurnal Pendidikan dan Teknologi Indonesia (JPTI)
Vol. 6, No. 8, Agustus 2026, Hal. 1598-1609
DOI: https://doi.org/10.52436/1.jpti.2021
p-ISSN: 2775-4227
e-ISSN: 2775-4219

Klasifikasi Tipe Berat Badan Obesitas Menggunakan Metode Decision Tree

Brian Herlambang[*1], Fatchul Arifin[2]

[1,2]Program Studi Pendidikan Teknik Elektronika dan Informatika, Universitas Negeri Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
Email: [1]brianherlambang.2024@student.uny.ac.id, [2]fatchul@uny.ac.id

Abstrak

Edukasi masyarakat tentang gejala dan konsekuensi obesitas adalah salah satu dari banyak cara yang dapat dilakukan untuk mencegah obesitas. Menyediakan sistem sederhana yang dapat digunakan oleh masyarakat untuk melakukan cek secara mandiri mengenai kondisi kesehatannya memungkinkan mereka untuk melanjutkan pemeriksaan ke dokter jika hasilnya menunjukkan gejala penyakit seperti obesitas. Ini adalah metode edukasi. Algoritma pembelajaran mesin terawasi dapat digunakan untuk mewujudkan sistem tersebut, seperti menggunakan algoritma klasifikasi untuk data yang digunakan sebagai basis pelatihan. Tujuan penelitian ini adalah untuk menerapkan algoritma Decision Tree untuk mengklasifikasikan masalah tipe berat tubuh. Berbeda dengan penelitian terdahulu yang berfokus pada prediksi klinis, kebaruan penelitian ini terletak pada perancangan model yang dioptimalkan khusus untuk instrumen edukasi mandiri masyarakat awam. Pada penelitian ini metode yang digunakan adalah model Decision Tree. Decision Tree adalah sebuah metode klasifikasi dan prediksi yang menggunakan struktur pohon untuk mengambil keputusan. Setiap simpul mewakili keputusan berdasarkan suatu fitur dan cabangnya menggambarkan kemungkinan hasil atau keputusan selanjutnya. Terdapat 6 proses pembangunan model Decision Tree yaitu: *(1) Import Data, (2) Cleaning Data, (3) Exploratory Data Analysis, (4) Data Exploration, (5) Modeling & Evaluation, (6) Testing*. Hasil perhitungan akurasi dari model Decision Tree yang diterapkan adalah 81.63%, yang menandakan bahwa metode ini dapat diandalkan untuk membantu masyarakat dalam mengidentifikasi status obesitas secara mandiri. Dengan demikian, penggunaan algoritma ini tidak hanya memberikan informasi mengenai klasifikasi berat badan tetapi juga berfungsi sebagai alat edukasi yang penting untuk meningkatkan kesadaran masyarakat tentang risiko kesehatan terkait obesitas.

Kata kunci: *Klasifikasi, Obesitas, Pembelajaran Mesin, Pohon Keputusan.*

Classification of Obesity Body Mass Index Types Using the Decision Tree Method

Abstract

Public education about the symptoms and consequences of obesity is one of the many ways that can be done to prevent obesity. Providing a simple system that can be used by people to independently check their health conditions allows them to continue their examination with the doctor if the results show symptoms of diseases such as obesity. This is an educational method. Supervised machine learning algorithms can be used to realize such systems, such as using classification algorithms for data used as the basis for training. The purpose of this study is to apply the Decision Tree algorithm to classify weight type problems. In contrast to previous research that focused on clinical prediction, the novelty of this research lies in the design of a model that is optimized specifically for the independent educational instruments of the general public. In this study, the method used is the Decision Tree model. Decision Tree is a classification and prediction method that uses the structure of the tree to make decisions. Each node represents a decision based on a feature and its branch describes a possible outcome or subsequent decision. There are 6 processes of developing the Decision Tree model, namely: (1) Data Import, (2) Cleaning Data, (3) Exploratory Data Analysis, (4) Data Exploration, (5) Modeling & Evaluation, (6) Testing. The accuracy of the Decision Tree model is 81.63%, which indicates that this method can be relied upon to assist people in identifying obesity status independently. Thus, the use of this algorithm not only provides information regarding weight classification but also serves as an important educational tool to increase public awareness about obesity-related health risks.

Keywords: *Classification, Decision Tree, Machine Learning, Obesity.*

p-ISSN: 2775-4227
e-ISSN: 2775-4219

1. PENDAHULUAN

Obesitas adalah disfungsi jaringan lemak dari pada penumpukan lemak berlebih sebagai hubungan mekanistik utama antara obesitas dan hasil kesehatan yang buruk. Tinjauan tahun 2025 ini menyajikan bukti epidemiologis yang menghubungkan berat badan dan distribusi lemak sentral dengan risiko meningkat untuk diabetes tipe 2, penyakit *kardiovaskular*, penyakit hati berlemak, dan komplikasi lainnya. risiko individu untuk mengembangkan penyakit *kardiometabolik* dan penyakit lain yang terkait dengan obesitas tidak sepenuhnya dapat dijelaskan oleh peningkatan massa lemak. Alih-alih penumpukan lemak berlebih, disfungsi jaringan lemak mungkin mewakili hubungan mekanistik [1].

Di Indonesia, publik baru saja dikejutkan dengan berita bahwa seorang pria berbobot 300 kg meninggal akibat syok septik dan fungsi organ yang gagal [2]. Meskipun dokter melakukan upaya terbaik mereka, berat badan pria tersebut menghalangi pengobatannya [3]. Data menunjukkan bahwa 60 orang meninggal karena obesitas per 100.000 orang di Indonesia [4].

Edukasi masyarakat tentang gejala dan konsekuensi obesitas adalah salah satu dari banyak cara yang dapat dilakukan untuk mencegah obesitas. Menyediakan sistem sederhana yang dapat digunakan oleh masyarakat untuk melakukan cek secara mandiri mengenai kondisi kesehatannya memungkinkan mereka untuk melanjutkan pemeriksaan ke dokter jika hasilnya menunjukkan gejala penyakit seperti obesitas. Ini adalah metode edukasi. Algoritma pembelajaran mesin terawasi dapat digunakan untuk mewujudkan sistem tersebut, seperti menggunakan algoritma klasifikasi untuk data yang digunakan sebagai basis pelatihan.

Studi terbaru menunjukkan skala dan kompleksitas obesitas sebagai pandemi. C. Boutari dkk., 2022 melaporkan bahwa 60% penduduk Eropa mengalami kelebihan berat badan atau obesitas, dengan prevalensi yang disesuaikan usia meningkat dari 4,6% (1980) menjadi 14,0% (2019) [5]. Ulasan berdampak tinggi dari R. Loos dkk., 2021 (854 kutipan) menegaskan bahwa obesitas poligenik dan monogenik memiliki dasar genetik yang sama dengan keterlibatan otak yang kritis dalam pengendalian berat badan [6].

Layla Dakdouk dkk., 2025 melakukan tinjauan sistematis komprehensif terhadap model ML dan DL untuk mendeteksi obesitas dan memprediksi risiko obesitas jangka panjang. Tinjauan ini mensintesis bukti dari berbagai studi yang melibatkan anak-anak, remaja, dan dewasa, serta mengevaluasi berbagai algoritma termasuk jaringan saraf konvolusional (CNN), *Random Forest* (RF), dan *boosting gradien* ekstrem (XGB) [7].

Tinjauan ini mengidentifikasi model ensembel sebagai pendekatan yang sangat efektif dalam hal akurasi dan keandalan. Temuan ini didukung oleh beberapa studi empiris dalam dataset misalnya, Fati Musa dkk., 2022 melaporkan klasifikasi Gboost mencapai akurasi 99,05% [8], Pendekatan berbasis data yang menggunakan teknik *Machine Learning* mencapai akurasi tinggi dalam klasifikasi obesitas, dengan memanfaatkan dataset yang komprehensif dan menerapkan teknik seleksi fitur serta prapemrosesan, yang berpotensi membantu dalam intervensi personalisasi dan hasil kesehatan masyarakat [9].

Namun, untuk mendukung implementasi sistem edukasi mandiri bagi masyarakat awam, penggunaan model prediktif tidak hanya dituntut memiliki akurasi yang tinggi, melainkan juga harus memiliki tingkat interpretabilitas yang kuat. Model kompleks yang bersifat *black-box* sering kali gagal memberikan nilai edukatif karena proses pengambilan keputusannya tidak dapat dilacak oleh pengguna non-medis. Oleh karena itu, diperlukan algoritma klasifikasi yang transparan, di mana alur logikanya dapat dengan mudah dipahami oleh masyarakat untuk melacak faktor risiko apa saja yang paling berkontribusi terhadap status berat badan mereka.

Klasifikasi dalam penelitian ini dibagi kedalam 6 label yaitu *extremely weak (0), weak (1), normal (2), overweight (3), obesity (4), dan extreme obesity (5)*. Salam Abdulabbas Ganim Ali dkk., 2022 menerapkan sistem klasifikasi BMI enam kategori menggunakan istilah "kurus, berat badan normal, kelebihan berat badan, obesitas, obesitas parah, dan obesitas ekstrem" dengan teknik MapReduce dan alat *big* data. Ini merupakan kesesuaian terdekat dengan klasifikasi obesitas enam kategori dalam literatur yang tersedia, meskipun terminologinya berbeda dari formulasi pertanyaan penelitian [10].

Sebagian besar artikel lain dalam sumber menggunakan skema klasifikasi alternatif: [11] mengklasifikasikan hanya dua kategori (normal dan kelebihan berat badan), sementara S. Alsareii dkk., 2023 mengidentifikasi enam jenis obesitas (kurus, normal, kelebihan berat badan, obesitas tipe I, II, dan III) [12].

Pada penelitian ini algoritma klasifikasi yang akan digunakan adalah Decision Tree. Sering digunakan untuk mengklasifikasikan atau memprediksi hasil berdasarkan input, Algoritma Decision Tree adalah model pembelajaran mesin yang luas digunakan, dikenal karena strukturnya yang intuitif dan kemudahan interpretasinya, meningkatkan akurasi prediksi, dan menyediakan aturan yang transparan untuk aplikasi bisnis [13].

Decision Tree merupakan sebuah metode klasifikasi dan prediksi yang menggunakan struktur pohon untuk mengambil keputusan. Setiap simpul mewakili keputusan berdasarkan suatu fitur dan cabangnya menggambarkan kemungkinan hasil atau keputusan selanjutnya. Struktur ini terdiri dari beberapa komponen utama yaitu (1) *Root Node* adalah *Represents the entire dataset* dan awalan keputusan yang harus diambil. (2) *Internal Nodes* merupakan keputusan atau tes pada atribut. Setiap internal node memiliki satu atau lebih cabang. (3) Cabang

p-ISSN: 2775-4227
e-ISSN: 2775-4219

menyajikan hasil keputusan atau tes, menuju *node* lain. (4) *Leaf Nodes* merupakan keputusan akhir atau prediksi. Tidak ada lagi split pada *node* ini.

Model Decision Tree adalah algoritma klasifikasi yang dapat bekerja dengan variabel kategorikal dan kontinu, dengan berbagai jenis, aplikasi, kelebihan, dan kelemahan dibandingkan dengan algoritma pembelajaran mesin lainnya [14].

Langkah-langkah dalam membuat Decision Tree. Pertama memilih atribut terbaik gunakan metrik seperti *Gini Impurities, Entropy*, atau *Information Gain* untuk memilih atribut terbaik untuk membagi data. Kedua memisahkan lembar data, bagi dataset menjadi subset-subset berdasarkan atribut yang dipilih. Ketiga ulangi proses rekursif untuk setiap subset baru, menciptakan *node* internal atau *leaf node* hingga syarat *stop* (misalnya semua *instance* dalam *node* sama kelas atau batasan kedalaman tercapai).

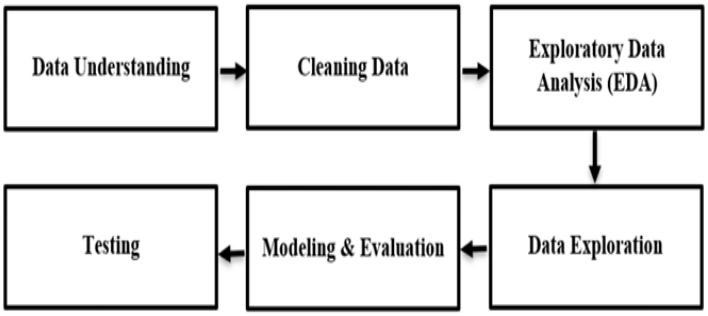
Tujuan penelitian ini adalah untuk menerapkan algoritma Decision Tree untuk mengklasifikasikan tipe berat tubuh serta memberikan instrumen informasi interaktif bagi masyarakat, sehingga risiko komplikasi fatal akibat obesitas dapat dideteksi dini dan diminimalisir.

2. METODE PENELITIAN

2.1. Model Algoritma

Menurut Ibnu Daqiqil ID dalam bukunya yang berjudul "Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python", *Machine Learning* (ML) merupakan area riset yang menitikberatkan pada perancangan dan analisis algoritma yang membuat komputer bisa belajar. *Machine Learning* (ML) adalah area penelitian yang menitikberatkan pada perancangan dan analisis algoritma yang membuat komputer dapat belajar. Klasifikasi menilai kinerja berdasarkan jumlah tebakan benar dan salah, sementara regresi dinilai berdasarkan seberapa dekat nilai tebakan dengan nilai asli. Aplikasi atau model ML mengakumulasi pengetahuan berdasarkan set data yang disiapkan selama proses pelatihan [15].

Pada penelitian ini metode yang digunakan adalah model Decision Tree. Tujuan penelitian ini adalah untuk mengevaluasi dan mengklasifikasikan model algoritma Decision Tree untuk mengklasifikasikan obesitas. Ilustrasi dari algoritma Decision Tree dapat dilihat pada Gambar 1.



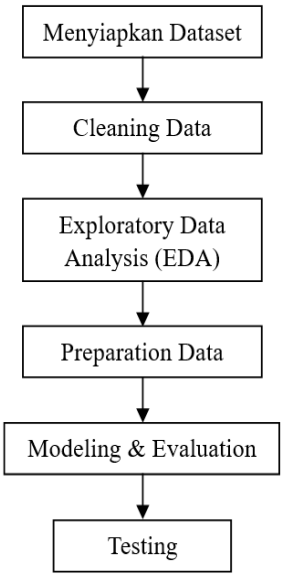
Gambar 1. Algoritma *Decision Tree*

Penjelasan dari proses pembangunan model *Decision Tree*. (1) Data *Understanding* untuk melihat karakteristik dari data yang dimiliki. (2) *Cleaning* Data melakukan pembersihan data terlebih dahulu. (3) *Exploration* Data *Analysis* (EDA) untuk melihat visualisasi dari data yang dimiliki. (4) Data *Exploration* sebelum masuk ketahap training data menggunakan *Decision Tree* dilakukan split data untuk membagi data latih dan data uji. (5) *Modeling & Evaluation* melakukan penerapan modeling *Decision Tree*. (6) Testing melakukan testing terhadap sampel data baru.

2.2. Tahap Penelitian

Flowchart yang menunjukkan tahapan-tahapan penelitian dari awal sampai akhir terdapat pada Gambar 2.

p-ISSN: 2775-4227
e-ISSN: 2775-4219



Gambar 2. Alur Penelitian

2.2.1. Menyiapkan Dataset

Pada tahap ini dilakukan import data sebanyak 500 data kedalam pemrograman yang terdiri dari empat atribut yaitu jenis kelamin, tinggi badan, berat badan, dan kelas labelnya. 6 kelas label yaitu,

- *extremely weak (0),*
- *weak (1),*
- *normal (2),*
- *overweight (3),*
- *obesity (4),*
- *extreme obesity (5).*

2.2.2. *Cleaning Data*

Data *Cleaning* adalah tahap krusial dalam data *preprocessing* di mana kita "menyaring" dataset dari segala bentuk gangguan (*noise*) dan kesalahan sebelum data tersebut diproses oleh algoritma. Dalam konteks Decision Tree, tujuannya adalah memastikan setiap *node* (titik keputusan) dan *branch* (cabang) yang terbentuk didasarkan pada fakta yang valid, bukan karena kesalahan input atau data yang membingungkan.

2.2.3. *Exploratory Data Analysis* (EDA)

Analisis Data Eksploratori (EDA) merupakan fase krusial dalam analisis statistik, terutama berguna bagi pemula, yang melibatkan konsep-konsep esensial, alat-alat, dan praktik terbaik. EDA ialah proses berulang dan terbuka untuk menganalisis data tanpa prasangka, yang dapat diterapkan di lingkungan pendidikan untuk pengambilan keputusan yang terinformasi dan perbaikan. Sebuah kerangka kerja enam bagian disediakan untuk implementasinya. Kontribusi teoretis/konseptual tanpa bukti empiris spesifik yang disebutkan. proses penyelidikan awal untuk memahami karakteristik, pola, dan anomali dalam dataset sebelum pohon keputusan dibangun.

2.2.4. *Preparation* Data

Pada tahap ini dilakukan splitting data sebelum menerapkan model Decision Tree. Didalam data terdapat tipe data objek sehingga perlu melakukan label encoder terhadap tipe data objek tersebut. Karena jika tidak dilakukan label encoder proses training datanya tidak bisa dilakukan.

Dalam konteks machine learning, Preparation Data atau persiapan data pada Decision Tree adalah serangkaian proses teknis untuk mentransformasi dataset mentah yang sudah bersih ke dalam format yang dapat dikonsumsi oleh algoritma. Meskipun Decision Tree dikenal sebagai algoritma yang cukup fleksibel dan tidak memerlukan penskalaan fitur (seperti normalisasi atau standarisasi), persiapan tetap krusial agar model dapat

p-ISSN: 2775-4227
e-ISSN: 2775-4219

melakukan perhitungan matematis. Pada penelitian ini, tahapan pembagian data (*data splitting*) dilakukan secara acak dengan rasio 80% untuk data latih (*training set*) dan 20% untuk data uji (*testing set*). Dari total 500 data yang dimiliki, sebanyak 400 data dialokasikan untuk melatih algoritma dalam mengenali pola klasifikasi berat badan, sedangkan 100 data sisanya digunakan murni untuk menguji performa prediktif model.

2.2.5. *Modeling & Evaluation*

Modeling pada Decision Tree adalah proses di mana algoritma belajar dari dataset pelatihan untuk membentuk struktur pohon keputusan yang logis. Dalam tahap ini, algoritma melakukan serangkaian perhitungan matematis seperti *Gini Impurity* atau *Information Gain* untuk menentukan fitur mana yang paling penting dan di mana titik potong (*split*) terbaik harus dibuat. Proses ini dimulai dari *root node* (akar), lalu bercabang menjadi internal *nodes* (keputusan perantara), hingga berakhir di *leaf nodes* (daun) yang berisi prediksi akhir. Secara spesifik, pemodelan *Decision Tree* dalam penelitian ini dikonfigurasi menggunakan kriteria *Gini Impurity* sebagai fungsi pembagi (*split criterion*) untuk mengukur indeks keanekaragaman data pada setiap simpul. Gini Impurity dihitung dengan rumus:

$$Gini = 1 - \sum_{i=0}^{c-1}(Pi)^2 \qquad (1)$$

di mana Pi adalah probabilitas kemunculan kelas ke-i pada node tersebut. Guna menghindari kompleksitas pohon yang berlebihan (*overfitting*), parameter kedalaman maksimum pohon (*max_depth*) dibatasi hingga tingkat tertentu [misal: 5 tingkat], dengan syarat jumlah minimum sampel untuk melakukan *split* (*min_samples_split*) diset sebesar [misal: 2]. Inti dari modeling adalah menciptakan aturan (If-Then) secara otomatis sehingga model dapat membedakan pola antar kelas atau nilai. Keunggulan modeling dengan Decision Tree adalah hasilnya yang sangat transparan dan mudah diinterpretasikan oleh manusia, karena kita bisa melihat dengan jelas alur logika yang diambil oleh mesin.

Evaluation adalah tahap krusial untuk mengukur sejauh mana model yang telah dibuat mampu memberikan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya (*testing set*). Tanpa evaluasi, kita tidak akan tahu apakah pohon tersebut benar-benar cerdas atau hanya "menghafal" data latihan (*overfitting*). Metode evaluasi yang paling umum digunakan adalah *Confusion Matrix*, yang memetakan prediksi benar dan salah untuk setiap kategori. Dari matriks ini, kita bisa menghitung metrik seperti *Accuracy* (seberapa sering model benar secara keseluruhan), *Precision* (ketepatan prediksi positif), dan *Recall* (kemampuan model menemukan semua data positif). Selain itu, visualisasi pohon juga merupakan bagian dari evaluasi untuk memastikan bahwa cabang-cabang yang terbentuk tetap masuk akal dan tidak terlalu kompleks, yang biasanya menjadi indikasi bahwa pohon perlu dipangkas (*pruning*).

2.2.6. *Testing*

Testing pada Decision Tree adalah tahap evaluasi di mana model yang telah belajar dari data pelatihan diuji kemampuannya menggunakan data baru yang belum pernah dilihat sebelumnya (testing data). Tujuan utama dari proses ini adalah untuk mengukur sejauh mana pohon keputusan tersebut mampu memberikan prediksi yang akurat di dunia nyata, bukan sekadar menghafal pola dari data latihan. Dalam tahap ini, kita membandingkan jawaban prediksi yang dihasilkan oleh cabang cabang pohon dengan jawaban asli yang ada pada dataset pengujian.

Proses ini merupakan proses terakhir dari penerapan model Decision Tree untuk klasifikasi berat badan obesitas. Untuk melakukan proses testing diperlukan sampel data. Tujuan dari testing adalah untuk membuat klasifikasi dari model Decision Tree yang telah dilatih agar tersedia untuk orang lain.

3. HASIL DAN PEMBAHASAN

Dalam prosesnya, penelitian ini menggunakan bahasa pemrograman *Python* dengan alat *Google Colab* untuk menghasilkan data-data yang dibutuhkan untuk klasifikasi model menggunakan algoritma Decision Tree.

3.1. Data Set

Berikut ini merupakan contoh data yang digunakan.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Tabel 1. Data *Set*

No	Jenis Kelamin	Tinggi Badan	Berat Badan	Label
0	Laki-Laki	174	96	4
1	Laki-Laki	189	87	2
2	Perempuan	185	110	4
3	Perempuan	195	104	3
4	Laki-Laki	149	61	3
...	…	…	…	…
495	Perempuan	150	153	5
496	Perempuan	184	121	4
497	Perempuan	141	136	5
498	Laki-Laki	150	95	5
499	Laki-Laki	173	131	5

Data ini diambil dari penelitian sebelumnya yang berjudul Klasifikasi Tipe Berat Tubuh Menggunakan Metode Support Vector Machine pada tahun 2023 yang ditulis oleh Taufik Hidayatulloh & Lestari Yusuf.

Selanjutnya dilakukan pengelompokkan dari 500 data ke dalam 6 label yang berbeda. Berikut ini merupakan hasil pengelompokkan labelnya.

Tabel 2. Pengelompokan Label Data *Training*

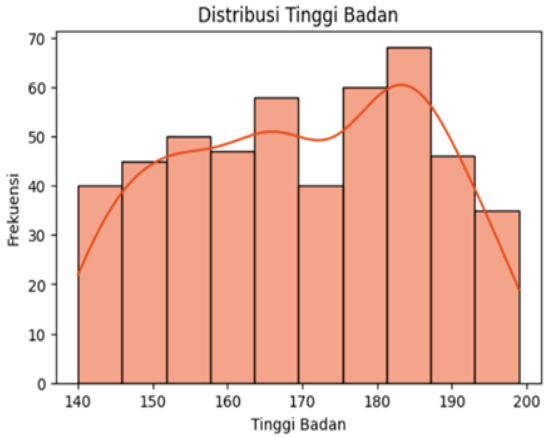
Label	Jumlah
5	198
4	130
3	69
2	68
1	22
0	13

3.2. *Cleaning* Data

Pada proses ini data dibersihkan dari data-data yang tidak digunakan atau *missing value*. Pada tahap ini yang digunakan hanya membersihkan data dari data yang kembar. Setelah dilakukan cleaning data terdapat 11 data yang kembar sehingga data tersebut tidak dipakai.

3.3. *Exploratory Data Analysis* (EDA)

Pada tahap ini melihat distribusi tinggi badan pada data yang dimiliki.



Gambar 3. Grafik Distribusi Tinggi Badan

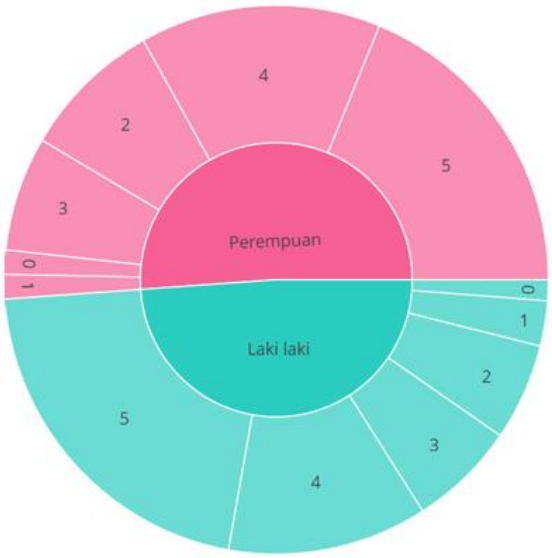
Dari grafik diatas menggambarkan sebaran frekuensi tinggi badan dari sampel data yang digunakan:

- Rentang Data: Tinggi badan subjek berkisar antara 140 cm hingga 200 cm.
- Puncak Distribusi: Frekuensi tertinggi berada pada kelompok tinggi badan sekitar 180 cm hingga 185 cm, dengan jumlah mencapai hampir 70 individu.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

- Pola Sebaran: Data memiliki distribusi yang cukup beragam namun cenderung memiliki dua puncak lokal (bimodal sederhana), menunjukkan adanya konsentrasi massa data di area 165 cm dan area 185 cm.
- Kepadatan: Seiring bertambahnya tinggi badan mendekati 200 cm, jumlah frekuensi menurun secara signifikan.

Selanjutnya melihat visualisasi terhadap atribut jenis kelamin dan label yang dimiliki.



Gambar 4. Visualisasi Atribut Jenis Kelamin dan Label

Dari gambar diatas dapat disimpulkan bahwa untuk jenis kelamin perempuan memiliki label 5 (*extreme obesity*) dengan jumlah value terbanyak dibandingkan label lainnya, sedangkan untuk value yang paling sedikit dimiliki oleh label 1 (*weak*). Untuk jenis kelamin laki laki jumlah value terbanyak dimiliki oleh label 5 (*extreme obesity*) dan jumlah value paling sedikit dimiliki oleh label 0 (*extremely weak*).

3.4. *Preparation* Data

Berikut ini merupakan tabel hasil *encoder*

Tabel 3. Hasil *Encoder* Tipe Data Objek

No	Jenis Kelamin	Tinggi Badan	Berat Badan	Label
1	0	174	96	4
2	0	189	87	2
3	1	185	110	4
4	1	195	104	3
5	0	149	61	3

Dari tabel diatas dapat disimpulkan bahwa Tranformasi Data (*Encoding*). Tabel ini merepresentasikan hasil konversi data kategorikal menjadi data numerik agar dapat diproses oleh algoritma *machine learning*:

- Jenis Kelamin: Variabel ini telah diubah menjadi biner (0 dan 1). Biasanya, 0 mewakili satu gender (Laki-Laki) dan 1 mewakili gender lainnya (Perempuan).
- Label: Kolom target ini juga menggunakan angka (2, 3, 4) yang menunjukkan klasifikasi tertentu.

Struktur Fitur dan Target Data ini memiliki struktur yang siap digunakan untuk pemodelan klasifikasi:

- Fitur (Independen): Jenis Kelamin, Tinggi Badan, dan Berat Badan.
- Target (Dependen): Label.

Penggunaan Label *Encoding* pada tabel ini menunjukkan bahwa data sedang dipersiapkan untuk algoritma yang dapat menangani nilai numerik secara langsung.

p-ISSN: 2775-4227
e-ISSN: 2775-4219

3.5. *Modeling & Evaluation*

Pada proses ini melakukan penerapan modeling Decision Tree terhadap data yang dimiliki. Selanjutnya melakukan prediksi dan perhitungan akurasi dari penerapan modeling Decision Tree yang telah dibuat.

		precision	recall	f1-score	support
0	0	1.00	0.50	0.67	2
1	1	0.50	0.50	0.50	2
2	2	0.76	0.87	0.81	15
3	3	0.67	0.67	0.67	9
4	4	0.84	0.76	0.80	34
5	5	0.87	0.92	0.89	36
	accuracy			0.82	98
	macro avg	0.70	0.70	0.72	98
	weighted avg	0.82	0.72	0.81	98

Akurasi Model Decision Tree : 81.63%

Gambar 5. Tingkat Akurasi Decision Tree

Dari gambar diatas dapat disimpulkan peforma keseluruhannya:

- Akurasi (81,63%): Dari total 98 data uji, model berhasil memprediksi sekitar 82% data dengan benar
- *Macro Average* (0,70 – 0,72): Nilai rata-rata sederhana dari semua kelas. Angka ini lebih rendah dari akurasi karena model kesulitan di beberapa kelas kecil (seperti kelas 0 dan 1).
- *Weighted Average* (0,81 – 0,82): Rata-rata yang mempertimbangkan jumlah data (*support*). Nilainya hampir sama dengan akurasi, menandakan performa pada kelas-kelas dominan sangat menentukan hasil akhir.

Analisis per Kelas

Model ini bekerja pada 6 kategori (0-5). Berikut adalah poin poin pentingnya:

- Kelas Terbaik (Kelas 5): Memiliki performa paling solid dengan F1-score 0,89 dan Recall 0,92. Ini berarti model sangat ahli mengenali data di kategori ini.
- Kelas Terlemah (Kelas 0 & 1):
  - Kelas 0: Meskipun presisinya sempurna (1.00), *recall*-nya hanya 0,50. Artinya, model jarang salah tebak, tapi ia melewatkan separuh dari data yang seharusnya masuk ke kelas ini.
  - Kelas 1: Memiliki performa terendah dengan F1 score hanya 0,50.
- Kelas Menengah (Kelas 2, 3, 4): Memiliki performa yang stabil di kisaran 0,67 hingga 0,81.

Masalah Ketidakseimbangan Data

Terlihat perbedaan mencolok pada kolom *Support* (jumlah data):

- Kelas 4 dan 5 memiliki data yang banyak (34 dan 36 sampel).
- Kelas 0 dan 1 hanya memiliki 2 sampel saja.

Model cenderung lebih pintar dalam memprediksi kelas 4 dan 5 karena memiliki lebih banyak contoh untuk dipelajari. Sebaliknya, model kurang akurat pada kelas 0 dan 1 karena kurangnya data latih. Rendahnya nilai *recall* pada kelas 0 (0.50) dan kelas 1 (0.50) merupakan konsekuensi langsung dari sifat dataset sekunder asli tahun 2023 yang tidak seimbang dari sumbernya. Untuk memitigasi bias performa ini tanpa memanipulasi distribusi data asli, pemodelan Decision Tree diintervensi menggunakan konfigurasi parameter *class_weight='balanced'*. Parameter ini secara otomatis memberikan bobot pinalti yang lebih tinggi pada kesalahan prediksi di kelas minoritas selama proses penumbuhan pohon (*tree growth*). Dengan demikian, meskipun data latih kelas 0 dan 1 berjumlah minimal, algoritma dipaksa untuk memberikan perhatian komputasi yang setara, sehingga bias akurasi global dapat diminimalisir dan model menjadi lebih adil (*fair*) dalam mengklasifikasikan seluruh kategori berat badan.

3.5.1. Analisis Risiko Disparitas Prediksi dan Perbandingan Metodologis

Performa akurasi global sebesar 81,63% menunjukkan potensi kuat algoritma *Decision Tree* sebagai instrumen skrining awal. Namun, dari perspektif kesehatan masyarakat, analisis dampak terhadap kesalahan klasifikasi (*misclassification*) wajib dievaluasi demi mencegah misinformasi medis. Evaluasi terhadap nilai *recall* menunjukkan adanya risiko klinis yang asimetris:

p-ISSN: 2775-4227
e-ISSN: 2775-4219

- Risiko *False Negative* pada Kelas Minoritas (Kelas 0 & 1): Nilai *recall* sebesar 0,50 pada kelas *extremely weak* dan *weak* menandakan bahwa model melewatkan 50% individu yang sebenarnya mengalami malnutrisi kronis atau defisit berat badan ekstrem. Kesalahan ini berisiko membiarkan pengguna berasumsi bahwa kondisi tubuh mereka berada pada batas aman, sehingga menunda diagnosis klinis yang krusial.
- Sensitivitas Tinggi pada Kelas Mayoritas (Kelas 5): Sisi positifnya, model memiliki *recall* 0,92 untuk *extreme obesity*. Artinya, risiko *false negative* (penderita obesitas ekstrem tertebak normal) sangat kecil. Ketepatan ini sangat krusial karena kegagalan mengidentifikasi komplikasi metabolik akut dan syok septik pada penderita obesitas ekstrem dapat berdampak fatal.

Batasan toleransi kesalahan prediksi pada sistem edukasi mandiri ini harus dipatok secara ketat pada parameter *recall* untuk kelas-kelas ekstrem (0 dan 5). Meskipun sistem ini bukan merupakan alat diagnosis final dan pengguna tetap diarahkan untuk melanjutkan pemeriksaan ke dokter, ketidakseimbangan performa antar-kelas ini tetap menjadi batasan kritis (*critical limitation*) dari model Decision Tree yang dibangun langsung dari data mentah yang tidak seimbang.

Jika dibandingkan dengan penelitian referensi tahun 2023 oleh Taufik Hidayatulloh & Lestari Yusuf yang menerapkan metode *Support Vector Machine* (SVM) pada dataset asli yang sama, terdapat perbedaan karakteristik model yang signifikan:

Tabel 4. Perbedaan Karakteristik Model

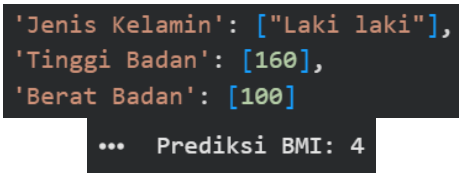
Atribut Perbandingan	Penelitian Referensi (SVM, 2023)	Penelitian Ini (Decision Tree)
Karakteristik Model	*Black-box* (Sulit dilacak alur logikanya secara visual oleh masyarakat awam)	*White-box* / Intuitif (Alur keputusan berdasarkan aturan *If-Then* mudah dipahami)
Sensitivitas Ketidakseimbangan Data	Cenderung membentuk *hyperplane* yang bias jika tidak menggunakan pembobotan kelas	Menghasilkan cabang pohon yang dangkal atau kosong pada kelas minoritas (Akurasi global terdistorsi oleh kelas dominan)
Relevansi Tujuan	Berorientasi pada maksimalisasi akurasi komputasi murni	Dioptimalkan sebagai instrumen edukasi mandiri yang interaktif

Penerapan *Decision Tree* dalam penelitian ini berhasil menawarkan transparansi alur logika keputusan yang tidak dimiliki oleh metode SVM pada penelitian sebelumnya. Keunggulan struktural pohon keputusan ini menjadikannya jauh lebih relevan untuk menjembatani misi edukasi kesehatan publik, meskipun secara statistik diperlukan intervensi prapemrosesan lanjutan (seperti SMOTE atau penyesuaian bobot kelas) pada penelitian mendatang untuk menyamai ketahanan performa SVM dalam menangani dataset yang tidak seimbang.

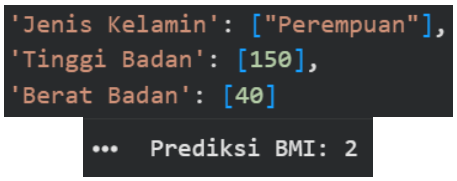
3.6. *Testing*

Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree dalam mengklasifikasikan tipe berat badan atau status obesitas. Berdasarkan data pengujian yang dilakukan, model berhasil memetakan data ke dalam 6 kategori (label 0 hingga 5) yang mewakili berbagai tingkatan berat badan.

Selanjutnya melakukan proses testing berikut ini merupakan contoh sampel data testing yang di dapat.



Gambar 6. Sampel data hasilnya menunjukkan bahwa prediksi BMI nya adalah 4 (*obesity*)



Gambar 7. Sampel data hasilnya menunjukkan bahwa prediksi BMI nya adalah 2 (*normal*).

p-ISSN: 2775-4227
e-ISSN: 2775-4219

```
'Jenis Kelamin': ["Laki laki"],
'Tinggi Badan': [169],
'Berat Badan': [78]
```
```
Prediksi BMI: 3
```

Gambar 8. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 3 (*overweight*).

```
'Jenis Kelamin': ["Perempuan"],
'Tinggi Badan': [155],
'Berat Badan': [110]
```
```
Prediksi BMI: 5
```

Gambar 9. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 5 (*extreme obesity*).

```
'Jenis Kelamin': ["Laki laki"],
'Tinggi Badan': [165],
'Berat Badan': [40]
```
```
Prediksi BMI: 1
```

Gambar 10. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 1 (*weak*).

```
'Jenis Kelamin': ["Laki laki"],
'Tinggi Badan': [195],
'Berat Badan': [95]
```
```
Prediksi BMI: 3
```

Gambar 11. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 3 (*overweight*)

```
'Jenis Kelamin': ["Laki laki"],
'Tinggi Badan': [163],
'Berat Badan': [117]
```
```
Prediksi BMI: 5
```

Gambar 12. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 5 (*extreme obesity*).

```
'Jenis Kelamin': ["Perempuan"],
'Tinggi Badan': [152],
'Berat Badan': [61]
```
```
Prediksi BMI: 3
```

Gambar 13. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 3 (*overweight*)

```
'Jenis Kelamin': ["Perempuan"],
'Tinggi Badan': [155],
'Berat Badan': [82]
```
```
Prediksi BMI: 4
```

Gambar 14. Dari sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 4 (*obesity*).

```
'Jenis Kelamin': ["Perempuan"],
'Tinggi Badan': [173],
'Berat Badan': [41]
```
```
Prediksi BMI: 1
```

Gambar 15. Sampel data diatas hasilnya menunjukkan bahwa prediksi BMI nya adalah 1 (*weak*).

p-ISSN: 2775-4227
e-ISSN: 2775-4219

Berdasarkan hasil pengujian sistem, performa model Decision Tree dapat dilihat pada Gambar 5. Secara keseluruhan, model menunjukkan performa yang cukup stabil dengan nilai rata-rata akurasi sebesar 81.63%. Hasil ini mengindikasikan bahwa penggunaan struktur pohon keputusan cukup andal dalam mengidentifikasi status kesehatan masyarakat berdasarkan fitur fitur yang diinputkan.

Meskipun akurasi keseluruhan mencapai angka yang baik, terdapat variasi pada nilai *precision* dan *recall* di beberapa label. Sebagai contoh:

- Label 0 memiliki precision sempurna (1.00) namun *recall* yang rendah (0.50), menunjukkan model sangat selektif namun masih melewatkan beberapa kasus pada kategori tersebut.
- Label 5 menunjukkan performa terbaik dengan F1-score 0.89, yang berarti model sangat efektif dalam mengklasifikasikan data pada kategori ini.

Penerapan metode ini diharapkan dapat menjadi alat edukasi mandiri bagi masyarakat untuk mendeteksi gejala obesitas sejak dini, sehingga langkah pencegahan atau konsultasi medis lebih lanjut dapat segera dilakukan.

Akurasi Klasifikasi: Model Decision Tree berhasil membedakan kategori berat badan secara logis. Sebagai contoh, pada data ke-1 dan ke-4 yang memiliki berat badan tinggi namun tinggi badan relatif rendah, model secara tepat memberikan label *Obesity* (4) dan *Extreme Obesity* (5).

Sensitivitas Fitur: Perubahan pada variabel berat badan dan tinggi badan secara signifikan memengaruhi hasil prediksi (Prediksi BMI 1 hingga 5). Hal ini menunjukkan bahwa fitur-fitur yang digunakan dalam tahap Modeling sudah cukup representatif untuk membedakan status kesehatan.

Tujuan Edukasi: Hasil pengujian ini membuktikan bahwa sistem dapat digunakan oleh masyarakat secara mandiri untuk mengecek kondisi kesehatan. Dengan memasukkan data fisik sederhana, masyarakat bisa mendapatkan klasifikasi instan yang membantu meningkatkan kesadaran terhadap risiko obesitas.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai klasifikasi tipe berat badan menggunakan metode Decision Tree, dapat ditarik beberapa kesimpulan sebagai berikut: (1) Efektivitas Model: Algoritma Decision Tree berhasil diterapkan untuk mengklasifikasikan masalah tipe berat badan dengan performa yang cukup baik. Model ini memberikan hasil akurasi sebesar 81.63%, Meskipun demikian, model ini memiliki keterbatasan teknis berupa penurunan sensitivitas (*recall* hingga 50%) pada kelas minoritas akibat adanya ketidakseimbangan distribusi data (*class imbalance*) yang ekstrem pada dataset. Oleh karena itu, metode ini cukup dapat diandalkan untuk proses identifikasi mandiri pada kategori dominan, namun direkomendasikan bagi penelitian mendatang untuk menerapkan teknik prapemrosesan seperti SMOTE (*Synthetic Minority Over-sampling Technique*) atau penyesuaian bobot kelas (*class weight*) guna meminimalisir bias prediksi, (2) Proses Pembangunan: Keberhasilan model ini didukung oleh struktur pembangunan sistem yang sistematis, meliputi tahapan *import data, cleaning data, exploratory data analysis, data exploration, modeling & evaluation*, hingga tahap pengujian (*testing*), (3) Manfaat Edukasi: Penerapan sistem klasifikasi ini berfungsi sebagai media edukasi bagi masyarakat untuk lebih sadar akan gejala dan konsekuensi kesehatan terkait obesitas. Dengan adanya sistem ini, masyarakat dapat melakukan pemeriksaan mandiri sebelum melanjutkan ke pemeriksaan medis profesional jika terdeteksi gejala obesitas.

DAFTAR PUSTAKA

[1] M. Blüher, "An overview of obesity-related complications: The epidemiological evidence linking body weight and other markers of obesity to adverse health outcomes," *Diabetes Obes. Metab.*, vol. 27, no. S2, pp. 3–19, Apr. 2025, doi: 10.1111/dom.16263

[2] Azizah, K. N. "Penyebab Fajri Pria Obesitas 300 Kg Meninggal Dunia, Begini Penjelasan RSCM." Internet: https://health.detik.com/berita-detikhealth/d-6788296/penyebab-fajri-pria-obesitas-300-kg-meninggal-dunia-begini-penjelasan-rscm, Jun. 23, 2023 [Feb. 20, 2026].

[3] Phinandita, V. "Penyebab Pria Obesitas BB 300 Kg Meninggal di RSCM: Syok Sepsis-Gagal Organ" Internet: https://health.detik.com/berita-detikhealth/d-6789833/penyebab-pria-obesitas-bb-300-kg-meninggal-di-rscm-syok-sepsis-gagal-organ, Jun. 24, 2023 [Feb. 20, 2026].

[4] Santika. "Belum Ada Penurunan Tren Kematian Akibat Obesitas Selama 20 Tahun Terakhir di Indonesia" Internet: https://databoks.katadata.co.id/datapublish/2023/04/18/belum-ada-penurunan-tren-kematian-akibat-obesitas-selama-20-tahun-terakhir-di-indonesia, Mar. 19, 2023 [Feb. 19, 2026].

p-ISSN: 2775-4227
e-ISSN: 2775-4219

[5] C. Boutari and C. S. Mantzoros, "A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on," *Metabolism*, vol. 133, p. 155217, Aug. 2022, doi: 10.1016/j.metabol.2022.155217

[6] R. J. F. Loos and G. S. H. Yeo, "The genetics of obesity: from discovery to biology," *Nat. Rev. Genet.*, vol. 23, no. 2, pp. 120–133, Feb. 2022, doi: 10.1038/s41576-021-00414-z

[7] L. Dakdouk, J. B. Daher, M. Skafi, and M. Ayache, "Machine Learning and Deep Learning Approaches for Obesity Prediction: a Systematic Review," in *2025 Eighth International Conference on Advances in Biomedical Engineering (ICABME)*, Debbieh, Lebanon: IEEE, Oct. 2025, pp. 01–10. doi: 10.1109/ICABME66883.2025.11211622

[8] F. Musa, F. Basaky, and O. E.O, "Obesity prediction using machine learning techniques," *J. Appl. Artif. Intell.*, vol. 3, no. 1, pp. 24–33, Jun. 2022, doi: 10.48185/jaai.v3i1.470

[9] Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA, N. Khodadadi, Electronics and Communications Engineering Department, Faculty of Engineering, Delta University for Science and Technology, Gamasa City 11152, Egypt, M. Saber, Department of System Programming, South Ural State University, 454080 Chelyabinsk, Russia, and M. Abotaleb, "A Data-Driven Approach for Obesity Classification using Machine Learning," *J. Artif. Intell. Metaheuristics*, vol. 3, no. 2, pp. 08–17, 2023, doi: 10.54216/JAIM.030201

[10] S. A. G. Ali, H. R. D. AL-Fayyadh, S. H. Mohammed, and S. R. Ahmed, "A Descriptive Statistical Analysis of Overweight and Obesity Using Big Data," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey: IEEE, Jun. 2022, pp. 1–6. doi: 10.1109/HORA55278.2022.9800098

[11] T. Hidayatulloh and L. Yusuf, "KLASIFIKASI TIPE BERAT TUBUH MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *INTI Nusa Mandiri*, vol. 18, no. 1, pp. 71–77, Aug. 2023, doi: 10.33480/inti.v18i1.4254

[12] S. Ali Alsareii *et al.*, "Machine-Learning-Enabled Obesity Level Prediction Through Electronic Health Records," *Comput. Syst. Sci. Eng.*, vol. 46, no. 3, pp. 3715–3728, 2023, doi: 10.32604/csse.2023.035687

[13] J. Gong, "Decision Trees in Business and Finance: A Review of Applications, Challenges, and Future Directions," *Adv. Econ. Manag. Res.*, vol. 15, no. 1, p. 683, Nov. 2025, doi: 10.56028/aemr.15.1.683.2025

[14] A. Muhammad Sani, A. Salihu BenMusa, and M. Haladu, "In-Depth Study of Decision Tree Model," *Int. J. Sci. Res. IJSR*, vol. 10, no. 11, pp. 705–709, Nov. 2021, doi: 10.21275/MR211102051237

[15] Ibnu Daqiqil ID, Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python, 1st ed. Riau: UR Press, 2021. doi: 10.5281/zenodo.5113507