

## KLASIFIKASI LEVEL OBESITAS MENGGUNAKAN DECISION TREE C45 DALAM MENENTUKAN AKURASI PADA KRITERIA INFORMATION GAIN, GAIN RATIO, GINI INDEX

M. Yhogha Ismail Ibn Ibrahim<sup>1</sup> Sandi Badiwibowo Atim<sup>2</sup>

<sup>1-2</sup> Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung,

<sup>1</sup> Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1, Kota Bandar Lampung, <sup>2</sup> Jl. Prof. Dr. Ir.  
Sumantri Brojonegoro No.1, Kota Bandar Lampung.

<sup>1</sup> yogaismail@fmipa.unila.ac.id, <sup>2</sup> sandibadiwibowoatim@fmipa.unila.ac.id.

**Abstrak** : Obesitas adalah penumpukan lemak ekstra yang disebabkan oleh ketidakcocokan jangka panjang antara asupan kalori dan pengeluaran energi. Menurut indeks RPJMN 2015-2019, 13,5 persen orang dewasa di Indonesia di atas usia 18 tahun kelebihan berat badan, 28,7 persen mengalami obesitas ( $BMI > 25$ ), bahkan 15,4 persen mengalami obesitas ( $BMI > 27$ ). 18,8% anak-anak berusia 5 hingga 12 tahun kelebihan berat badan, dan 10,8% dari mereka mengalami obesitas. Tindakan memilih, memeriksa, dan memodelkan sejumlah besar data untuk menemukan pola dan tren yang biasanya tidak diperhatikan. Algoritma DT C4.5 adalah evolusi dari algoritma ID3 yang menggunakan entropi informasi, atribut kontinu dan diskrit, atribut kategoris dan numerik, dan nilai yang hilang. Hasil yang diperoleh setelah menggunakan algoritma DT. C45 dengan Software Rapid Miner adalah 95,2%.

**Kata Kunci** : Obesitas, Data Mining, Decision Tree C4.5 Algoritma.

**Abstract:** Obesity is the buildup of extra fat brought on by a long-term mismatch between caloric intake and energy expenditure. According to the 2015-2019 RPJMN index, 13.5 percent of adults in Indonesia over the age of 18 are overweight, 28.7 percent are obese ( $BMI > 25$ ), and even 15.4 percent are obese ( $BMI > 27$ ). 18.8% of children aged 5 to 12 were overweight, and 10.8% of them were obese. The act of choosing, examining, and modeling vast amounts of data in order to discover patterns and trends that are not typically noticed.. The C4.5 algorithm is an evolution of the ID3 algorithm that uses information entropy, continuous and discrete attributes, categorical and numerical attributes, and missing values. The result obtained after using the DT algorithm. C45 with Software Rapid Miner is 95.2%.

**Keywords:** Obesity, Data Mining, Decision Tree C4.5 Algorithm

### PENDAHULUAN

Obesitas, yang merupakan akumulasi kelebihan lemak akibat ketidakseimbangan antara energi yang dikonsumsi dan energi yang dikeluarkan dalam jangka panjang, telah menjadi isu kesehatan global yang signifikan. Kelebihan berat badan dan obesitas tidak hanya ditandai dengan penambahan lemak tubuh yang abnormal, tetapi juga merupakan indikator penting

dari berbagai masalah kesehatan. Indeks Massa Tubuh (BMI), yang digunakan untuk mengukur berat badan relatif terhadap tinggi badan, menunjukkan bahwa nilai di atas 25 dianggap sebagai berat badan berlebih, sementara nilai di atas 30 mengindikasikan obesitas. Pada tahun 2017, data global menunjukkan lebih dari 4 juta kematian terkait obesitas,

menandakan dampak serius dari masalah ini terhadap kesehatan masyarakat.

Di Indonesia, situasi ini juga cukup memprihatinkan. Data dari RPJMN 2015-2019 mencatat bahwa 13,5% orang dewasa berusia di atas 18 tahun mengalami berat badan berlebih, sementara 28,7% mengalami obesitas dengan BMI lebih dari 25, dan 15,4% mengalami obesitas lebih berat dengan BMI lebih dari 27. Anak-anak juga tidak luput dari masalah ini; sekitar 18,8% anak usia 5-12 tahun mengalami kelebihan berat badan, dan 10,8% mengalami obesitas. Angka-angka ini menunjukkan tren yang mengkhawatirkan baik pada orang dewasa maupun anak-anak, yang mencerminkan peningkatan prevalensi obesitas di seluruh dunia.

Masalah obesitas tidak hanya berkaitan dengan faktor genetik atau pola makan, tetapi juga dipengaruhi oleh berbagai variabel seperti jenis kelamin, usia, tingkat pendidikan, sikap terhadap obesitas, asupan nutrisi (termasuk karbohidrat, kalori, dan protein), serta faktor stres dan aktivitas fisik. Penurunan aktivitas fisik seiring bertambahnya usia berkontribusi pada berkurangnya massa otot, yang pada gilirannya mengurangi pembakaran kalori dan meningkatkan risiko akumulasi lemak tubuh. Tanpa adanya pengaturan kalori yang memadai, energi yang tidak digunakan akan disimpan dalam bentuk lemak, menyebabkan obesitas.

Fenomena obesitas kini tidak terbatas pada negara-negara berpenghasilan tinggi saja, tetapi juga telah merambah ke negara-negara berpenghasilan rendah dan menengah, terutama di daerah perkotaan. Di wilayah Afrika sub-Sahara dan Asia, obesitas kini merupakan bagian dari masalah kekurangan gizi, mengubah

persepsi bahwa obesitas hanya menjadi masalah di negara-negara kaya. Penelitian menunjukkan bahwa prevalensi obesitas lebih tinggi di negara-negara kaya dibandingkan dengan negara-negara miskin, dan prevalensi ini juga meningkat di kalangan anak-anak di negara-negara dengan penghasilan rendah dan menengah.

Dalam konteks ini, penggunaan teknologi informasi dan ilmu komputer dapat memberikan kontribusi signifikan dalam memecahkan masalah obesitas. Analisis data besar (*big data*), model prediksi berbasis kecerdasan buatan, dan aplikasi kesehatan digital dapat membantu dalam pemantauan, pencegahan, dan pengelolaan obesitas. Oleh karena itu, penting untuk mengeksplorasi pengembangan dan pendayagunaan ilmu komputer dalam konteks kesehatan, terutama untuk meningkatkan efektivitas intervensi dan solusi yang ada untuk mengatasi obesitas secara lebih efektif.

Salah satu cara di mana teknologi informasi dan ilmu komputer berperan adalah melalui analisis data besar (*big data*). Dengan adanya volume data kesehatan yang terus berkembang, baik yang berasal dari rumah sakit, aplikasi kebugaran, perangkat wearable, maupun survei kesehatan, analisis data besar memungkinkan untuk memproses dan menganalisis informasi dalam jumlah yang sangat besar dengan lebih cepat dan akurat. *Big data* dapat digunakan untuk mengidentifikasi pola-pola yang tidak terlihat sebelumnya, seperti hubungan antara faktor-faktor gaya hidup, pola makan, aktivitas fisik, dan faktor genetik yang berkontribusi terhadap obesitas. Dengan kemampuan untuk menganalisis berbagai sumber data secara simultan, dapat ditemukan tren atau faktor risiko

yang dapat membantu dalam pengembangan kebijakan kesehatan yang lebih tepat sasaran.

Selain itu, model prediksi berbasis kecerdasan buatan (AI) juga memainkan peran penting dalam penanganan obesitas. Dengan memanfaatkan machine learning dan deep learning, AI dapat digunakan untuk memprediksi kemungkinan seseorang mengalami obesitas berdasarkan data yang tersedia, seperti usia, jenis kelamin, tingkat aktivitas, kebiasaan makan, dan faktor-faktor lain yang relevan. Model prediksi ini dapat menjadi alat yang sangat berguna untuk dokter atau penyedia layanan kesehatan dalam mengidentifikasi individu yang berisiko tinggi mengalami obesitas, bahkan sebelum kondisi tersebut berkembang lebih lanjut. Teknologi ini juga dapat digunakan untuk memantau perubahan dalam status kesehatan seseorang, memberikan peringatan dini tentang potensi masalah, dan menyarankan intervensi atau perubahan gaya hidup yang diperlukan.

Lebih lanjut, aplikasi kesehatan digital seperti aplikasi pelacakan diet, kebugaran, dan berat badan yang terhubung dengan perangkat wearable (seperti smartwatch atau fitness tracker) juga berkontribusi besar dalam pengelolaan obesitas. Aplikasi-aplikasi ini tidak hanya memungkinkan individu untuk melacak asupan makanan, aktivitas fisik, dan perubahan berat badan mereka, tetapi juga dapat memberikan saran berbasis data dan umpan balik yang dipersonalisasi. Dengan teknologi ini, individu dapat lebih mudah mengakses informasi yang relevan dan mengontrol kesehatan mereka dengan cara yang lebih terukur dan terdokumentasi. Aplikasi kesehatan digital juga dapat membantu dalam memotivasi

perubahan perilaku dengan memberikan notifikasi dan pengingat, yang sangat berguna dalam mengatasi obesitas, yang sering kali berhubungan dengan kebiasaan jangka panjang.

Selain dari sisi pencegahan dan pemantauan, teknologi informasi juga dapat mendukung program intervensi medis dan kesehatan masyarakat dalam skala yang lebih besar. Dengan platform telemedicine dan aplikasi berbasis cloud, ahli gizi, pelatih kebugaran, dan dokter dapat berkomunikasi dengan pasien secara langsung, memberikan nasihat atau program penurunan berat badan yang lebih sesuai dengan kebutuhan individu, serta memantau perkembangan pasien secara lebih mudah dan efisien. Ini akan mempercepat akses terhadap layanan kesehatan, terutama di daerah-daerah yang kurang terlayani oleh fasilitas medis, dan membantu individu mengatasi obesitas dengan cara yang lebih terarah dan terstruktur.

Secara keseluruhan, pendayagunaan ilmu komputer dalam konteks kesehatan, terutama dalam menangani masalah obesitas, memiliki potensi besar untuk meningkatkan efektivitas berbagai intervensi dan solusi yang ada. Teknologi dapat memperkuat upaya pencegahan dan pengelolaan obesitas dengan menyediakan alat yang lebih canggih untuk memantau kesehatan individu, mengembangkan model prediksi yang lebih akurat, serta memberikan solusi berbasis data yang lebih personalized. Oleh karena itu, penting untuk terus mengeksplorasi dan mengembangkan aplikasi teknologi informasi dan ilmu komputer dalam sektor kesehatan, yang tidak hanya akan membantu dalam mengurangi prevalensi obesitas, tetapi juga berkontribusi pada peningkatan

kualitas hidup dan kesehatan masyarakat secara keseluruhan.

## **KAJIAN PUSTAKA DAN LANDASAN TEORI**

### **1. Data mining**

Salah satu bidang yang mendapat manfaat signifikan dari perkembangan teknologi informasi adalah Pembelajaran Data (Data Mining).[1]

Data mining adalah sebuah metode untuk melakukan akuisisi pengetahuan.[2]

Dengan menggunakan berbagai teknik dan algoritma, data mining membantu dalam mengidentifikasi tren, mengklasifikasikan data, dan membuat prediksi yang berguna untuk pengambilan keputusan. Ini melibatkan tahap-tahap seperti preprocessing data, pemodelan, evaluasi, dan interpretasi hasil. [3]

Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan.[4]

Data mining dapat didefinisikan sebagai proses eksplorasi dan analisis data besar dengan tujuan untuk menemukan pola yang berguna atau informasi yang dapat dimanfaatkan untuk meningkatkan keputusan atau strategi dalam berbagai bidang. Dalam konteks yang lebih luas, data mining tidak hanya terbatas pada pengolahan data terstruktur, tetapi juga mencakup pengolahan data tidak terstruktur, seperti teks, gambar, dan audio.

Secara umum, data mining mengacu pada proses menggunakan algoritma statistik dan teknik pembelajaran mesin untuk mengekstrak informasi yang relevan dari data. Hal ini tidak hanya melibatkan pencarian pola dalam data, tetapi juga pembuatan model yang dapat memprediksi nilai atau kelas baru berdasarkan data

yang ada. Beberapa sumber menyebutkan bahwa data mining adalah langkah penting dalam knowledge discovery yang menghubungkan pengumpulan data dengan keputusan berbasis data.

### **2. Klasifikasi**

Salah satu teknik data mining adalah klasifikasi.[5]

Klasifikasi adalah proses untuk menemukan model yang dapat membedakan kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.[6]

Tujuan dari klasifikasi adalah untuk menemukan model dari data latih yang akan membedakan atribut ke dalam kategori atau kelas yang sesuai model. [7]

### **3. Obesitas**

Obesitas adalah kondisi medis yang ditandai dengan penumpukan lemak tubuh yang berlebihan dan dapat mempengaruhi kesehatan secara signifikan. Menurut World Health Organization (WHO), obesitas diukur menggunakan Indeks Massa Tubuh (IMT), yaitu rasio berat badan terhadap kuadrat tinggi badan ( $BMI = \text{berat badan (kg)} / \text{tinggi badan (m)}^2$ ) (WHO, 2020). Klasifikasi obesitas menurut BMI adalah sebagai berikut:

- Normal: BMI 18,5 - 24,9
- Kelebihan Berat Badan: BMI 25 - 29,9
- Obesitas Kelas 1: BMI 30 - 34,9
- Obesitas Kelas 2: BMI 35 - 39,9
- Obesitas Kelas 3: BMI  $\geq 40$

Obesitas adalah penyakit fisik kronis yang paling umum di masyarakat modern.[8]

Faktor penyebab obesitas pada anak bersifat multifaktorial diantaranya ketidakseimbangan antara pemasukan energi dengan pengeluaran energi, faktor genetik, faktor psikologis, dan kondisi lainnya

### **4. Decision Tree C4.5**

Konsep dasar dari algoritma C4.5 yaitu mengubah data menjadi pohon keputusan dan aturan-aturan keputusan (rule). Kelebihan dari algoritma C4.5 yaitu mudah dimengerti, fleksibel dan menarik karena mampu divisualisasikan dalam gambar. [7]

Pohon keputusan merupakan output dari perhitungan entropi dan perolehan informasi setiap kali atribut pada pohon mendapatkan kelas. Proses ini melibatkan iterasi berulang, di mana setiap langkah melibatkan pemilihan atribut terbaik berdasarkan entropi dan perolehan informasi. Langkah-langkah ini terus diulang hingga tidak ada lagi operasi komputasi yang dapat dilakukan, dan setiap cabang pohon memiliki kelas yang ditetapkan. [9]

*Decision Tree* (Pohon Keputusan) adalah salah satu metode yang paling umum digunakan dalam data mining dan machine learning untuk tugas klasifikasi dan regresi. Metode ini menyajikan keputusan dan konsekuensinya dalam bentuk pohon, di mana setiap node internal mewakili atribut atau fitur yang diambil untuk memisahkan data, dan setiap cabang mewakili keputusan atau nilai yang dicapai berdasarkan atribut tersebut. Di ujung pohon (leaf nodes), terdapat label atau nilai prediksi yang menunjukkan hasil atau keputusan akhir.

Decision tree adalah model yang berupa struktur pohon yang digunakan untuk mengambil keputusan berdasarkan data yang ada. Pohon ini memiliki:

- *Root Node*: Node pertama di pohon yang mewakili seluruh dataset yang akan dibagi berdasarkan fitur atau atribut tertentu.
- *Internal Nodes*: Node yang membagi dataset lebih lanjut dengan menggunakan atribut tertentu.
- *Leaf Nodes (Daun)*: Node terakhir yang menunjukkan hasil atau prediksi akhir untuk kelas atau nilai yang ingin diprediksi.

## 5. Penerapan Decision Tree dalam Klasifikasi Kesehatan

Penggunaan pohon keputusan dalam bidang kesehatan, termasuk dalam klasifikasi obesitas, telah terbukti efektif. Misalnya dalam Penyakit Paru Paru untuk menemukan pola pada kanker paru-paru, yang memungkinkan pendeteksian awal penyakit kanker paru-paru. Salah satu teknik klasifikasi yang banyak digunakan adalah Decision Tree.[10]

Penerapan Decision Tree dalam Klasifikasi Kesehatan adalah salah satu contoh penggunaan machine learning untuk menganalisis data medis dan membantu pengambilan keputusan terkait diagnosis, prognosis, atau kebijakan kesehatan lainnya. Decision Tree dapat digunakan untuk mengklasifikasikan pasien ke dalam kategori tertentu berdasarkan berbagai fitur medis yang dimiliki, seperti usia, jenis kelamin, gejala, hasil tes laboratorium, dan faktor risiko lainnya.

Dalam konteks obesitas, Decision Tree dapat digunakan untuk mengklasifikasikan individu ke dalam berbagai kategori obesitas berdasarkan berbagai atribut seperti umur, jenis kelamin, tinggi badan, berat badan, dan faktor gaya hidup. Pengembangan teknologi diagnostik penting untuk deteksi dini.[11]

## METODE

### 1. Pengumpulan Data

Pengumpulan Data adalah proses sistematis dalam mengumpulkan informasi atau data yang relevan untuk tujuan analisis atau penelitian. Dalam konteks data mining, pengumpulan data sangat penting karena kualitas dan keberagaman data yang dikumpulkan akan sangat mempengaruhi hasil dan akurasi model yang dibangun. Data yang dikumpulkan akan digunakan untuk menemukan pola, tren, atau hubungan yang dapat diinterpretasikan dan digunakan dalam pengambilan keputusan.

Data adalah komponen krusial dalam setiap penelitian. Dalam konteks klasifikasi level obesitas, data berasal dari situs

[www.kaggle.com](http://www.kaggle.com). Dengan jumlah  
 sebanyak 16 atribut

Tabel 1.1 Atribut

Atribut	Deskripsi	Kriteria
Jenis Kelamin	- Wanita - Pria	- Wanita - Pria
Usia	Usia	14 – 61
Tinggi	Tinggi	145 – 198
Berat	Berat	39 – 173
Riwayat Keluarga Memiliki Obesitas	Keluarga yang menderita obesitas	1 = Tidak 2 = Ya
FAVC	Atribut yang menunjukkan seringnya konsumsi makanan dengan highcalories	1 = No 2 = Yes
FCVC	Atribut yang menunjukkan frekuensi konsumsi sayuran	1 = Tidak pernah 2 = Kadang-kadang 3 = Selalu
NCP	Jumlah makanan pokok yang dikonsumsi setiap harinya	1 = Satu 2 = Dua 3 = Tiga 4 = lebih dari tiga
CAEC	Frekuensi konsumsi makanan di antara waktu makan	1 = Tidak. 2 = Kadang-kadang 3 = Sering 4 = Selalu
Smoke	Perokok atau tidak	1 = Tidak. 2 = Ya
CH2O	Jumlah air untuk diminum setiap hari	1 = Kurang dari satu liter 2 = Antara 1 dan 2 L 3 = Lebih dari 2 L

SCC	Pantau kalori yang dikonsumsi setiap hari	1 = Tidak. 2 = Ya
FAF	Waktu Menggunakan Perangkat	0 = Saya tidak memiliki 1 = 1 Or 2 hari 2 = 2 atau 4 hari 3 = 4 atau 5 hari
TUE	Waktu menggunakan perangkat	0 = 0 - 2 jam 1 = 3 - 5 jam 2 = Lebih dari 5 jam
CALC	Frekuensi minum alcohol	1 = Tidak. 2 = Kadang-kadang 3 = Sering 4 = Selalu
MTrans	Transportasi yang sering digunakan setiap hari	1 = Mobil 2 = Sepeda motor 3 = Sepeda 4 = Transportasi Umum 5 = Berjalan

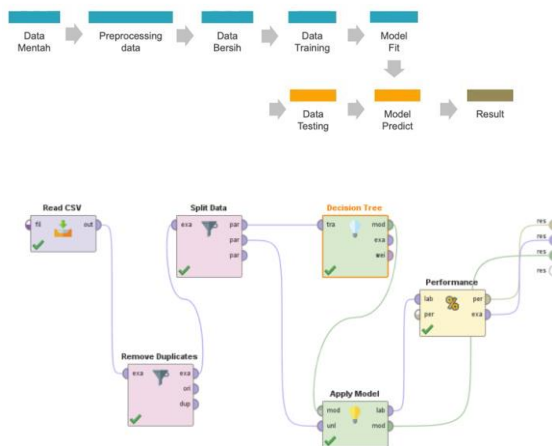
## 2. Pembagian Data

Pembagian data ini sangat penting untuk memastikan bahwa model yang dibangun tidak hanya menghafal data pelatihan, tetapi juga dapat menggeneralisasi dengan baik ke data yang belum pernah dilihat sebelumnya (data testing). Pembagian data yang tepat akan membantu dalam menilai kinerja model dan memastikan bahwa model tersebut dapat diterapkan pada situasi dunia nyata.

Pada tahap ini dilakukan pengujian Algoritma DT C.45 dengan menggunakan 70 % Data Training dan sebanyak 30% Data Testing (7:3) untuk melihat hasil akurasi maksimal dari masing kriteria.

### 3. Penerapan Algoritma Decision Tree C4.5

Pada tahap ini, penerapan pemodelan menggunakan Algoritma C.45 dilakukan dengan memanfaatkan tiga kriteria utama yang digunakan untuk memilih atribut terbaik dalam membagi dataset, yaitu *Information Gain*, *Gain Rasio*, dan *Gini Index*. Tujuan utama dari penerapan ketiga kriteria ini adalah untuk mengevaluasi seberapa efektif masing-masing dalam memisahkan data dan menentukan kualitas prediksi yang dihasilkan oleh algoritma C.45.



Gambar 3.1 Model Decision Tree

#### Information Gain

*Information Gain* adalah sebuah metrik yang digunakan dalam *data mining* dan *machine learning*, khususnya pada algoritma *decision tree* (seperti ID3 dan C4.5), untuk mengukur sejauh mana sebuah atribut dalam dataset memberikan informasi untuk memisahkan data ke dalam kelas-kelas tertentu.

accuracy: 95.20%

	true Normal...	true Overweig...	true Overweig...	true Obesity...	true Insuffici...	true Obesity...	true Obesity...	class precision
pred Normal...	74	4	1	0	2	0	0	91.36%
pred Overwel...	8	75	2	0	0	0	0	98.24%
pred Overwel...	0	4	83	0	0	0	0	95.40%
pred Obesity...	0	0	0	102	0	3	0	97.14%
pred Insuffici...	3	0	0	0	78	0	0	95.30%
pred Obesity...	0	0	0	3	0	86	0	95.63%
pred Obesity...	0	0	0	0	0	0	97	100.00%
class recall	87.06%	90.36%	95.51%	97.14%	97.50%	95.63%	100.00%	

Gambar 3.2 Akurasi Information Gain

#### Gain Ratio

merupakan algoritma seleksi fitur yang banyak digunakan peneliti karena handal

dan mampu berjalan pada dimensi data yang tinggi. Tahapan seleksi fitur sebenarnya adalah menghitung kepentingan dari keseluruhan atribut data yang ada untuk nantinya dijadikan patokan dalam tahap berikutnya yaitu klasifikasi. Hasil akhir dari tahapan ini adalah atribut yang memiliki tingkat kepentingan yang tinggi selanjutnya akan digunakan sedangkan atribut dengan tingkat kepentingan yang rendah tidak akan digunakan dalam tahap berikutnya.[12]

accuracy: 84.64%

	true Normal...	true Overweig...	true Overweig...	true Obesity...	true Insuffici...	true Obesity...	true Obesity...	class precision
pred Normal...	61	2	0	0	5	0	0	89.71%
pred Overwel...	17	46	4	0	0	0	0	68.66%
pred Overwel...	3	35	80	4	0	2	0	64.52%
pred Obesity...	0	0	2	88	0	15	0	85.22%
pred Insuffici...	4	0	0	0	75	0	0	94.94%
pred Obesity...	0	0	0	3	0	72	0	95.00%
pred Obesity...	0	0	0	0	0	0	97	100.00%
class recall	71.76%	55.42%	93.02%	93.33%	93.75%	80.90%	100.00%	

Gambar 3.3 Akurasi Gain Ratio

#### Gini Index

adalah salah satu metode yang digunakan dalam data mining dan machine learning, khususnya dalam algoritma pohon keputusan (*decision tree*), seperti CART (*Classification and Regression Trees*). Gini Index digunakan untuk mengukur ketidakmurnian atau keragaman dalam suatu dataset, yang sangat penting dalam pemisahan data untuk membangun pohon keputusan.

accuracy: 94.56%

	true Normal...	true Overweig...	true Overweig...	true Obesity...	true Insuffici...	true Obesity...	true Obesity...	class precision
pred Normal...	76	3	1	0	3	0	0	91.57%
pred Overwel...	5	77	5	0	0	0	0	95.51%
pred Overwel...	0	3	77	2	0	0	0	93.90%
pred Obesity...	0	0	3	103	0	5	0	92.79%
pred Insuffici...	4	0	0	0	77	0	0	95.06%
pred Obesity...	0	0	0	0	0	84	0	100.00%
pred Obesity...	0	0	0	0	0	0	97	100.00%
class recall	89.41%	92.77%	89.52%	98.10%	95.25%	94.38%	100.00%	

Gambar 3.4 Akurasi Gini Index

Berikut adalah penjelasan yang lebih panjang dan terperinci berdasarkan teks yang Anda berikan:

Berdasarkan hasil yang diperoleh dari pengujian dan evaluasi yang dilakukan pada setiap kriteria pemilihan atribut dalam **Algoritma C.45**, dapat dilihat bahwa setiap kriteria memberikan hasil yang berbeda dalam hal tingkat akurasi model. Pada pengujian ini, tiga kriteria yang digunakan untuk menilai efektivitas pemisahan data adalah *Information Gain*, *Gain Rasio*, dan

*Gini Index*. Setelah proses penerapan dan evaluasi yang mendalam, diperoleh hasil yang menunjukkan bahwa *Information Gain* memberikan tingkat akurasi tertinggi di antara ketiga kriteria yang diuji.

## KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan dengan menggunakan Algoritma Decision Tree (DT) C.45 yang diterapkan melalui aplikasi RapidMiner, diperoleh beberapa temuan yang signifikan terkait kinerja dan efektivitas model dalam mengklasifikasikan dataset yang diuji sebagai berikut :

Criterion	Max Depth	Apply Pruning	Apply PrePruning	Result (Accurate %)
Information Gain	10	No	No	95,2%
Gain Rasio	10	Yes	Yes	85,6%
Gini Index	10	Yes	Yes	94,56%

Penelitian ini telah mengeksplorasi penerapan algoritma Decision Tree (DT) menggunakan metode C.45 dalam menentukan level obesitas seseorang, dengan tujuan untuk membangun model klasifikasi yang efektif dan akurat. Dalam eksperimen ini, data yang digunakan dibagi menjadi dua bagian, yaitu 70% untuk data pelatihan (training) dan 30% untuk data pengujian (testing). Hasil dari evaluasi model menunjukkan bahwa algoritma C.45 yang diterapkan dengan kriteria *Information Gain* menghasilkan tingkat akurasi yang sangat baik, yaitu sebesar 95%.

Metode C.45 adalah salah satu algoritma pembelajaran mesin berbasis pohon keputusan yang populer dan terbukti efektif dalam menangani masalah klasifikasi yang melibatkan data kategori. Keunggulan

utama dari C.45 terletak pada kemampuannya untuk membangun pohon keputusan yang tidak hanya dapat mengklasifikasikan data dengan baik, tetapi juga mampu mengatasi data yang bersifat biner maupun multikelas dengan tingkat kompleksitas yang lebih rendah. Dalam penelitian ini, penerapan C.45 dengan menggunakan *Information Gain* sebagai kriteria pemisahan atribut menghasilkan model yang mampu memisahkan kelas dengan jelas berdasarkan fitur-fitur yang tersedia.

Penting untuk dicatat bahwa penelitian ini tidak mengaktifkan fitur Pruning maupun Pre-Pruning, yang biasanya digunakan untuk menghindari overfitting dan memperbaiki generalisasi model. Meskipun demikian, hasil akurasi yang tinggi sebesar 95% menunjukkan bahwa model yang dibangun masih mampu memberikan prediksi yang akurat meskipun dengan risiko potensi overfitting. Hal ini menunjukkan bahwa dalam konteks dataset ini, penggunaan C.45 tanpa penerapan teknik pruning tidak menyebabkan penurunan performa yang signifikan.

*Information Gain* yang digunakan dalam algoritma C.45 berfungsi untuk memilih atribut yang paling informatif dalam setiap langkah pemisahan. Dengan menghitung pengurangan entropi pada setiap cabang pohon keputusan, C.45 mampu memilih atribut yang memiliki kekuatan diskriminasi tertinggi, yang dalam hal ini berhubungan langsung dengan tingkat obesitas berdasarkan variabel yang ada. Penggunaan *Information Gain* ini terbukti efektif dalam membangun model klasifikasi yang dapat memetakan level obesitas dengan tingkat akurasi yang tinggi.



Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma Decision Tree C.45 adalah metode yang sangat baik untuk diterapkan dalam pemodelan untuk menentukan level obesitas seseorang. Meskipun akurasi yang dicapai sudah sangat baik, penelitian lebih lanjut dapat dilakukan dengan mengaktifkan teknik Pruning untuk meminimalkan risiko overfitting, serta melakukan eksperimen dengan data lebih besar atau fitur-fitur tambahan yang mungkin dapat meningkatkan performa model.

Selain itu, untuk aplikasi dunia nyata, model ini dapat digunakan untuk mendukung keputusan medis dalam mendeteksi tingkat obesitas pada individu, yang dapat berkontribusi dalam upaya pencegahan dan pengelolaan masalah kesehatan terkait obesitas. Dalam pengembangan lebih lanjut, penerapan teknik ensemble atau model yang lebih kompleks seperti Random Forest atau Gradient Boosting bisa menjadi arah yang menarik untuk meningkatkan ketahanan model terhadap variabilitas data dan meningkatkan tingkat akurasi lebih lanjut.

Dengan demikian, penelitian ini tidak hanya memberikan wawasan tentang penerapan algoritma C.45 dalam klasifikasi level obesitas, tetapi juga membuka peluang untuk penelitian dan pengembangan lebih lanjut dalam bidang data mining dan klasifikasi medis yang lebih luas..

## REFERENSI

- [1] S. Shalsabilla, P. Rachmawati, K. Vidya Prakusa, and S. Rihastuti, "Penerapan Data Mining dengan Metode Decision Tree untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka," Nov. 2023.
- [2] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [3] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [4] R. Saputra and A. J. P. Sibarani, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," 2020. [Online]. Available: <http://jurnal.mdp.ac.id>
- [5] H. I. Islam, M. K. Mulyadien, and U. Enri, "Penerapan Algoritma C4.5 dalam Klasifikasi Status Gizi Balita," *Ilmiah Wahana Pendidikan*, vol. 8, no. 10, pp. 116–125, 2022, [Online]. Available: <http://jurnal.peneliti.net/index.php/JIWP/article/view/1839/1425>
- [6] I. G. A. Purnajiwa Arimbawa and N. A. Sanjaya ER, "Penerapan Metode Adaboost Untuk Multi-Label Classification Pada Dokumen Teks," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 9, no. 1, p. 127, 2020, doi: 10.24843/jlk.2020.v09.i01.p13.
- [7] D. R. Ente, S. A. Thamrin, and H. Kuswanto, "Klasifikasi faktor-faktor penyebab penyakit diabetes melitus di rumah sakit unhas menggunakan algoritma c4.5 \*," *Indonesian*

- Journal of Statistics and Its Applications*, vol. Vol 4 No 1, pp. 80–88, 2020.
- [8] N. M. Murtane, “Obesitas dan Depresi pada Orang Dewasa,” *Jurnal Ilmiah Kesehatan Sandi Husada*, vol. 10, no. 1, pp. 88–93, 2021, doi: 10.35816/jiskh.v10i1.515.
- [9] A. Irma Purnamasari and T. Suprpti, “PENERAPAN ALGORITMA DECISION TREE DALAM KLASIFIKASI PENYAKIT STROKE OTAK,” 2024.
- [10] H. Widya, N. Surya Putra, V. Atina, and J. Maulindar, “Penerapan Algoritme Decision Tree Pada Klasifikasi Penyakit Kanker Paru-Paru,” Dec. 2023, [Online]. Available: <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer>,
- [11] R. Antika, A. Rifa, F. Dikananda, D. Indriya Efendi, and R. Narasati, “PENERAPAN ALGORITMA DECISION TREE BERBASIS POHON KEPUTUSAN DALAM KLASIFIKASI PENYAKIT JANTUNG,” 2023.
- [12] M. F. Kurniawan, J. Nugraha, I. Stmik, and W. P. Pekalongan, “Peningkatan Performa Algoritma Naive Bayes dengan Gain Ratio untuk Klasifikasi Keganasan Kanker Payudara,” *Oktober*, vol. XIII, no. 2, 2018, [Online]. Available: <http://ejournal.stmik-wp.ac.id>