SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika

https://ejurnal.itats.ac.id/snestik dan https://snestik.itats.ac.id

Informasi Pelaksanaan :
SNESTIK V - Surabaya, 26 April 2025
Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:
DOI : 10.31284/p.snestik.2025.7619

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Metode Decision Tree Untuk Klasifikasi Berat Badan Obesitas

Tutuk Indriyani*, Chelvin Suprapto, Iqbal Izha Mahendra, Raditya Pratama, Muchamad Nurdin Wahyudi

Institut Teknologi Adhi Tama Surabaya

e-mail: tutuk @itats.ac.id

ABSTRACT

Obesity is a serious health problem and is one of the main factors causing chronic diseases such as diabetes, heart disease, and hypertension. Handling obesity requires a holistic and multidisciplinary approach, starting from notifying individuals at high risk of obesity. One method that can be used to classify obesity risk is to consider the individual's weight. In this study, the Decision Tree method was used to classify weight and identify factors that contribute to obesity. The C4.5 Algorithm decision tree is used to create a decision tree, the decision tree of an object classified in the tree must be tested for its Entropy and Gain values. The Entropy value is then calculated for the Gain value of each attribute, then the attribute with the highest Gain is selected as the test attribute of a node. The process carried out during classification is that the dataset is divided into two parts, namely Training data and Testing data, then the Entropy and Gain calculations are carried out to produce the Decision Tree C4.5 decision tree. The previous process produced classification results from the Testing data.The data setistaken from https://www.kaggle.com/datasets/denih1dayat/obesitas. The amount of data used for this study is 355. Based on the classification results, testing was carried out using Confusion Matrix to calculate Accuracy, Precision, Recall, F1_Score. The model successfully detected about 92% of all actual class examples in the dataset and the f1-score result was 0.89 which means the f1-score was 89%,

Keywords: *classification; Decision tree C4.5; Obesity, confusion matrix.*

ABSTRAK

Obesitas merupakan masalah kesehatan yang serius dan telah menjadi salah satu faktor utama

penyebab penyakit kronis seperti diabetes, penyakit jantung, dan hipertensi. Pengelolaan obesitas memerlukan pendekatan yang holistik dan multidisiplin, dimulai dari identifikasi individu yang berisiko tinggi mengalami obesitas. Salah satu metode yang dapat digunakan untuk melakukan klasifikasi risiko obesitas adalah dengan mempertimbangkan berat badan individu.Dalam penelitian ini, metode *Decision Tree* digunakan untuk melakukan klasifikasi berat badan dan mengidentifikasi faktor-faktor yang berkontribusi terhadap obesitas. Decision tree C4.5 algoritma yang digunakan untuk membuat decision tree (pohon keputusan), pohon keputusan sebuah obyek yang diklasifikasikan pada pohon wajib dites nilai Entropy dan Gainnya. Nilai Entropy tersebut kemudian dihitung nilai Gain masing-masing atribut, kemudian atribut dengan Gain tertinggi dipilih menjadi test atribut dari suatu node. Proses yang dilakukan saat klasifikasi adalah dataset dibagi menjadi dua bagian, data Training dan data Testing, kemudian menghitung Entropy dan Gain, sehingga menghasilkan pohon keputusan Decision Tree C4.5. Dari proses sebelumnya menghasilkan hasil klasifikasi dari data Testing. Dataset diambil dari https://www.kaggle.com/datasets/denih1dayat/obesitas. Jumlah data yang digunakan untuk penelitian berjumlah 355.Berdasarkan hasil klasifikasi di uji menggunakan Confusion Matrix untuk menghitung Accuracy, Precision, Recall, F1_Score. Model berhasil mendeteksi sekitar 92% dari semua contoh kelas yang sebenarnya ada dalam dataset dan hasil *f1-score* adalah 0.89 yang berarti *f1-score* bernilai 89%.

Kata kunci: klasifikasi;Decision tree C4.5; Obesitas, confusion matrix

PENDAHULUAN

Obesitas merupakan masalah kesehatan yang semakin meningkat secara global dan menjadi perhatian utama dalam bidang kesehatan masyarakat [1]. Menurut Organisasi Kesehatan Dunia (WHO), obesitas telah mencapai tingkat epidemiologi yang mengkhawatirkan, dengan prevalensi yang meningkat pesat di berbagai negara, termasuk Indonesia. Obesitas tidak hanya meningkatkan risiko berbagai penyakit kronis seperti diabetes, penyakit jantung, dan stroke, tetapi juga berdampak pada kualitas hidup seseorang dan meningkatkan biaya perawatan kesehatan [2].

Pengelolaan obesitas memerlukan pendekatan yang holistik dan multidisiplin, dimulai dari identifikasi individu yang berisiko tinggi mengalami obesitas. Salah satu metode yang dapat digunakan untuk melakukan klasifikasi risiko obesitas adalah dengan mempertimbangkan berat badan individu. Penentuan kategori obesitas berdasarkan berat badan biasanya mengacu pada indeks massa tubuh (Body Mass Index/BMI), yang dihitung berdasarkan perbandingan antara berat badan (kg) dengan tinggi badan (m) kuadrat[3]. Penelitian penelitian sebelumnya meliputi: melakukan klasifikasi dengan metode decision tree fuzzy yang diterapakan pada objek data penjualan menghasilkan akutrasi dibawah 85% [4], metode pengelompokan k-Means pertama-tama digunakan untuk mempartisi instans pelatihan menjadi k kluster menggunakan kesamaan jarak Euclidean dimana setiap kluster, yang mewakili wilayah kepadatan instans normal atau anomali, kami membangun pohon keputusan menggunakan algoritma pohon keputusan C4.5 menghasilkan akurasi 88% [5]. Pada saat yang sama, perkembangan dalam bidang ilmu komputer telah membuka peluang baru dalam analisis data dan prediksi. Salah satu metode yang populer digunakan dalam klasifikasi data adalah metode decision tree [6]. Decision tree adalah model prediktif yang menggunakan struktur pohon keputusan untuk memetakan data input ke output dengan cara yang berbasis aturan [7]. Penerapan metode decision tree dalam klasifikasi berat badan obesitas memiliki potensi untuk menjadi alat yang efektif dalam mengidentifikasi individu yang berisiko tinggi mengalami obesitas berdasarkan berat badan mereka [8]. Dengan memahami hubungan antara berat badan dan kategori

obesitas secara lebih terperinci, intervensi dan tindakan preventif yang tepat dapat diambil untuk mengurangi risiko obesitas dan penyakit terkaitnya. penelitian ini akan difokuskan pada beberapa pertanyaan pokok yang akan menjadi fokus utama dalam penelitian. bagaimana metode decision tree dapat diterapkan dalam klasifikasi berat badan obesitas dan bagaimana akurasi aplikasi hasil klasifikasi berat badan obesitas menggunakan metode decision tree dalam konteks penanganan dan pencegahan obesitas. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi penerapan metode decision tree dalam klasifikasi berat badan obesitas, dengan harapan dapat memberikan kontribusi dalam pengembangan metode analisis yang lebih efektif dalam penanganan masalah obesitas serta bertujuan untuk mendapatkan pemahaman yang lebih mendalam tentang potensi metode decision tree dalam klasifikasi berat badan obesitas, serta untuk menyumbangkan pengetahuan baru dalam pengembangan strategi pengelolaan obesitas yang lebih efektif. Memiliki kontribusi dalam Bidang Kesehatan Masyarakat: Penelitian ini akan memberikan pemahaman yang lebih baik tentang penggunaan metode decision tree dalam klasifikasi berat badan obesitas. Informasi ini dapat menjadi dasar bagi para praktisi kesehatan masyarakat untuk mengidentifikasi individu yang berisiko tinggi mengalami obesitas dan mengambil tindakan pencegahan yang sesuai [9]. Kontribusi dalam Pengembangan Ilmu Pengetahuan: Penelitian ini dapat memberikan kontribusi yang signifikan dalam pengembangan ilmu pengetahuan dan teknologi, khususnya dalam bidang analisis data dan kesehatan masyarakat. Temuan dan metodologi yang dihasilkan dari penelitian ini dapat menjadi sumbangan berharga bagi pengetahuan yang ada dan menjadi dasar untuk penelitian lanjutan di masa depan. Dengan manfaat-manfaat tersebut, diharapkan penelitian ini dapat memberikan dampak positif yang signifikan dalam upaya penanganan dan pencegahan obesitas, serta dalam pengembangan metodologi analisis data yang lebih efektif dan relevan dalam konteks kesehatan masyarakat.

METODE

Dalam penelitian ini menerapkan prediksi berat badan pada pasien yang nantinya dapat mengetahui beberapa pasien yang terkena berat badan obesitas atau tidak. Metode yang digunakan pada tugas akhir ini adalah *Decision tree* untuk klasifikasi data berat badan obesitas. Didalam kalsifikasi ini terdapat berbagai proses yang dimulai dari data *input,* proses, dan *output* berupa informasi prediksi berat badan obesitas yang berguna untuk memudahkan kerja dokter saat melakukan pemeriksaan pada pasien yang mengalami obesitas [4]. Namun pada saat melakukan prediksi dibutuhkan terlebih dahulu yaitu alur atau tahapan-tahapan yang membantu jalannya penelitian, sehingga penelitian dapat berjalan sesuai renacana. Alur program untuk penelitian ini direpresentasikan pada diagram blok pada Gambar1.



Gambar 1. Desain sistem Penelitian

Dari Gambar 1, maka tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:
1.Pengumpulan dataset.
2. Data di klasifikasi untuk menentukan label category obesitas
3. Melakukan pengujian menggunakan *confuison matrix*.
4. Menampilkan output dari hasil klasifikasi.

1. Pengumpulan dataset

Dataset berat badan obesitas yang digunakan dalam penelitian ini diambil dari https://www.kaggle.com/datasets/denih1dayat/obesitas Dataset ini mencakup informasi tentang parameter–parameter seperti jenis kelamin,, berat badan, tinggi badan, label. Jumlah data yang tersedia dalam dataset ini berjumlah 500, dataset berat badan obesitas disajikan dalam table berikut :

Tabel 1. Dataset Obesitas

id	Jenis Kelamin	Tinggi Badan	Berat Badan	Label
1	Laki - Laki	170	60	1
2	Laki - Laki	190	80	1
3	Perempuan	150	80	4
4	Perempuan	165	50	1
5	Laki - Laki	160	90	5
6	Perempuan	170	70	2
7	Laki - Laki	180	80	2
……	…….	……	……	……..
494	Perempuan	160	40	1
495	Laki - Laki	178	81	3
496	Perempuan	160	80	4
497	Perempuan	172	73	3
498	Laki - Laki	168	70	3
499	Perempuan	170	60	2
500	Laki - Laki	160	100	5

2. Klasifikasi Algoritma Decision Tree

Melakukan Klasifikasi dengan menggunakan algoritma *Decision Tree.* Pengolahan data berat badan menggunakan algoritma *Decision Tree*[10]. Data Obesitas diambil dari Kaggle. Tahapan proses *Decision Tree* di jelaskan pada gambar di bawah



Gambar 2. Kasifikasi Decision Tree

Langkah-langkah atau alur dalam penyelesaian perhitungan manual data mining menggunakan *Decision Tree* sebagai berikut

Alur Algoritma Decision Tree dan contoh perhitungan manualnya

Langkah 0 : Menggunakan Atribut Jenis kelamin, tinggi badan, berat badan dan label

Langkah 1 : Memasukkan sampel data sebagai bahan perhitungan

Tabel 2. Sampel Data

1d	Jenis Kelamin	Tinggi Badan	Berat Badan	Label
1	Laki – Laki	174	96	4
2	Laki – laki	189	87	2
3	Perempuan	185	110	4
4	Perempuan	195	104	3
5	Laki – Laki	149	61	3



Gambar 3 Label kategori obesitas

Langkah 2 : Menghitung *entropy* dari "total" jumlah kasus. Menghitung *entropy* total menggunakan rumus maka akan akan kembali ke awal memasukkan sampel data [5].

Entropy (S) = $\sum_{i=1}^{n} - pi^* pi$ .....(1)

Rumus menghitung nilai entropy

Keterangan :

S : himpunan kasus

A : fitur

n : Jumlah partisi S

pi : proporsi dari Si terhadap S

Gain merupakan salah satu *attribute selection measure* yang digunakan

Entropy (total) = (- 3/14*(10/14))+(- 11/14*(4/14)) = 1.524

Langkah 3 : Klasifikasi data dari atribut :

- id (urutan) : id 0 sampai id 13.
- Jenis kelamin : 0 = laki laki ,1 = perempuan
- Tinggi badan : dari tinggi 147 cm sampai 195 cm
- Berat badan : 51 Kg sampai 111Kg

Langkah 4 : Klasifikasi jumlah kasus, totalnya ada 14 jumlah kasus yang terdiri dari :
id : 0 = 1 kasus, 1-4 = 3 kasus, 5-8 = 4 kasus, 9 = 1 kasus, 10 = 0 kasus, 11-12 = 2 kasus, 13 = 0 kasus Jenis Kelamin 0 = 6 kasus, 1 = 5 kasus
Tinggi Badan 149 – 175 = 7 kasus, 176 – 195 = 4 kasus
Berat badan 50 -70 = 1 kasus, 71 – 90 = 2 kasus, 91 – 120 = 8 kasus,

Langkah 5 : - Klasifikasi ya dan tidak dari 14 total jumlah kasus yang terdiri
dari 11 y dan 3 tidak dari atribut : id 0 = 1 y, 1 = 1 T, 2 – 4 = 3 y, 5 – 8 = 4 y, 9 – 13 = 3 ia dan 2 tidak. Jenis kelamin 0 = 6 ia dan 3 tidak, 1 = 5 ia
Tinggi badan 147 = 1 y, 149 = 1 y, 154 = 1 y, 155 = 1 T, 159 = 1 y, 1= 1 y, 174 = 2 y, 185 = 1 y, 189 = 1 y 1 tidak, 192 = 1 y, 195 = 1 y, 132= 1 tidak,
Berat badan 50 – 70 = 1 , 1 tidak, 71 – 90 = 2 tidak, 91 – 120 = 8 y.

Langkah 6 : Menghitung *entropy id* tiap atribut, Menghitung gain tiap atribut.

3. Pengujian Menggunakan Confusion Matrix

Melakukan pengujian menggunakan *confision matrix* [11], [12]. Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai tersebut adalah *True Positive* (TP), *True Negative* (TN), *False Negative* (FN). Nilai *True Negative* (TN) merupakan data negatif yang terdeteksi benar kelas negatif, sedangkan *False Negative* (FN) merupakan kebalikan dari *False Positive*, sehingga data negatif, namun terdeteksi sebagai data positif. Sementara itu, *True Positive* (TP) merupakan data positif yang terdeteksi benar kelas positif, sedangkan *False Positive* (FP) merupakan kebalikan dari *False Negative*, sehingga data postif, namun terdeteksi sebagai data negatif[13], [14]. Berikut cara untuk menghitung *confusion matrix* :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

$$\text{Presisi} = \frac{TP}{FP+TP} * 100\%$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\%$$

$$\text{F1-score} = 2 * \frac{(recall*presisi)}{(recall+presisi)} * 100\%$$

Dimana :

- TP adalah *True positive*, yaitu jumlah data positif yang terklasifikasi dengan benar adalah data postif oleh sistem.
- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar data negatif oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi data postif oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi data negatif oleh sistem.

HASIL DAN PEMBAHASAN

Pembagian Dataset

Data dalam penelitian akan dibagi menjadi 2 bagian. Yaitu data *training* dan data *testing*. Data *training* nantinya akan digunakan untuk melatih algoritma dalam mencari model yang terbaik. Model tersebut akan diterapkan pada data testing. Hal ini bertujuan untuk menguji kinerja algoritma. Proporsi data *training* dalam *machine learning* umumnya mempuyai komposisi yang lebih besar dari pada data *testing*. Dalam hal ini besaran data yang akan di ambil sebagai data *training* adalah sebanyak 80% dari total data. Sementara itu, data testing memiliki proporsi yang lebih sedikit jika dibandingkan data *training*. Dari kesuluruhan dataset digunakanlah 20% data untuk data testing.

Data *Training*

Menampilkan data *training* pengujian untuk melakukan klasifikasi dari program seperti berikut:

Tabel 3. Data *training*

Id	Jenis kelamin	Tinggi badan	Berat badan
60	1	191	54
226	0	166	70
180	1	162	74
280	0	193	65
17	0	140	129
....	....	....	....
188	1	161	134

71	0	178	117
106	0	166	160
270	0	145	130
102	0	161	155

Dari tabel 3 diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Data *training* adalah fase pada data yang digunakan untuk membangun sebuah model. Data *training* yang berasal dari file *.csv*, kemudian dataset berhasil dibagi ke data training, maka program akan menampilakan sebuah tabel yang berisi data training sebesar 80%, sehingga didapat sebanyak 268 data *training* dan data yang dihasilkan dari program urutanya acak.

Data *Testing*

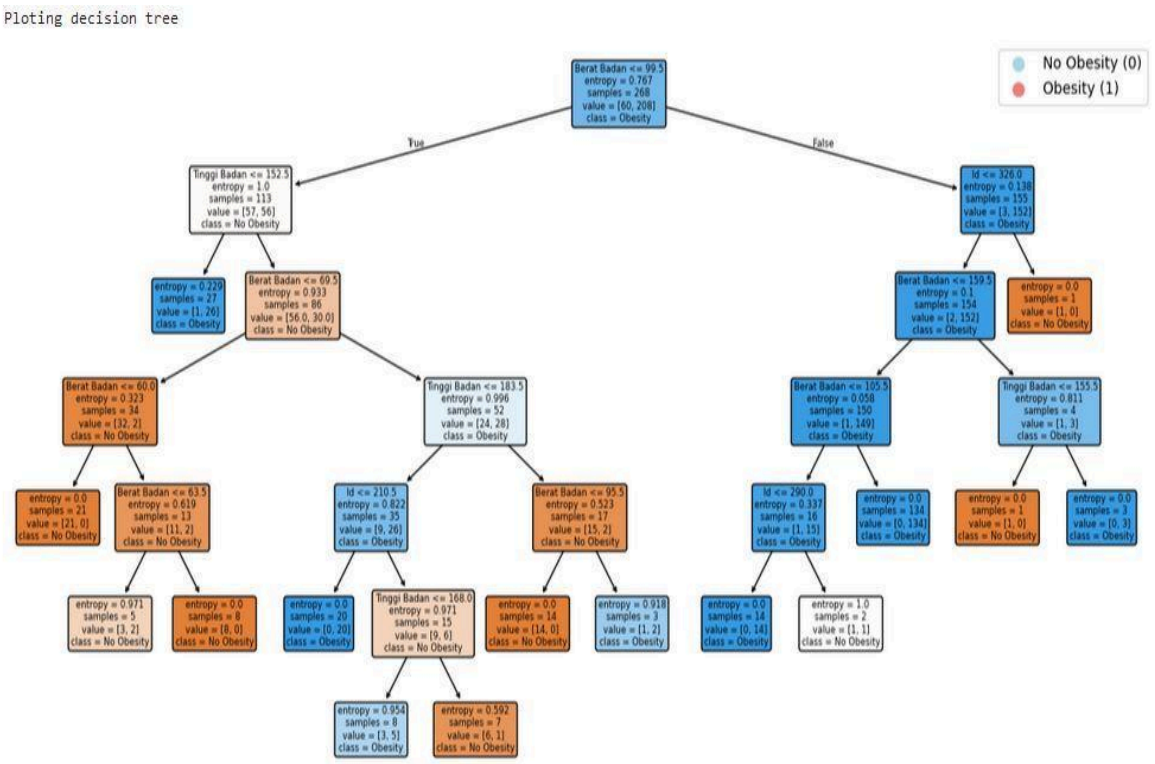
Menampilkan data *testing* pengujian dari program untuk melakukan klasifikasi seperti berikut:

Tabel 4 Data *testing*

id	Jenis kelamin	Tinggi badan	Berat badan
25	1	151	64
109	0	149	66
76	0	165	104
196	0	157	74
57	0	178	127
....	....	....	....
298	1	187	92
294	1	168	143
116	1	165	57
78	1	176	122
244	1	183	50

Dari Tabel 4. diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Proses *testing* merupakan sebuah proses pengujian dimana data yang tidak diketahui kelasnya dimasukkan kedalam program yang telah di bentuk pada saat proses *training* data yang bertujuan untuk prediksi apakah tergolong dalam kelas-kelas obesitas yang telah ditentukan. Pada saat proses data testing ini juga dilakukan dengan menggunakan data yang berasal dari suatu file yang formatnya *.csv,* lalu setelah dataset berhasil dibagi ke data *testing*, maka program akan menampilakan sebuah tabel yang berisi data *testing* sebesar 20%, sehingga didapat sebanyak 67 data *testing* dan data yang dihasilkan dari program urutanya acak.

Menampilkan hasil plot gambar pohon keputusan *decision tree C4.5* dari program seperti berikut :



Gambar 3. Pohon hasil keputusan decision tree C4.5 pada penelitian ini

Menampilkan hasil klasifikasi pada data testing dari program seperti berikut:

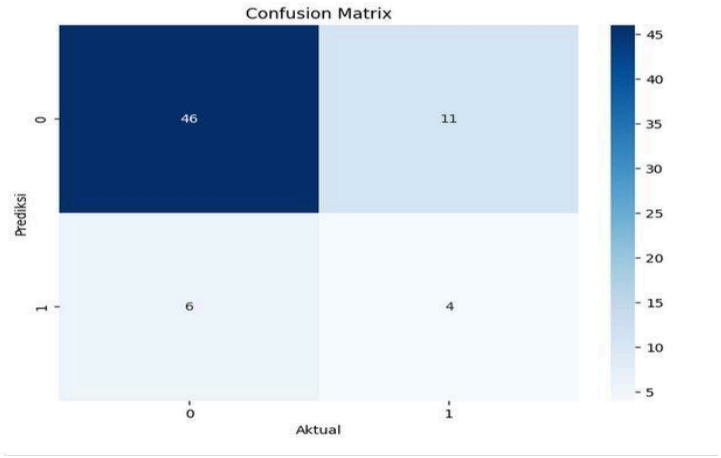
Tabel 5 Hasil klasifikasi

No	Id Acak	Jenis kelamin	Tinggi badan	Berat badan	Label	Predict
1	25	1	151	64	0	1
2	109	0	149	66	1	1
3	76	0	165	104	1	1
4	196	0	157	74	1	1
5	57	0	178	127	1	1
.....	.....	.....	.....	.....	.....	.....
6	298	1	187	92	0	0
7	294	1	168	143	1	1
8	116	1	165	57	0	0
9	78	1	176	122	1	1
10	244	1	183	50	0	0

Dari tabel 5 diatas yang di tampilkan di program hanya *head* dan *tail* nya saja yaitu 5 teratas dan 5 terbawah. Pada saat proses data testing ini juga dilakukan dengan menggunakan data yang berasal dari suatu file yang formatnya *.csv*. Proses klasifikasi pada data testing menghasilkan data yang berbeda, misalnya pada data nomor 25 pasien sebelumnya tidak masuk kedalam kategori obesitas di hasil (label) tapi setelah di klasifikasi ternyata pasien menunjukkan angka 1 di *predicted* yang berarti pasien nomor 25 masuk kedalam kategori obesitas dan seterusnya. Kemudian kita akan menguji hasil klasifikasi menggunakan metode *confusion matrix*.

Confusion Matrix

Hasil implementasi algoritma dapat dilihat melalui *confusion matrix*. Gambar berikut menunjukkan hasil uji algoritma yang ditampilkan dalam *confusion matrix* berikut.



Gambar 4. Hasil pengujian dengan confusion Matrik

Dari gambar *matrix* tersebut dapat diketahui dari 67 data terdapat 37 data yang masuk kedalam kategori obesitas dan 30 data yang tidak masuk kedalam kategori obesitas. Berikut hasil pengujian algoritma yang ditunjukkan sebagai berikut:

$$\text{Accuracy} = \frac{(46+11)}{(46+4+6+11)} = 0,85$$

$$\text{Precision} = \frac{46}{(6+46)} = 0,88$$

$$\text{Recall} = \frac{46}{(4+46)} = 0,92$$

$$\text{F1_Score} = 2 * \frac{(0,92*0,88)}{(0,92+0,88)} = 0,89$$

Precision adalah 0.88, yang berarti dari semua prediksi yang model buat sekitar 88% adalah benar. Hasil *recall* adalah 0.92, yang berarti model berhasil mendeteksi sekitar 92% dari semua contoh kelas yang sebenarnya ada dalam dataset. Semakin tinggi nilai *f1-score* maka model semakin baik. Hasil *f1-score* adalah 0.89. akurasi adalah 0.85, yang berarti model Anda memprediksi dengan benar sekitar 85% dari semua contoh dalam dataset pengujian.

KESIMPULAN

Algoritma mengklasifikasi dengan benar saat hasil uji menggunakan confusion matrix menunjukkan nilai *Accuracy* 0,85 yang menunjukkan klasifikasi dataset pengujian ini sebesar 85% dan 15% tidak terklasifikasi dengan benar. Secara kesuluruhan, algoritma mengklasifikasi dataset berat badan obesitas dengan baik. Hal ini di indikasikan dengan nilai rata rata seperti precision 0,88 yang berarti dari semua prediksi yang model dibuat sekitar 88% adalah benar, kemudian hasil *recall* adalah 0.92, yang berarti model berhasil mendeteksi sekitar 92% dari semua contoh kelas yang sebenarnya ada dalam dataset dan hasil *f1-score* adalah 0.89 yang berarti *f1-score* bernilai 89%, semakin tinggi nilai *f1-score* maka model semakin baik.

DAFTAR PUSTAKA

[1] J. Glass *et al.*, "Management and impact of obesity in Canada: A real-world survey of people with obesity and their physicians," *Obesity Pillars*, vol. 14, Jun. 2025, doi: 10.1016/j.obpill.2025.100171.

[2] T. Indriyani, M. Kurniawan, G. Yuliastuti, C. Prabiantissa, and R. Kembang, "An Improve KNN Method for Classification of Sexually Transmitted Diseases," *2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, vol. 3, pp. 315–319, Sep. 2023.

[3] S. Christensen and C. Nelson, "Chronicity of obesity and the importance of early treatment to reduce cardiometabolic risk and improve body composition," *Obesity Pillars*, p. 100175, Apr. 2025, doi: 10.1016/j.obpill.2025.100175.

[4] Y. Li, E. Herrera-Viedma, G. Kou, and J. A. Morente-Molinera, "Z-number-valued rule-based decision trees," *Inf Sci (N Y)*, vol. 643, Sep. 2023, doi: 10.1016/j.ins.2023.119252.

[5] A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading k-Means clustering and C4.5 decision tree algorithm," in *Procedia Engineering*, 2012, pp. 174–182. doi: 10.1016/j.proeng.2012.01.849.

[6] P. Li, F. Xiong, X. Huang, and X. Wen, "Construction and optimization of vending machine decision support system based on improved C4.5 decision tree," *Heliyon*, vol. 10, no. 3, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25024.

[7] A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading k-Means clustering and C4.5 decision tree algorithm," in *Procedia Engineering*, 2012, pp. 174–182. doi: 10.1016/j.proeng.2012.01.849.

[8] T. Mahmudiono *et al.*, "Integrating Traffic Light Diet System via food analysis in Android app for adolescent nutrition education: A strategy to reduce sugar, salt, and fat consumption," *Clinical Nutrition Open Science*, vol. 59, pp. 206–215, Feb. 2025, doi: 10.1016/j.nutos.2024.10.014.

[9] M. Ng *et al.*, "National-level and state-level prevalence of overweight and obesity among children, adolescents, and adults in the USA, 1990–2021, and forecasts up to 2050," *The Lancet*, Dec. 2024, doi: 10.1016/S0140-6736(24)01548-4.

[10] H. Bin Wang and Y. J. Gao, "Research on C4.5 algorithm improvement strategy based on MapReduce," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 160–165. doi: 10.1016/j.procs.2021.02.045.

[11] T. Indriyani, I. Utoyo, and R. Rulaningtyas, "Comparison of image edge detection methods on potholes road images," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1742-6596/1613/1/012067.

[12] T. Indriyani, S. Nurmuslimah, A. Taufiqurrahman, R. K. Hapsari, C. N. Prabiantissa, and A. Rachmad, "Steganography on Color Images Using Least Significant Bit (LSB) Method," 2023, pp. 39–48. doi: 10.2991/978-94-6463-174-6_5.

[13] T. Indriyani, M. I. Utoyo, and R. Rulaningtyas, "A New Watershed Algorithm for Pothole Image Segmentation," *Studies in Informatics and Control*, vol. 30, no. 3, pp. 131–139, 2021, doi: 10.24846/v30i3y202112.

[14] K. Qazanfari and R. Safabakhsh, "A new steganography method which preserves histogram: Generalization of LSB++," *Inf Sci (N Y)*, vol. 277, pp. 90–101, Sep. 2014, doi: 10.1016/j.ins.2014.02.007.