

## Klasifikasi Obesitas dengan menggunakan Metode Algoritma Decision Tree Interactive Dichotomiser 3 berbasis RapidMiner

Saifulloh<sup>\*1</sup>, Erfia Nadia Safari<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Universitas PGRI Madiun, Madiun

\*Email penulis yang sesuai: [saifulloh@unipma.ac.id](mailto:saifulloh@unipma.ac.id)

### Abstract.

One of the most common health problems worldwide is obesity, a condition that continues to increase every year. Prevention and control of obesity is very important. Identifying the correct obesity category is crucial. The purpose of this study is to create an obesity classification model using the ID3 Decision Tree Algorithm method, based on RapidMiner. This method was chosen because of its ability to create simple yet effective decision tree models by utilizing information gain measurements during the tree formation process. Age, body mass index (BMI), eating habits, physical activity level, and medical history were all obesity risk factors used in this study. RapidMiner is a data analytics platform that enables data processing and analysis with a visual interface. The results show that the ID3 Decision Tree algorithm model can classify obesity categories with high accuracy based on information gain, with operator performance reaching 100% on a sample data set of 110 data attributes. In addition, the resulting model can identify BMI and eating habits, which are the most influential components in obesity classification. These results are expected as the proposed approach model can help health professionals detect the possibility of obesity more efficiently.

Keywords: Health, Obesity, Decision Tree ID3 method, RapidMiner

### Abstrak

Salah satu masalah kesehatan yang paling umum di seluruh dunia adalah obesitas, dimana kondisi ini terus meningkat setiap tahunnya. Pencegahan dan pengendalian obesitas sangat penting. Identifikasi kategori obesitas yang tepat sangat penting. Tujuan penelitian ini adalah untuk membuat model klasifikasi obesitas dengan menggunakan metode Algoritma *Decision Tree ID3*, yang berbasis *RapidMiner*. Metode ini dipilih karena kemampuannya untuk membuat model pohon keputusan yang sederhana namun efektif dengan memanfaatkan pengukuran *gain* informasi selama proses pembentukan pohon. Usia, indeks massa tubuh (IMT), kebiasaan makan, tingkat aktivitas fisik, dan riwayat kesehatan adalah semua faktor risiko obesitas yang digunakan dalam penelitian ini. *RapidMiner* adalah *platform* analitik data yang memungkinkan pengolahan dan analisis data dengan antarmuka visual. Hasil penelitian menunjukkan bahwa model metode algoritma *Decision Tree ID3* dapat mengklasifikasikan kategori obesitas dengan akurasi yang tinggi berdasarkan information gain dengan operator performance mencapai 100% pada data sampel 110 data atribut. Selain itu, model yang dihasilkan dapat menemukan IMT dan kebiasaan makan, yang merupakan komponen yang paling berpengaruh terhadap klasifikasi obesitas. Hasil ini diharapkan sebagai model pendekatan yang diusulkan dapat membantu profesional kesehatan menemukan kemungkinan obesitas secara lebih efisien.

### 1. Pendahuluan

Obesitas telah menjadi salah satu masalah kesehatan utama di Indonesia, dengan *prevalensi* yang terus meningkat setiap tahunnya. Data Riset Kesehatan Dasar (Riskesdas) 2018 menunjukkan bahwa prevalensi obesitas pada orang dewasa di Indonesia mencapai 21,8%, meningkat signifikan dibandingkan tahun-tahun sebelumnya (Kementerian Kesehatan

RI, 2018). Menurut data *World Health Organization* (WHO, 2022), lebih dari 1 miliar orang dewasa di dunia mengalami kelebihan berat badan, dan sekitar 650 juta di antaranya tergolong obesitas. Fenomena ini tidak hanya berdampak pada aspek kesehatan individu, tetapi juga memberikan beban besar terhadap sistem kesehatan nasional serta menurunkan produktivitas masyarakat secara keseluruhan. Kondisi ini tidak hanya mempengaruhi kualitas hidup individu tetapi juga meningkatkan risiko penyakit kronis seperti diabetes, hipertensi, dan penyakit kardiovaskular. Oleh karena itu, klasifikasi obesitas berdasarkan faktor risiko yang komprehensif diperlukan untuk mendukung upaya pencegahan dan pengelolaan kondisi ini secara efektif. Berdasarkan data riset tersebut, dimana kenaikan yang signifikan ini dipicu oleh berbagai faktor, antara lain pola makan tinggi kalori, kurangnya aktivitas fisik, stres, gaya hidup sedentari, dan faktor genetik [1]. Obesitas yang tidak ditangani dengan tepat dapat memicu penyakit tidak menular (PTM) seperti diabetes melitus, penyakit jantung koroner, stroke, dan kanker tertentu, yang berisiko menyebabkan kematian dini.

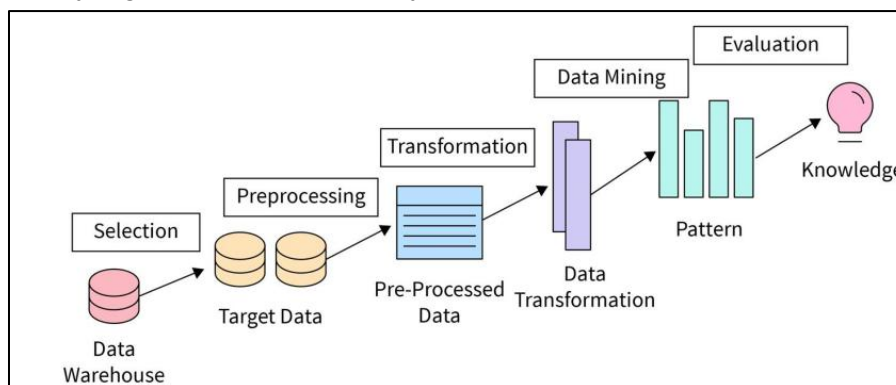
Dalam era digital, analisis data berbasis machine learning telah banyak diterapkan untuk mendukung pengambilan keputusan di bidang kesehatan. Salah satu metode yang sering digunakan adalah *Algoritma Decision Tree*, khususnya ID3 (Iterative Dichotomiser 3), yang mampu membangun model klasifikasi berdasarkan konsep *information gain* [2], [3]. Penelitian sebelumnya oleh Gunawan, dkk (2023) menunjukkan bahwa metode *Decision Tree* memiliki tingkat akurasi yang tinggi dalam klasifikasi data kesehatan, sekaligus menawarkan interpretasi yang mudah dipahami oleh praktisi kesehatan [2].

*RapidMiner* merupakan salah satu perangkat lunak yang banyak digunakan untuk implementasi algoritma *machine learning* [4]. Dengan fitur antarmuka yang intuitif, *RapidMiner* memungkinkan pengolahan data yang cepat dan efisien, sehingga ideal untuk penelitian di bidang kesehatan. Perangkat lunak data science berbasis GUI (*Graphical User Interface*) yang memfasilitasi pembangunan model machine learning tanpa harus menulis kode pemrograman. *RapidMiner* sangat cocok digunakan dalam lingkungan penelitian dan pendidikan karena mendukung pemrosesan data besar, visualisasi proses, dan integrasi dengan berbagai metode analitik [5], [6].

Penelitian ini bertujuan untuk mengembangkan sebuah model klasifikasi obesitas menggunakan metode algoritma *Decision Tree ID3* berbasis *RapidMiner*, nantinya akan mengkaji apa saja faktor-faktor yang berkontribusi terhadap obesitas meliputi Indeks Massa Tubuh (IMT), pola makan, dan tingkat aktivitas fisik, untuk menghasilkan model prediktif yang akurat. Dengan harapan hasil dari penelitian ini dapat menjadi referensi bagi praktisi kesehatan dalam mendeteksi dan menangani obesitas secara lebih efektif.

## 2. Metode

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksploratif dan eksperimental. Tujuan utama dari penelitian ini adalah untuk mengklasifikasikan tingkat obesitas seseorang berdasarkan atribut-atribut tertentu menggunakan Algoritma *Decision Tree Iterative Dichotomiser 3* (ID3) yang diimplementasikan pada perangkat lunak *RapidMiner* [7]. Pendekatan menggunakan metode KDD (*Knowledge Discovery in Database*), terdapat lima (5) tahapan yang dilakukan, diantaranya:



Gambar 1. Metode KDD (*Knowledge Discovery in Database*) [8]

Metode KDD terdiri dari lima (5) tahapan utama, diantaranya:

1. *Selection Data*

Data yang relevan dikumpulkan untuk mendukung proses klasifikasi obesitas. Data dapat berasal dari dataset terbuka seperti UCI *Machine Learning Repository*, atau dari hasil survei kesehatan dengan atribut-atribut yang berkaitan dengan obesitas. Atribut yang dipilih biasanya mencakup: Usia, Jenis kelamin, Indeks Massa Tubuh (IMT), Pola makan (frekuensi makan cepat saji, konsumsi sayur/buah), Aktivitas fisik (frekuensi olahraga), Riwayat keluarga (genetik), Konsumsi alkohol atau merokok. Seleksi dilakukan dengan tujuan menyaring data yang paling relevan dan berkualitas tinggi untuk proses selanjutnya.

2. *Preprocessing Data*

Data yang telah dikumpulkan sering kali mengandung nilai kosong (*missing values*), data duplikat, atau kesalahan input. Oleh karena itu, pada tahap ini dilakukan kegiatan seperti: (1) Pembersihan data (*data cleaning*): Menghapus atau mengisi data yang hilang dengan teknik tertentu seperti mean, modus, atau interpolasi; (2) Normalisasi/standarisasi: Jika diperlukan, data numerik seperti IMT atau usia dapat dinormalisasi agar tidak mendominasi perhitungan dalam algoritma; Pengecekan konsistensi dan validitas data: Untuk memastikan tidak ada data yang berada di luar batas logis, misalnya IMT bernilai negatif atau usia di atas 150 tahun. Pra-pemrosesan ini penting untuk memastikan bahwa data yang digunakan benar-benar siap untuk dianalisis oleh model *machine learning*.

3. *Transformation Data*

Transformasi bertujuan untuk mengubah data mentah menjadi bentuk yang sesuai dengan model klasifikasi. Beberapa kegiatan pada tahap ini antara lain: (1) Pengkodean atribut kategorikal (*encoding*): Misalnya, jenis kelamin dikodekan menjadi “0” untuk laki-laki dan “1” untuk perempuan, atau kebiasaan merokok dikodekan sebagai “ya” dan “tidak”; (2) Pembuatan label kelas (*class labeling*): Data diberi label seperti “Normal”, “Overweight”, “Obese” berdasarkan nilai IMT sesuai dengan standar WHO. Ini menjadi target klasifikasi (kelas output); (3) Penyusunan format data: Menyesuaikan struktur dataset sesuai format input yang diterima oleh *RapidMiner*. Tahapan akan memastikan data dalam format optimal untuk diterapkan pada metode ID3.

4. *Data Mining*

Tahap ini adalah inti dari proses KDD, yaitu penerapan metode *Decision Tree ID3* untuk membangun model klasifikasi obesitas. Aktivitas utama: (1) Penggunaan *RapidMiner* untuk membangun model ID3; (2) Training dan testing model: Dataset dibagi menjadi data latih dan data uji (misalnya dengan *k-fold cross-validation*) untuk mengukur akurasi model; (3) Proses pemilihan atribut berdasarkan information gain: ID3 akan secara otomatis memilih atribut terbaik sebagai simpul akar dan menyusun pohon keputusan [8], [9], [10]. Sehingga, Model ini akan mempelajari pola dari data latih untuk dapat memprediksi kelas obesitas dari data baru.

5. *Evaluation*

Setelah model dibangun, tahap ini bertujuan untuk mengevaluasi performa dan memahami hasil klasifikasi: (1) Evaluasi akurasi model: Menggunakan metrik seperti akurasi, *precision*, *recall*, dan *confusion matrix*; (2) Visualisasi pohon keputusan: *RapidMiner* menyediakan tampilan visual dari pohon keputusan yang terbentuk, memudahkan interpretasi; (3) Analisis atribut penting: Mengetahui atribut mana yang paling berpengaruh dalam menentukan status obesitas, misalnya IMT atau aktivitas fisik. Interpretasi ini sangat penting untuk mengubah hasil data mining menjadi wawasan yang dapat digunakan untuk pengambilan keputusan di bidang Kesehatan [11].

Data yang digunakan dalam penelitian ini berasal dari dataset “*Obesity Levels*” yang tersedia secara publik di platform *Kaggle* (<https://www.kaggle.com/datasets>). Dataset *Obesity Levels* di Kaggle berisi sekitar 2111 data entri dan pada penelitian ini menggunakan rasio 70:30 (Training: 1477 data, Testing: 634 data).

### 3. Hasil dan Pembahasan

Hasil dan diskusi dapat dilakukan secara keseluruhan yang berisi temuan dan penjelasan penelitian.

#### 3.1. Penyajian Data

Data yang digunakan dalam penelitian ini berasal dari dataset publik berjudul “*Obesity Levels Dataset*” yang tersedia di *UCI Machine Learning Repository*. Dataset ini terdiri dari 2111 data responden dengan berbagai atribut yang memengaruhi status obesitas dapat dilihat pada tabel 1 dibawah ini:

Tabel 1. Data Atribut

No	Umur	Jenis Kelamin	Tinggi Badan	Berat Badan	BMI	Kategori
1	25	Laki-laki	175	80	25.3	Normal Weight
2	30	Perempuan	160	60	22.5	Normal Weight
3	35	Laki-laki	180	90	27.3	Overweight
4	40	Perempuan	150	50	20	Underweight
5	45	Laki-laki	190	100	31.2	Obese
...	...	...	...	...	...	...
105	57	Perempuan	120	25	10	Underweight
106	11	Laki-laki	175	10	3.9	Underweight
107	16	Perempuan	160	10	3.9	Underweight
108	21	Laki-laki	180	15	5.6	Underweight
109	26	Perempuan	150	15	5.6	Underweight
110	31	Laki-laki	190	20	8.3	Underweight

#### 3.2. Pemrosesan Data

Data yang sudah terkumpul perlu melakukan data cleaning atau pembersihan data. Proses ini mencakup menghilangkan data ganda, memastikan data konsisten, dan memperbaiki kesalahan. Pada variabel “No” perlu dihilangkan karena tidak memberikan kontribusi signifikan dalam analisis data dan justru bisa mengurangi efektivitas pemodelan.

Tabel 2. Pembersihan Data

Variabel	Missing Value
Umur	0
Jenis Kelamin	0
Tinggi Badan	0
Berat Badan	0
BMI	0
Kategori	0

Beberapa variabel pada dataset diberikan *role* (peran) agar algoritma dapat fokus hanya pada variabel yang relevan untuk analisis. Peran variabel mempengaruhi bagaimana variabel tersebut diproses, seperti normalisasi pada input atau *encoding* pada variabel target. Pada variabel kategori diberikan *role label* dan variabel berat badan diberikan *role weight*.

Setelah dilakukan pelabelan data, data perlu dikelompokkan berdasarkan kategori yang akan diteliti. Dalam penelitian ini, data dikelompokkan menjadi empat kategori berdasarkan jenis berat badan, yaitu:

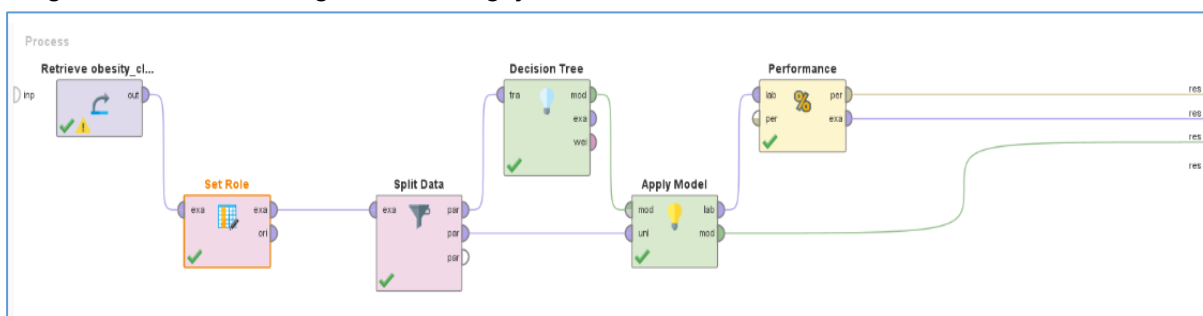
1. *Normal Weight*: menunjukkan berat badan normal.
2. *Underweight*: menunjukkan kekurangan berat badan atau berat badan kurang ideal.
3. *Overweight*: menunjukkan kelebihan berat badan.
4. *Obese*: menunjukkan berat badan memiliki kategori obesitas.

	ID <i>integer</i>	Umur <i>integer</i>	Jenis Kelamin <i>polynomial</i>	Tinggi Badan <i>integer</i>	Berat Badan <i>integer weight</i>	BMI <i>real</i>	Kategori <i>polynomial label</i>
1	1	25	Laki-laki	175	80	25.300	Normal Weight
2	2	30	Perempuan	160	60	22.500	Normal Weight
3	3	35	Laki-laki	180	90	27.300	Overweight
4	4	40	Perempuan	150	50	20.000	Underweight
5	5	45	Laki-laki	190	100	31.200	Obese
6	6	50	Perempuan	140	40	16.700	Underweight
7	7	55	Laki-laki	200	110	34.200	Obese
8	8	60	Perempuan	130	30	13.300	Underweight
9	9	65	Laki-laki	210	120	37.200	Obese
10	10	70	Perempuan	120	20	10.000	Underweight
11	11	18	Laki-laki	175	70	23.400	Normal Weight
12	12	23	Perempuan	160	50	20.000	Underweight
13	13	28	Laki-laki	180	80	25.300	Normal Weight
14	14	33	Perempuan	150	60	22.500	Normal Weight
15	15	38	Laki-laki	190	90	27.300	Overweight
16	16	43	Perempuan	140	50	20.000	Underweight
17	17	48	Laki-laki	200	100	31.200	Obese
18	18	53	Perempuan	130	40	16.700	Underweight

Gambar 2. Pemberian *Role* dan Pelabelan variabel

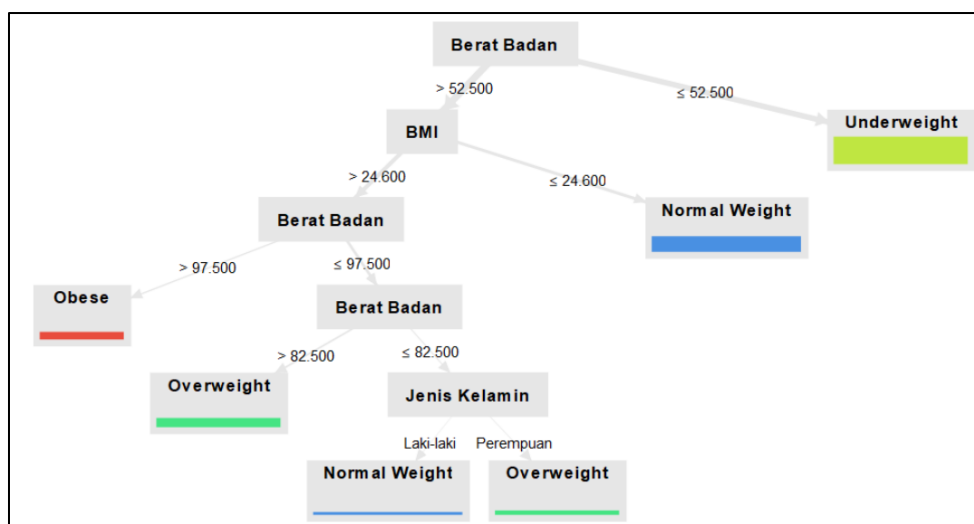
### 3.3. Pemodelan Algoritma Decision Tree

Pengujian klasifikasi obesitas pada *RapidMiner* menggunakan komponen *Set Role*, *Split Data*, *Algoritma Decision Tree*, *Apply Model*, dan *Performance (Classification)*. Pada komponen *Set Role*, penentuan variabel kategori diklasifikasikan ke dalam *role* label. Sementara pada variabel lainnya seperti BMI, umur, tinggi badan, berat badan dan jenis kelamin diklasifikasikan ke dalam atribut regular. *Split data* yang digunakan untuk membagi dataset diatur rasionya menjadi 70% data Training dan 30% data Testing. *Split data* tersebut berguna untuk membangun dan menguji model.



Gambar 3. Proses Pengolahan (Rapidminer)

Selanjutnya, pada komponen *Decision Tree* menggunakan parameter *Information Gain*. Parameter tersebut digunakan untuk menentukan tingkat pengaruh setiap data terhadap hasil akhir [12], [13]. Pohon keputusan menggunakan *Information Gain* untuk memilih atribut dengan kontribusi terbesar dalam mengurangi ketidakpastian (*entropi*). Berat Badan dipilih sebagai *root node* karena memiliki *Information Gain* terbesar. BMI digunakan untuk data yang berada di kisaran tertentu. Sedangkan jenis kelamin digunakan sebagai atribut terakhir pada subset tertentu.



Gambar 4. Decision Tree

Pohon yang dihasilkan dari ranting decision tree dapat dinyatakan dalam bentuk IF THEN sebagai berikut. Dari ranting decision tree didapat rule pola baru yaitu:

1. RULE 1 IF Berat Badan  $\leq 52.500$  THEN Kategori = Underweight
2. RULE 2 IF Berat Badan  $> 52.500$  AND BMI  $\leq 24.600$  Kategori = Normal Weight
3. RULE 3 IF Berat Badan  $> 52.500$  AND BMI  $> 24.600$  AND Berat Badan  $> 97.500$  Kategori = Obese
4. RULE 4 IF Berat Badan  $> 52.500$  AND BMI  $> 24.600$  AND Berat Badan  $\leq 97.500$  AND Berat Badan  $> 82.500$  THEN Kategori = Overweight
5. RULE 5 IF Berat Badan  $> 52.500$  AND BMI  $> 24.600$  AND Berat Badan  $\leq 82.500$  AND Jenis Kelamin = Laki-laki THEN Kategori = Normal weight
6. RULE 6 IF Berat Badan  $> 52.500$  AND BMI  $> 24.600$  AND Berat Badan  $\leq 82.500$  AND Jenis Kelamin = Perempuan THEN Kategori = Overweight

Model ini menghasilkan sejumlah aturan logika *if-then* yang mudah dipahami untuk membantu dokter mengelompokkan pasien berdasarkan kategori berat badan seperti *underweight*, *normal weight*, *overweight*, dan *obese*. Hasil studi menunjukkan bahwa metode *Decision Tree ID3*, yang digunakan oleh *RapidMiner*, adalah alat yang efektif dan efisien dalam analisis data kesehatan.

### 3.4. Evaluasi

Dalam penelitian ini, *algoritma decision tree* dengan kriteria pemisahan data berdasarkan *information gain* dengan operator *performance* yang digunakan menunjukkan bahwa model yang dihasilkan memiliki tingkat akurasi sebesar 100,00%.

accuracy: 100.00%					
	true Normal Weight	true Overweight	true Underweight	true Obese	class precision
pred. Normal Weight	9	0	0	0	100.00%
pred. Overweight	0	6	0	0	100.00%
pred. Underweight	0	0	14	0	100.00%
pred. Obese	0	0	0	4	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	

Gambar 5. Hasil Akurasi Operator Performance

Berdasarkan hasil akurasi *operator Performance (Classification)* di *RapidMiner* menunjukkan akurasi 100,00% dapat diartikan: 1) Semua data uji (testing data) berhasil diklasifikasikan dengan benar oleh model, 2) Tidak ada prediksi yang salah dari model → *zero error* pada set testing. Kondisi tersebut bisa mengalami *overfitting* jika data tidak terdistribusi normal pada sampel dataset dan jika terjadinya *overfitting* maka harus dilakukan *one hot and coding* pada datasetnya.



#### 4. Kesimpulan

Pemodelan menggunakan metode *Decision Tree ID3* yang dibangun oleh *RapidMiner*, penelitian ini menghasilkan model klasifikasi obesitas yang sangat akurat dengan tingkat akurasi 100% namun kondisi tersebut bisa mengalami overfitting jika data tidak terdistribusi normal pada sampel dataset dan jika terjadinya overfitting maka harus dilakukan *one hot and coding* pada datasetnya. Proses klasifikasi didasarkan pada faktor risiko obesitas, termasuk Indeks Massa Tubuh (IMT), berat badan, tinggi badan, umur, dan jenis kelamin. Dalam membangun pohon keputusan, algoritma ID3 menggunakan information gain untuk memilih fitur yang paling relevan. Variabel berat badan dan IMT adalah fitur utama dalam klasifikasi obesitas.

#### 5. Referensi

- [1] A. P. Ayudhitama and Utomo Pujiyanto, "Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer," *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, 2020, doi: 10.33795/jip.v6i2.274.
- [2] I. M. Agus Oka Gunawan, I. D. A. Indah Saraswati, I. D. G. Riswana Agung, and I. P. Eka Putra, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 2, pp. 73–83, 2023, doi: 10.47233/jteksis.v5i2.775.
- [3] M. R. Andriansyah, E. Santoso, and Sutrisno, "Klasifikasi Risiko Hipertensi Menggunakan Fuzzy Decision Tree Iterative Dichotomiser 3 (ID3)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. Vol 2 No 12 (2018), pp. 7088–7096, 2018.
- [4] M. Fadhillah, R. Wandri, A. Hanafiah, P. R. Setiawan, Y. Arta, and S. Daulay, "Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa," *J. Informatics Manag. Inf. Technol.*, vol. 5, no. 1, pp. 40–47, 2025, doi: 10.47065/jimat.v5i1.473.
- [5] P. P. P. A. N. . F. I. R. . Zer, D. Hartama, and S. R. Andani, "Analisa Faktor Dominan Mahasiswa Kesulitan Memahami Bahasa Pemrograman Menggunakan Metode C4.5," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 492, 2019, doi: 10.30645/senaris.v1i0.55.
- [6] I. N. Kholifah, "Memprediksi Tingkat Kelulusan Peserta Kursus Pertama dengan Algoritma Data Mining C4.5 dan RapidMiner," *J. Inform. dan Sist. Inf.*, vol. 6, no. 1, pp. 1–10, 2020.
- [7] R. A. Saputra *et al.*, "Detecting Alzheimer's Disease by the Decision Tree Methods Based on Particle Swarm Optimization," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, pp. 1–7, 2020, doi: 10.1088/1742-6596/1641/1/012025.
- [8] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [9] T. Hendrickx, B. Cule, P. Meysman, S. Naulaerts, K. Laukens, and B. Goethals, "Mining association rules in graphs based on frequent cohesive itemsets," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9078, no. 3, pp. 637–648, 2015, doi: 10.1007/978-3-319-18032-8\_50.
- [10] I. Junaedi, N. Nuswantari, and V. Yasin, "Perancangan Dan Implementasi Algoritma C4 . 5 Untuk Data Mining," *J. Inf. Syst. Informatics Comput.*, vol. 3, no. 1, pp. 29–44, 2019.
- [11] M. Muhsi, "Model dan Analisa Faktor Eksternal Aktifitas Siswa Kelas X TKJ SMKN 1 Pakong Pamekasan Menggunakan Algoritma Decision Tree," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 92–106, 2021, doi: 10.31102/jatim.v2i2.1239.
- [12] M. Muhsi and S. Suprpto, "Jurnal Informatika : Jurnal pengembangan IT Modifikasi Gain Ratio Pada Algoritma C4 . 5 dengan Nilai Koefisien Determinasi untuk Prediksi Kelulusan Mahasiswa ( Studi Kasus : Universitas Islam Madura )," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 164–177, 2025, doi: 10.30591/jpit.v9ix.xxx.

- [13] M. Mirqotussa'adah, M. A. Muslim, E. Sugiharti, B. Prasetyo, and S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 2, pp. 135–143, 2017, doi: 10.24843/lkjiti.2017.v08.i02.p07.