

IMPLEMENTASI *DATA MINING* UNTUK KLASIFIKASI PENYAKIT ASAM URAT MENGUNAKAN ALGORITMA C4.5

Implementation of Data Mining for Gout Classification Using C4.5 Algorithm

Gianovita Talarima^{1*}, Ferry Kondo Lembang², Norisca Lewaherilla³, Johan Bruiyf Bension⁴

^{1,2,3} Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pattimura

⁴ Program Studi Ilmu Pendidikan Kedokteran, Fakultas Kedokteran, Universitas Pattimura
Jl. Ir. M. Putuhena, Kampus Unpatti Poka, Ambon, 97233, Indonesia

E-mail Coresponding Author: talarimagianovita@gmail.com

Abstrak: Penyakit asam urat merupakan masalah kesehatan di seluruh dunia khususnya Indonesia dengan prevalensi mencapai angka 11,9%. Angka tersebut akan semakin meningkat jika adanya ketidaktahuan masyarakat umum tentang faktor-faktor yang dapat memicu terkena penyakit asam urat. Penelitian ini menggunakan data *Screening Civitas Akademika UKIM* yang dilakukan untuk membuat model klasifikasi menggunakan *data mining* Algoritma C4.5 yang menghasilkan sebuah pohon keputusan serta pengujian yang dilakukan dengan menggunakan program R. Dari seluruh data berjumlah 277 data dibagi menjadi 198 data *training* dan 79 data *testing*. Dalam penelitian ini terdapat beberapa atribut klasifikasi yaitu, jenis kelamin, konsumsi alkohol, konsumsi gula berlebihan, usia, nilai IMT, dan riwayat penyakit. Berdasarkan perhitungan diperoleh hasil akurasi setelah *post-pruning* adalah 78% untuk data *training* dan 70% untuk data *testing*.

Kata Kunci: Algoritma C4.5, Klasifikasi, Penyakit Asam Urat

Abstract: Gout is a health problem throughout the world, especially Indonesia, with a prevalence reaching 11.9%. This figure will increase if there is general public ignorance about the factors that can trigger gout. In this study, the Screening Data of the Academic Community UKIM was carried out to create a classification model using the C4.5 Algorithm data mining, which produced a decision tree and tests were carried out using the R program. From all data, 277 data were divided into 198 training data and 79 data. testing. In this study there were several classification attributes, namely, gender, alcohol consumption, excessive sugar consumption, age, BMI value, and medical history. Based on the calculation, the accuracy after post-pruning is 78% for training data and 70% for testing data.

Keywords: Algorithm C4.5, Classification, Gout

1. PENDAHULUAN

Data mining merupakan satu dari sekian banyak bentuk implementasi yang digunakan untuk mendapatkan sebuah model dan pola yang mampu melakukan prediksi pada suatu data berdasarkan data sebelumnya di periode waktu tertentu [1]. *Data mining* merupakan suatu proses menggali sekumpulan data dan mengubahnya dalam bentuk informasi yang bermanfaat bagi pengguna [2]. Umumnya beberapa pekerjaan yang dapat dilakukan oleh *data mining* antara lain deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi. Akan tetapi, pekerjaan yang paling sering digunakan dalam *data mining* yaitu prediksi dan klasifikasi dimana dilakukan analisis data yang mampu memberi gambaran kelas data atau untuk memprediksi data di masa yang akan datang [3].

Salah satu algoritma dalam *data mining* yang dapat digunakan untuk pekerjaan klasifikasi dan prediksi yaitu Algoritma C4.5, yang didasarkan pada pembentukan pohon keputusan [4]. Pohon keputusan merupakan model prediksi dan klasifikasi yang paling baik digunakan sebab menggunakan struktur pohon atau struktur berhirarki, dimana konstruksinya relatif cepat dan model yang dihasilkan sangat mudah dimengerti [5].

Prediksi terhadap penyakit asam urat menjadi kasus yang diangkat dalam penelitian ini. Asam urat adalah asam berbentuk kristal yang merupakan produk akhir dari metabolisme atau pemecahan purin [6]. Purin bersumber dari tubuh yang secara alamiah diproduksi dan juga dapat bersumber dari asupan makanan seperti tanaman atau hewan [7]. Peran asam urat dalam tubuh sebenarnya dapat sebagai antioksidan yang berguna meregenerasi sel. Akan tetapi, bisa menyebabkan masalah apabila kadar dalam tubuh melampaui batas normal

[8]. Beberapa kondisi fisik bagian tubuh yang terdampak akibat asam urat melampaui batas normal antara lain rasa nyeri dan bengkak pada persendian yang menyebabkan susah jalan [9]. Normalnya, kadar asam urat dalam darah manusia berkisar antara 3,4 – 7,0 mg/dL pada laki laki, pada perempuan antara 2,4 – 6,0 mg/dL, dan 2,0 – 5,5 mg/dL pada kelompok anak-anak [10]

Dari data *World Health Organization* (WHO), ditemukan prevalensi penyakit asam urat di dunia mencapai angka 34,2%. Kejadian asam urat paling banyak terjadi pada negara-negara maju, semisal Amerika Serikat dimana prevalensi mencapai angka 26, 3% dari keseluruhan penduduk. Tidak hanya negara-negara maju, prevalensi kejadian asam urat di negara-negara berkembang seperti Indonesia juga mengalami peningkatan [11]. Menurut data Riset Kesehatan Dasar disampaikan bahwa prevalensi penyakit asam urat berdasarkan diagnosa tenaga kesehatan di Indonesia mencapai angka 11,9%. Provinsi Maluku merupakan salah satu provinsi dengan prevalensi asam urat paling tinggi dari prevalensi rata-rata nasional, dimana terjadi peningkatan mencapai angka 8,9% sedangkan dari diagnosis dan gejala meningkat mencapai angka 18,8% [12]. Dampak buruk yang ditimbulkan oleh penyakit asam urat apabila tidak segera diobati antara lain dapat berakibat pada kelainan bentuk tulang serta komplikasi gangguan ginjal, jantung, diabetes mellitus, stroke, dan osteoporosis [13].

Berdasarkan hasil data praktek *screening* penyakit tidak menular oleh mahasiswa Program Studi Kesehatan Masyarakat UKIM, diperoleh hasil dari 250 sampel terdapat 167 sampel yang memiliki kadar asam urat di atas kadar normal. Hal ini menunjukkan bahwa prevelensi kasus kadar asam urat berat pada mahasiswa, bapak ibu pegawai, dan dosen di lingkup Civitas Akademika UKIM dapat dikategorikan cukup tinggi [14].

Beberapa penelitian terdahulu terkait diantaranya [15], yang merancang sistem pendiagnosa penyakit hiperkolesterol menggunakan metode pohon keputusan (*decision tree*) dengan menggunakan Algoritma C4.5. Sistem tersebut memberikan hasil berdasarkan beberapa kriteria yaitu IMT (Indeks Massa Tubuh), LDL (*Low Density Lipoprotein*), HDL (*High Density Lipoprotein*), TC (*Total Cholesterol*), TGA (Trigeliserida), dan tingkatan perokok, yang mana diperoleh hasil berupa penggolongan tingkatan kolesterol yaitu: Hiperkilomikronemia, Hiperkolesterol, Hiperlipoproteinemia tipe III, Hiperlipoproteinemia tipe IV dan Hiperlipoproteinemia Tipe V. Dari data *training* yang berjumlah 900 data dan data *testing* berjumlah 20 data diperoleh *error* sebesar 5%. Adapun penelitian lainnya terkait algoritma C4.5 untuk pengkalsifikasian cara keluar pasien ginjal kronis berdasarkan rekam medis BPJS Kesehatan di RSUAM [17]. Dihilangkan 11 aturan yang dapat dijadikan pola dalam menentukan pasien yang paling berpotensi meninggal dengan akurasi yang cukup besar yaitu 91% sebelum dilakukan post-pruning dan 88% setelah dilakukan post-pruning. Beranjak dari penelitian-penelitian tersebut, peneliti tertarik untuk tujuan menerapkan algoritma C4.5 dalam memprediksi secara diagnostik pasien penyakit asam urat berdasarkan beberapa input dan beberapa faktor pendukung. Penggunaan salah satu teknik *data mining* yaitu klasifikasi diharapkan mampu melakukan prediksi status seseorang apakah mengidap penyakit asam urat atau tidak [18]. Hasil penelitian ini diharapkan dapat membantu Dinas Kesehatan Provinsi Maluku untuk menentukan status asam urat seseorang.

2. METODOLOGI

Pada penelitian ini data yang digunakan merupakan data sekunder yang diperoleh dari data *screening* Civitas Akademika Universitas Kristen Indonesia Maluku (UKIM) pada 24 Juni 2022. Jumlah data yang digunakan untuk analisis sebanyak 277 data. *Event* pada penelitian ini adalah pasien penderita penyakit asam yang terbagi dalam tiga tingkatan kadar asam urat yaitu, berat, ringan dan tidak ada.

2.1. Variabel Penelitian

Variabel penelitian yang digunakan tercantum pada Tabel 1, yang dibagi berdasarkan standar yang ditetapkan oleh Kementerian Kesehatan RI. Tingkat kadar asam urat laki-laki normalnya berkisar pada 3,4-7,0 mg/dL. Jadi apabila seorang pasien jenis kelamin laki-laki dengan kadar asam urat < 3,4 maka dikategorikan dalam kelas “tidak ada”, berada dalam kisaran 3,4-7,0 dikategorikan dalam kelas “ringan”, dan jika > 7,0 maka dikategorikan dalam kelas “berat”.

Kejadian yang sama juga berlaku untuk pasien jenis kelamin perempuan. Tingkat kadar asam urat perempuan normalnya berkisar pada 2,4-6,0 mg/dL. Jadi apabila seorang pasien jenis kelamin perempuan dengan kadar asam urat < 2,4 maka dikategorikan dalam kelas “tidak ada”, berada dalam kisaran 2,4-6,0 dikategorikan dalam kelas “ringan”, dan jika > 6,0 maka dikategorikan dalam kelas “berat”.

Tabel 1. Variabel Penelitian

	Variabel	Defenisi	Pembagian Kelas	
Dependen	Y = Kadar asam urat	Tingkat kadar asam urat pasien (Ringan/Berat/Tidak ada)		
	X ₁ = Jenis kelamin	Jenis kelamin pasien (Perempuan/Laki-laki)		
	X ₂ = Konsumsi alkohol	Konsumsi alkohol pasien (Ya/Tidak)		
Independen	X ₃ = Konsumsi gula berlebihan	Tingkat konsumsi gula pasien (Normal/Kurang/Lebih)	<110mg/dL 110-180mg/dL >180mg/dL	Kurang Normal Lebih
	X ₄ = Usia	Usia pasien (Remaja/Dewasa/Lansia)	<26 tahun 26-45 tahun >45 tahun	Remaja Dewasa Lansia
	X ₅ = Nilai IMT	Berat IMT pasien (Normal/Kurus/Gemuk)	<18,5 18,5-25,0 >25,0	Kurus Normal Gemuk
	X ₆ = Riwayat penyakit	Riwayat penyakit pasien yang berpengaruh atau tidak (0=tidak ada, 1=penyakit yang berpengaruh, 2=penyakit yang tidak berpengaruh)		

2.2. Tahapan Penelitian

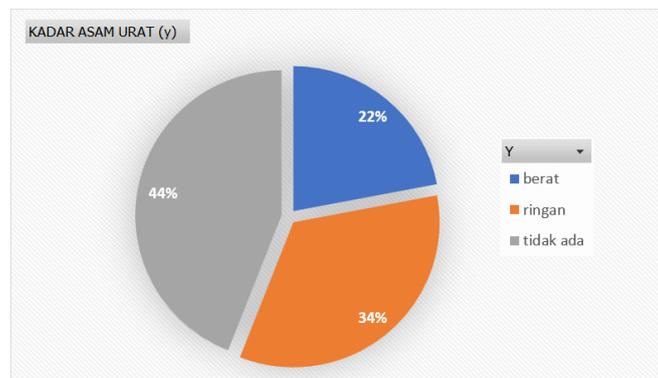
Tahapan dalam penelitian ini dijelaskan sebagai berikut:

- Melakukan pengumpulan data sekunder pada data *screening* asam urat Civitas Akademika UKIM.
- Mendeskripsikan karakteristik pasien berdasarkan faktor-faktor yang diduga mempengaruhi sebagai variabel penelitian.
- Melakukan tahapan *decision tree*.
- Menentukan data *training*.
- Menghitung nilai *entropy*.
- Menghitung nilai *gain*.
- Menentukan *root node*.
- Proses partisipasi pohon keputusan akan berhenti saat semua cabang dalam *node* N mendapat kelas yang sama.
- Melakukan pengujian model untuk membangun struktur (pola) pohon keputusan melalui metode *decision tree* C4.5.
- Menganalisis hasil klasifikasi menggunakan *confusion matrix*.
- Menarik kesimpulan dan saran.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Deskriptif

Analisis deskriptif untuk penelitian ini ditunjukkan pada Gambar 1 yang dibuat berdasarkan pengolahan data menggunakan Ms. Excel untuk ketujuh variabel. Berdasarkan Gambar 1, deskriptif dari variabel tingkat kadar asam urat (y) dengan total $n = 277$, dimana pasien dengan tingkat kadar asam urat ringan 34%, berat 22%, dan tidak ada 44%.



Gambar 1. Distribusi Pasien Berdasarkan Ringkat Kadar Asam Urat (y)

Selanjutnya, ditunjukkan pula tabel kontingensi untuk variable kadar asam urat dan jenis kelamin pada Tabel 2. Dimana dari 277 total pasien penyakit asam urat, sebagian besar penderita asam urat yaitu pada jenis kelamin perempuan dengan tingkat kadar asam urat ringan 56 pasien dan berat 29 pasien. Kemudian pada jenis kelamin laki-laki, tingkat asam urat ringan, 38 pasien dan berat 32 pasien.

Tabel 2. Tabel Kontingensi Variabel Kadar Asam Urat dan Jenis Kelamin

Kadar Asam Urat	Jenis Kelamin		Total
	Perempuan	Laki-laki	
Ringan	56	38	94
Berat	29	32	61
Tidak ada	78	44	122
Total	163	113	277

Sedangkan pada Tabel 3, dapat dilihat bahwa pada tabel kontingensi untuk variabel kadar asam urat dan konsumsi alkohol, sebagian besar pasien tidak mengonsumsi alkohol yaitu dengan jumlah 234 pasien. Hanya terdapat 13 pasien dengan kadar asam urat berat dan 11 pasien kadar asam urat ringan yang aktif mengonsumsi alkohol.

Tabel 3. Tabel Kontingensi Variabel Kadar Asam Urat dan Konsumsi Alkohol

Kadar Asam Urat	Konsumsi Alkohol		Total
	Ya	Tidak	
Ringan	11	83	94
Berat	13	48	61
Tidak ada	19	103	122
Total	43	234	277

Lebih lanjut pada Tabel 4, yaitu tabel kontingensi variabel kadar asam urat dan konsumsi gula berlebihan diketahui bahwa, pasien mengonsumsi gula berlebihan lebih banyak terdapat pada kondisi kadar asam urat ringan dengan total 94 pasien, dimana 61 pasien tingkat konsumsi gula normal, 29 pasien tingkat konsumsi gula kurang dan 4 pasien tingkat konsumsi gula lebih.

Tabel 4. Tabel kontingensi variabel kadar asam urat dan konsumsi gula berlebihan

Kadar Asam Urat	Konsumsi Gula Berlebihan			Total
	Normal	Kurang	Lebih	
Ringan	61	29	4	94
Berat	43	13	5	61
Tidak ada	101	16	5	122
Total	205	58	14	277

Tabel 5 berikut ini merupakan tabel kontingensi untuk variabel kadar asam urat, dan usia. Dimana pasien untuk usia remaja memiliki jumlah lebih banyak yaitu 147 pasien. Untuk kadar asam urat ringan 49 pasien, berat 8 pasien dan tidak ada 90 pasien.

Tabel 5. Tabel Kontingensi Variabel Kadar Asam Urat dan Usia

Kadar Asam Urat	Usia			Total
	Remaja	Dewasa	Lansia	
Ringan	49	29	16	94
Berat	8	23	30	61
Tidak ada	90	30	2	122
Total	147	82	48	277

Tabel 6 adalah tabel kontingensi untuk variabel kadar asam urat, dan nilai IMT. Dapat dilihat dari 277 pasien, paling banyak pasien penyakit asam urat dengan indeks massa tubuh normal. Dimana pasien dengan kadar asam urat ringan berjumlah 55 pasien, berat 27 pasien, dan tidak ada 61 pasien.

Tabel 6. Tabel Kontingensi Variabel Kadar Asam Urat dan Nilai IMT

Kadar Asam Urat	Nilai IMT			Total
	Normal	Kurus	Gemuk	
Ringan	55	11	28	94
Berat	27	2	32	61
Tidak ada	61	32	29	122
Total	143	45	89	277

Kemudian pada Tabel 7 ditunjukkan tabel kontingensi untuk variabel kadar asam urat dan riwayat penyakit. Bahwa pasien dengan kadar asam urat berat untuk penyakit yang berpengaruh adalah pasien paling banyak yaitu 61 pasien dan diikuti dengan pasien kadar asam urat ringan untuk yang tidak memiliki riwayat penyakit sebanyak 41 pasien.

Tabel 7. Tabel Kontingensi Variabel Kadar Asam Urat dan Riwayat Penyakit

Kadar Asam Urat	Riwayat Penyakit			Total
	0	1	2	
Ringan	41	14	39	94
Berat	0	61	0	61
Tidak ada	62	0	60	122

Hasil statistik uji antara variabel dependen dan setiap variabel independen dapat dilihat pada Tabel 8 berikut.

Tabel 8. Hasil Statistik Uji Antara Variabel Dependen dan Variabel Independen

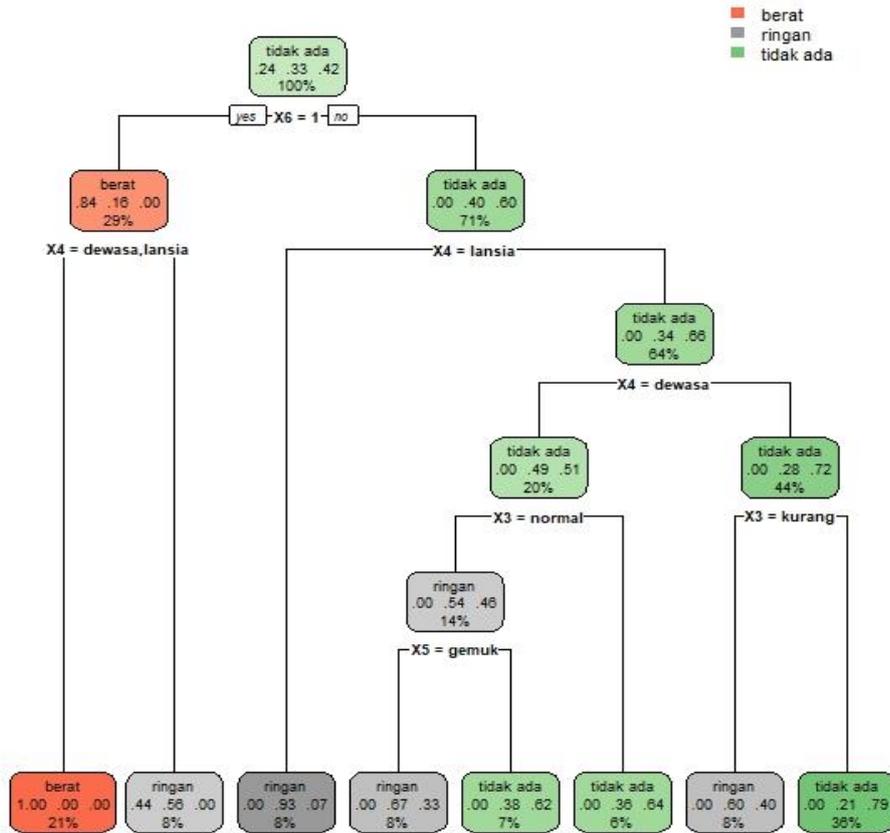
Variabel Independen	P-Value	Alpha (α)	Keputusan
Jenis kelamin	0.103	>	Gagal tolak H_0
Konsumsi alkohol	0.272	>	Gagal tolak H_0
Konsumsi gula berlebihan	0.018	<	Tolak H_0
Usia	0.000	<	Tolak H_0
Nilai IMT	0.000	<	Tolak H_0
Riwayat penyakit	0.000	<	Tolak H_0

3.2. Penerapan *Decision Tree* C4.5 Untuk Klasifikasi Penyakit Asam Urat

Hasil pembentukan pohon keputusan setelah melakukan proses *post-pruning* dapat dilihat pada Gambar 2, dimana terdapat *node number* dari model *decision tree* untuk membantu menjelaskan struktur dan alur keputusan yang dihasilkan oleh model ini. Dapat dilihat pada Gambar 2 bahwa kelas “tidak ada” menjadi *node number* 1 untuk variabel riwayat penyakit yang berpengaruh. Sementara hasil *confusion matrix* yang diperoleh untuk data *training* dapat dilihat pada Tabel 9.

Tabel 9. Hasil *Confusion Matrix* Data Training

Hasil Prediksi	Prediksi : Berat	Prediksi : Ringan	Prediksi : Tidak ada
Aktual : Berat	41	0	0
Aktual : Ringan	7	42	12
Aktual : Tidak ada	0	24	72



Gambar 2. Hasil Pohon Keputusan Setelah *Post-Pruning*

Berdasarkan *output* yang diperoleh, dapat diketahui bahwa kalkulasi TP, TN, FP dan FN untuk kelas berat adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{11} = 41$
- FN = $(X_{12} + X_{13}) = 0 + 0 = 0$
- FP = $(X_{21} + X_{31}) = 7 + 0 = 7$
- TN = $(X_{22} + X_{23} + X_{32} + X_{33}) = 42 + 12 + 24 + 72 = 150$

Kalkulasi TP, TN, FP dan FN untuk kelas ringan adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{22} = 42$
- FN = $(X_{21} + X_{23}) = 7 + 12 = 19$
- FP = $(X_{12} + X_{32}) = 0 + 24 = 24$
- TN = $(X_{11} + X_{13} + X_{31} + X_{33}) = 41 + 0 + 0 + 72 = 113$

Kalkulasi TP, TN, FP dan FN untuk kelas tidak ada adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{33} = 72$
- FN = $(X_{31} + X_{32}) = 0 + 24 = 24$
- FP = $(X_{13} + X_{23}) = 0 + 12 = 12$
- TN = $(X_{11} + X_{12} + X_{21} + X_{22}) = 41 + 0 + 7 + 42 = 90$

Hasil *confusion matrix* yang diperoleh untuk data *training* dapat dilihat pada Tabel 10 berikut ini.

Tabel 10. Hasil Confusion Matrix Data Testing

Hasil Prediksi	Prediksi : Berat	Prediksi : Ringan	Prediksi : Tidak ada
Aktual : Berat	12	0	0
Aktual : Ringan	1	13	8
Aktual : Tidak ada	0	15	30

Berdasarkan output tersebut, dapat diketahui bahwa kalkulasi TP, TN, FP dan FN untuk kelas berat adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{11} = 12$
- FN $= (X_{12} + X_{13}) = 0 + 0 = 0$
- FP $= (X_{21} + X_{31}) = 1 + 0 = 1$
- TN $= (X_{22} + X_{23} + X_{32} + X_{33}) = 12 + 8 + 15 + 30 = 66$

Kalkulasi TP, TN, FP dan FN untuk kelas ringan adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{22} = 13$
- FN $= (X_{21} + X_{23}) = 1 + 8 = 9$
- FP $= (X_{12} + X_{32}) = 0 + 15 = 15$
- TN $= (X_{11} + X_{13} + X_{31} + X_{33}) = 12 + 0 + 0 + 30 = 42$

Kalkulasi TP, TN, FP dan FN untuk kelas tidak ada adalah:

- TP untuk data aktual dan prediksi terdapat pada $X_{33} = 30$
- FN $= (X_{31} + X_{32}) = 0 + 15 = 15$
- FP $= (X_{13} + X_{23}) = 0 + 8 = 8$
- TN $= (X_{11} + X_{12} + X_{21} + X_{22}) = 12 + 0 + 1 + 13 = 26$

Berikut ditunjukkan perbandingan keduanya untuk data *training* dan *testing*.

Tabel 11. Perbandingan Dari Confusion Matrix Data Training dan Data Testing

	Training	Testing
Accuracy	78%	70%
Sensitivity	78%	73%
Specificity	88%	82%
Positive prediction value	81%	75%
Negative prediction value	89%	83%
Prevalence	33%	33%

3.3. Decision Tree C4.5 Perhitungan Manual

1. Menentukan *node* 1, pertama-tama hitung nilai *entropy* semua variabel. Dibawah ini terdapat contoh perhitungan *entropy* dari dua variabel pada data penyakit asam urat yaitu variabel jenis kelamin dan konsumsi alkohol.

a. Seluruh Data

Tabel 12. Total Seluruh Data

Total	Total	Ringan	Berat	Tidak ada
	198	64	43	91

$$Entropy(S) = Entropy(S) = \sum_{i=1}^a -P_i \times \log_2 P_i$$

$$\left(-\frac{64}{198} \times \log_2 \left(\frac{64}{198} \right) \right) + \left(-\frac{43}{198} \times \log_2 \left(\frac{43}{198} \right) \right) + \left(-\frac{91}{198} \times \log_2 \left(\frac{91}{198} \right) \right) = 0,457738783$$

b. Jenis kelamin

Diketahui bahwa variabel Jenis kelamin termasuk kedalam variabel kategori. Berikutnya contoh perhitungan *entropy* variabel jenis kelamin:

Tabel 13. Total Jenis Kelamin

	Total	Ringan	Berat	Ringan
Perempuan	117	42	17	58
Laki-laki	81	22	26	33

- Perempuan

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^a -P_i \times \log_2 P_i \\
 &= \left(-\frac{42}{117} \times \log_2 \left(\frac{42}{117} \right) \right) + \left(-\frac{17}{117} \times \log_2 \left(\frac{17}{117} \right) \right) + \left(-\frac{58}{117} \times \log_2 \left(\frac{58}{117} \right) \right) \\
 &= 0,432519832
 \end{aligned}$$

- Laki-laki

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^a -P_i \times \log_2 P_i \\
 &= \left(-\frac{22}{81} \times \log_2 \left(\frac{22}{81} \right) \right) + \left(-\frac{26}{81} \times \log_2 \left(\frac{26}{81} \right) \right) + \left(-\frac{33}{81} \times \log_2 \left(\frac{33}{81} \right) \right) \\
 &= 0,471033586
 \end{aligned}$$

c. Konsumsi alkohol

Diketahui bahwa variabel konsumsi alkohol termasuk kedalam variabel kategori. Berikutnya contoh perhitungan *entropy* variabel konsumsi alkohol:

Tabel 14. Total Konsumsi alkohol

	Total	Ringan	Berat	Tidak ada
Ya	37	9	11	17
Tidak	161	55	32	74

- Ya

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^a -P_i \times \log_2 P_i \\
 &= \left(-\frac{9}{37} \times \log_2 \left(\frac{9}{37} \right) \right) + \left(-\frac{11}{37} \times \log_2 \left(\frac{11}{37} \right) \right) + \left(-\frac{17}{37} \times \log_2 \left(\frac{17}{37} \right) \right) \\
 &= 0,461144054
 \end{aligned}$$

- Tidak

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^a -P_i \times \log_2 P_i \\
 &= \left(-\frac{55}{161} \times \log_2 \left(\frac{55}{161} \right) \right) + \left(-\frac{32}{161} \times \log_2 \left(\frac{32}{161} \right) \right) + \left(-\frac{74}{161} \times \log_2 \left(\frac{74}{161} \right) \right) \\
 &= 0,453981811
 \end{aligned}$$

2. Hitung *Gain*

a. Jenis kelamin

Berikut perhitungan *gain* variabel jenis kelamin:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

$$= 0,455846176 - \left(\left(\frac{117}{198} \times 0,432519832 \right) + \left(\frac{81}{198} \times 0,471033586 \right) \right)$$

$$= 0,009463324$$

b. Konsumsi alkohol

Berikut perhitungan *gain* variabel konsumsi alkohol:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i)$$

$$= 0,455846176 - \left(\left(\frac{37}{198} \times 0,461144054 \right) + \left(\frac{161}{198} \times 0,453981811 \right) \right)$$

$$= 0,002418573$$

Berikut merupakan hasil perhitungan untuk *Node* 1.

Tabel 15. Hasil Perhitungan *Node* 1

	Total	Ringan	Berat	Tidak ada	Entropy	Gain
	198	64	43	91	0,457738783	
	Jenis kelamin					
Perempuan	117	42	17	58	0,432519832	0,009463324
Laki-laki	81	22	26	33	0,471033586	
	Konsumsi alkohol					
Ya	37	9	11	17	0,461144054	0,002418573
Tidak	161	55	32	74	0,453981811	
	Konsumsi gula berlebihan					
Normal	144	38	30	76	0,441089941	0,014507523
Kurang	41	21	8	12	0,443478978	
Lebih	13	5	5	3	0,466169213	
	Usia					
Lansia	37	13	21	3	0,387681926	0,062378581
Remaja	111	35	7	69	0,362091852	
Dewasa	50	16	15	19	0,474897864	
	Nilai IMT					
Normal	107	41	18	48	0,446032164	0,016027319
Gemuk	58	16	21	21	0,473787819	
Kurus	33	7	4	22	0,371325297	
	Riwayat penyakit					
0	72	28	0	45	0	0,457738783
1	54	11	43	0	0	
2	72	25	0	46	0	

Setelah memperoleh nilai *gain* dari seluruh partisi, didapatkan bahwa variabel riwayat penyakit memiliki nilai *gain* terbesar dari seluruh variabel yang ada yaitu 0,457738783. Sehingga variabel ini akan menjadi *Node* 1. Jumlah kasus variabel riwayat penyakit adalah 64 pasien dengan kadar asam urat ringan, 43 pasien kadar asam urat berat dan 91 pasien yang tidak memiliki asam urat.

Perhitungan untuk *node* berikutnya sama dengan perhitungan untuk menentukan *node* 1 namun terdapat perbedaan pada data. Untuk perhitungan selanjutnya yaitu mem-*filter* data berdasarkan nilai variabel pada *Node* 1 (Riwayat penyakit). Proses partisipasi pohon keputusan akan berhenti saat cabang dalam *node* N mendapat kelas yang sama dan dapat dikatakan memiliki kelas yang sama yaitu dengan melihat apakah dalam variabel tersebut ada atau tidak ada pasien penderita penyakit asam urat.

4. KESIMPULAN

1. Hasil data *screening* pasien asam urat pada pemeriksaan yang dilakukan Civitas Akademika UKIM menghasilkan 277 orang pasien dengan mengambil beberapa variabel penting yang merupakan faktor-faktor penyebab terkena penyakit asam urat yaitu jenis kelamin, konsumsi alkohol, konsumsi gula berlebihan, usia, nilai IMT, riwayat penyakit serta tingkat kadar asam urat. Diperoleh bahwa sebagian

besar Civitas Akademika UKIM yang terkena asam urat berada pada usia remaja. Jenis kelamin perempuan lebih banyak terkena asam urat dengan total 85 pasien sedangkan laki-laki total 70 pasien. Namun, pada pemeriksaan konsumsi alkohol, hampir semua pasien tidak mengonsumsi alkohol. Untuk konsumsi gula berlebihan dan nilai IMT, sebagian besar masuk dalam kelas konsumsi “normal” dan riwayat penyakit menjadi faktor penyebab utama pasien terkena asam urat. Dengan demikian, dalam hasil penelitian ini diperoleh variabel yang berpengaruh dan tidak berpengaruh pasien dapat terkena penyakit asam urat. Seperti, jenis kelamin dan konsumsi alkohol adalah variabel yang tidak berpengaruh pada kasus ini.

2. Metode *decision tree* C4.5 yang digunakan dalam klasifikasi penyakit asam urat pada Civitas Akademika UKIM menghasilkan 8 aturan/*rule*. 8 aturan tersebut dapat dijadikan pola dalam menentukan pasien penderita asam urat ringan, berat atau tidak ada. Metode ini dapat digunakan untuk mengetahui hasil klasifikasi penyakit asam urat pada Civitas Akademika UKIM dengan tingkat keakuratan sebesar 78% setelah *post-pruning* dan berdasarkan evaluasi yang dilakukan dapat diketahui bahwa proses pembentukan pohon menggunakan teknik *pruning* tidak selalu memiliki akurasi yang lebih besar dan ditemukan bahwa penyebab terbesar pasien menderita penyakit asam urat adalah faktor riwayat penyakit.

DAFTAR PUSTAKA

- [1] I. Hasanah, "Implementasi Decision Tree C4.5 Untuk Klasifikasi Cara Keluar Pasien Ginjal Kronis Berdasarkan Rekam Medis BPJS Kesehatan di RSUDAM," *Universitas Islam Indonesia*, pp. 1-68, 2021.
- [2] Ermanto and D. Mahardi, "Penerapan Data Mining Untuk Klasifikasi Pandemi Covid-19 Di Indonesia Dengan Algoritma Naive Bayes," *Jurnal Teknologi Pelita Bangsa*, vol. 11, no. 1, pp. 2407-3903, 2020.
- [3] C. Sammut, "Encyclopedia Of Machine Learning," Australia, Springer, 2011.
- [4] A. R. Sukma, R. Halfis and A. Hermawan, "Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4.5," *Jurnal Teknik Komputer AMIK BSI*, vol. 5, no. 1, pp. 21-28, 2019.
- [5] J. Han and M. Kamber, "Data Mining. In: Concepts and Techniques," San Francisco, Morgan Kaufmann Publishers, 2006, pp. 1-772.
- [6] E. Widhiastuti, "Implementasi Data Mining Untuk Memprediksi Penyakit Hipertensi Dalam Kehamilan Menggunakan Algoritma C4.5 (Study Kasus: Puskesmas Rimba Melintang, Rokan Hilir)," *Yayasan Lembaga Pendidikan Islam Daerah Riau, Universitas Islam Riau*, 2021.
- [7] Noviyanti, "Hidup Sehat Tanpa Asam Urat," in *NOTEBOOK*, Yogyakarta, 2015.
- [8] K. N, "Asam Urat," Yogyakarta, B First, 2009.
- [9] I. Noya, I. V. Lawalata and B. Talarima, "Analisis Hubungan Peningkatan Kadar Asam Urat di Wilayah Kerja Puskesmas Waihoka Kota Ambon Tahun 2018," *Moluccas Health Journal*, vol. 1, no. 2, pp. 2686-1828, 2019.
- [10] "KEMENKES," Direktorat Jenderal Pelayanan Kesehatan, 2023. [Online]. Available: <https://yankes.kemkess.go.id>.
- [11] WHO, "Monitoring Health Organization," [Online].
- [12] RISKESDAS, Laporan Provinsi Maluku RISKESDAS 2018, Ambon: Badan Penelitian dan Pengembangan Kesehatan (LPB), 2018.
- [13] M. Nasir, "Gambaran Asam Urat Pada Lansia Di Wilayah Kampung Selayar Kota Makassar," *Jurnal Media Analisis Kesehatan*, vol. 8, no. 2, pp. 1-5, 2017.
- [14] UKIM, "Laporan Praktek Screening Penyakit Tidak Menular," 2011.

- [15] V. Fernitha, "Penerapan Algoritma C4.5 Untuk Deteksi Penyakit Kanker Serviks," Yogyakarta, Universitas Sanata Dharma, 2019, pp. 1-105.
- [16] D. Arianti, "Sistem Pendiagnosa Penyakit Hiperkolesterol Menggunakan Algoritma C4.5," Skripsi thesis, Universitas Islam Negeri Sultan Syarif Kasim Riau, 2010.
- [17] I. Hasanah, "Implementasi Decision Tree C4.5 Untuk Klasifikasi Cara Keluar Pasien Ginjal Kronis Berdasarkan Rekam Medis Bpjs Kesehatan Di RSUAM," Universitas Islam Indonesia, 2021.
- [18] H. Zuhriyah and D. J. E. Sari, "Pengaruh Pemberian Air Rebusan Daun Salam Terhadap Kadar Asam Urat Pada Lansia Di Puskesmas Pembantu Desa Manonggal Kecamatan Klampis Kabupaten Bangkalan," *Jurnal Ilmiah Multi Disiplin Indonesia*, vol. 2, no. 1, pp. 2809-1620, 2022.

