

Klasifikasi Diagnosis Penyakit Diabetes Menggunakan Algoritma C4.5

Indah Sagita Cahyani¹, Karlena Indriani², Monikka Nur Winnarto^{3*}

^{1,2,3}Universitas Bina Sarana Informatika
e-mail: ¹19236025@bsi.ac.id, ²karlena@bsi.ac.id, ³monikka.mnt@bsi.ac.id

Diterima	Direvisi	Disetujui
02-06-2025	04-06-2025	12-06-2025

Abstrak - Kesehatan memiliki peranan yang sangat penting untuk menunjang kehidupan manusia, dengan memiliki kesehatan yang baik, manusia dapat melakukan aktifitas dengan produktif dalam sosialisasi atau ekonomi untuk mencapai tujuan hidup. Salah satu penyakit yang dapat mengakibatkan komplikasi bahkan kematian adalah penyakit diabetes. Angka kematian yang tinggi yang diakibatkan oleh penyakit diabetes sangat mengkhawatirkan, diagnosis dini begitu penting dilakukan untuk menekan angka kematian. Selain itu diagnosis dini juga merupakan titik awal penderita untuk mencegah terjadinya diabetes lebih parah dengan melakukan pola hidup sehat agar tidak mengalami komplikasi. Diagnosa penyakit diabetes pada manusia salah satunya dapat dilakukan dengan menerapkan algoritma Machine Learning yaitu C4.5. Penelitian ini bertujuan melakukan diagnosa penyakit diabetes untuk menghindari keterlambatan diagnosis yang dapat mengakibatkan komplikasi pada pasien pengidap penyakit diabetes dengan menggunakan Algoritma C4.5 dalam mendiagnosa dan membuat model prediksi yang menghasilkan sebuah pohon keputusan serta pengujian terhadap hasil diagnosa penyakit diabetes. Dalam penelitian ini terdapat beberapa atribut klasifikasi yaitu gender, age, urea, creatinine, HbA1c, kolesterol, trigeliserida, HDL, LDL, VLDL, dan BMI. Hasil dari penelitian ini dijadikan sebagai acuan untuk dapat melihat apakah seseorang terkena diabetes atau tidak. Hasil akhir dari penelitian ini yaitu menghasilkan nilai akurasi sebesar 99.47%, dimana nilai ini menandakan bahwa algoritma C4.5 mampu melakukan diagnosis penyakit diabetes dengan baik.

Kata Kunci : Diabetes, Data Mining, Algoritma C4.5

Abstracts

Health plays a very important role in supporting human life, by having good health, humans can carry out productive activities in socialization or economy to achieve life goals. One of the diseases that can cause complications and even death is diabetes. The high mortality rate caused by diabetes is very worrying, early diagnosis is very important to reduce the mortality rate. In addition, early diagnosis is also the starting point for sufferers to prevent diabetes from getting worse by adopting a healthy lifestyle so as not to experience complications. One of the ways to diagnose diabetes in humans is by applying the Machine Learning algorithm, namely C4.5. This study aims to diagnose diabetes to avoid delays in diagnosis that can cause complications in patients with diabetes by using the C4.5 Algorithm in diagnosing and creating a prediction model that produces a decision tree and testing the results of the diabetes diagnosis. In this study there are several classification attributes, namely gender, age, urea, creatinine, HbA1c, cholesterol, triglycerides, HDL, LDL, VLDL, and BMI. The results of this study are used as a reference to see whether someone has diabetes or not. The final result of this study is to produce an accuracy value of 99.47%, where this value indicates that the C4.5 algorithm is able to diagnose diabetes well.

Keywords: Diabetes, Data Mining, C4.5 Algorithm

PENDAHULUAN

Kesehatan merupakan peranannya sangat penting untuk menunjang kehidupan manusia, dengan memiliki kesehatan yang baik, manusia dapat melakukan aktifitas dengan produktif dalam sosialisasi atau ekonomi untuk mencapai tujuan hidup. Salah satu penyakit yang dapat mengakibatkan komplikasi bahkan kematian adalah penyakit diabetes. Diabetes bukan hanya penyebab dari

kematian prematur di dunia, penyakit ini bisa menyebabkan kebutaan, gagal ginjal, dan bisa juga menyebabkan penyakit jantung (Robbani et al., 2022).

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi (*taksonomi*) merupakan proses penempatan objek atau konsep tertentu ke dalam satu set kategori berdasarkan objek yang digunakan. Salah satu teknik klasifikasi yang populer

digunakan adalah *decision tree*. Klasifikasi juga dapat diartikan sebagai sebuah proses menemukan suatu model atau fungsi yang menggambarkan dan membedakan klas data atau konsep dengan tujuan dapat menggunakan model untuk membuat prediksi kelas objek dimana kelas labelnya tidak diketahui. Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya (Bachtiar, 2023).

Menurut Han dan Kamber, data mining adalah proses penemuan pola atau informasi yang berguna dari basis data besar dengan menggunakan metode termasuk teknik statistik, matematika, dan kecerdasan buatan. Proses data mining melibatkan penggunaan berbagai teknik statistik, matematis, dan kecerdasan buatan untuk menganalisis data dengan cara yang sistematis dan otomatis. Hasil dari data mining dapat digunakan untuk mendukung pengambilan keputusan, mengidentifikasi tren pasar, meningkatkan efisiensi operasional atau merumuskan strategi bisnis (Sudipa et al., 2024).

“Data Mining adalah suatu proses pencarian data secara otomatis dapat mendapatkan sebuah model dari database yang besar” (Dewi & Rahayu, 2022). Penelitian sebelumnya yang menggunakan metode algoritma C4.5 memiliki akurasi yang sangat baik, maka dalam penelitian ini akan menerapkan algoritma C4.5 pada kasus diabetes, diharapkan agar dapat meningkatkan kinerja algoritma C4.5 dan menghasilkan informasi yang akurat.

“Algoritma C4.5 adalah algoritma yang digunakan untuk mengklasifikasikan dataset dengan membuat pohon keputusan yang dapat dimanfaatkan untuk memprediksi suatu keputusan sesuai dengan aturan pengambilan keputusan. Dalam pohon keputusan merupakan diagram alir dan setiap internal node menghimpun atribut akan diuji, pada setiap cabang mempresentasikan kelas-kelas tertentu” (Putra, 2021). “Algoritma C4.5 adalah algoritma klasifikasi dan prediksi yang cukup terkenal dan biasanya memiliki akurasi yang tinggi” (Najib et al., 2019).

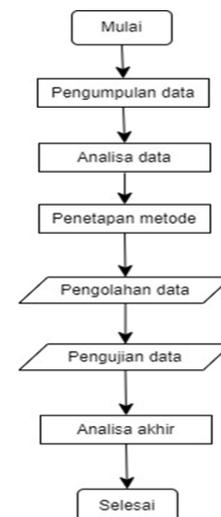
Menurut International Diabetes Federation yang mengidap penyakit diabetes pada tahun 2015 sebanyak 415 juta jiwa, dan diperkirakan meningkat sebanyak 227 juta jiwa atau menjadi 642 juta jiwa pada tahun 2040 yang mengidap penyakit tersebut. Di setiap negara jumlah Diabetes Melitus mengalami peningkatan dan kasus terbanyak orang yang

mengalami Diabetes Melitus berada di usia antara 40-59 tahun. Dilihat dari angka kematian yang tinggi yang diakibatkan oleh diabetes, diagnosis dini begitu penting dilakukan untuk menekan angka kematian. Selain itu diagnosis dini juga merupakan titik awal penderita untuk mencegah terjadinya diabetes lebih parah dengan melakukan pola hidup sehat agar tidak mengalami komplikasi. Sedangkan seseorang yang sudah lama mengalami diabetes yang tidak terdiagnosis dan tidak diobati, berkemungkinan besar kesehatan tubuhnya akan semakin buruk. (Fadhillah et al., 2022).

Berdasarkan data diatas pada penelitian ini klasifikasi diagnosis diabetes dilakukan dengan menggunakan teknik data mining yaitu algoritma C4.5. Penelitian ini mengarah pada diagnosis apakah seseorang menderita diabetes dengan mengklasifikasikan variabel-variabel yang telah ditentukan. Dua belas variabel diuji, salah satunya adalah variabel target, yaitu “positif” dan “negatif”.

METODE PENELITIAN

Proses penelitian diawali dengan pengumpulan data melalui Mendeley setelah itu dilakukan analisis data untuk pemahaman terhadap permasalahan yang ada, dalam menyelesaikan permasalahan peneliti melakukan tinjauan pustaka untuk mencari solusi tepat terhadap permasalahan. Kemudian dilakukan penetapan metode, pada penelitian ini menggunakan metode algoritma C4.5 selanjutnya dilakukan pengolahan data dan pengujian melalui aplikasi RapidMiner. Pada tahap terakhir peneliti melakukan analisa akhir.



Sumber: Penelitian, 2025

Gambar 1. Tahapan Penelitian

HASIL DAN PEMBAHASAN

Pengumpulan Data

Penelitian ini berlandaskan pada metode Algoritma C4.5 untuk meningkatkan akurasi diagnosis penyakit diabetes. Hasil dari Algoritma C4.5 adalah sebuah pohon keputusan atau *decision tree*. Pohon keputusan berperan untuk mengkaji data, menemukan ikatan tersembunyi antara sejumlah variabel masukan dan variabel sasaran. Proses dalam pohon keputusan adalah mengubah data yang berisi aturan-aturan menjadi struktur hierarki yang kemudian menghasilkan keputusan akhir yaitu berupa variabel sasaran atau label. Sehingga proses pengambilan keputusan dapat dilakukan dengan segera karena pohon keputusan memangkas proses yang kompleks menjadi lebih ringkas.

Langkah pertama yang dilakukan adalah mengakumulasi data, dalam penelitian ini menggunakan data dari mendeley <https://data.mendeley.com/datasets/wj9rwkp9c2/1> yaitu data dikumpulkan dari masyarakat Irak karena datanya diperoleh dari laboratorium Rumah Sakit *Medical City* dan (Pusat Spesialis Endokrinologi dan Rumah Sakit Pendidikan Diabetes-Al-Kindy). File pasien diambil dan data diekstraksi darinya dan dimasukkan ke dalam database untuk membuat kumpulan data diabetes. Sebanyak 947 data termuat di dalamnya, data tersebut diseleksi kemudian digolongkan menjadi 12 atribut, 11 diantaranya adalah atribut perbandingan dan satu atribut hasil untuk melihat status dari pasien tersebut. Data-data pasien dapat dilihat dalam gambar berikut:

ID	No	Pat	Gender	AGE	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	VLDL	BMI	CLASS
502	17975	F	50	4,7	46	4,9	4,2	0,9	2,4	1,4	0,5	24	Negative	
735	34221	M	26	4,5	62	4,9	3,7	1,4	1,1	2,1	0,6	23	Negative	
420	47975	F	50	4,7	46	4,9	4,2	0,9	2,4	1,4	0,5	24	Negative	
504	34223	M	33	7,1	46	4,9	4,9	1	0,8	2	0,4	21	Negative	
634	34224	F	45	2,3	24	4	2,9	1	1	1,5	0,4	21	Negative	
721	34225	F	50	2	50	4	3,6	1,3	0,9	2,1	0,6	24	Negative	
421	34227	M	48	4,7	47	4	2,9	0,8	0,9	1,6	0,4	24	Negative	
670	34229	M	43	2,6	67	4	3,8	0,9	2,4	3,7	1	21	Negative	
759	34230	F	32	3,6	28	4	3,8	2	2,4	3,8	1	24	Negative	
636	34231	F	31	4,4	55	4,2	3,6	0,7	1,7	1,6	0,3	23	Negative	
788	34232	F	33	3,3	53	4	4	1,1	0,9	2,7	1	21	Negative	
82	40815	F	30	3	42	4,1	4,9	1,3	1,2	3,2	0,5	22	Negative	
132	34234	F	45	4,6	54	5,1	4,2	1,7	1,2	2,2	0,8	23	Negative	
402	34235	F	50	3,5	39	4	4	1,5	1,2	2,2	0,7	24	Negative	
566	34236	M	50	5,5	74	5	3,6	1,1	1	2,1	0,5	21	Negative	
596	34237	F	50	5,9	53	5,4	5,3	0,8	1,1	4,1	0,3	21	Negative	
39	23984	M	45	5,3	77	11,2	3,9	1,5	1,3	2	10,4	29,5	Positive	
474	67036	M	31	3,4	55	6,5	4,9	1,6	1	3,2	0,7	24	Positive	
648	745226	M	45	5,3	77	11,2	3,9	1,5	1,3	2	10,4	29,5	Positive	
48	23987	M	30	3	40	6,9	4,5	1,8	1,2	2,6	12,7	24,6	Positive	
57	23990	M	35	4,8	64	7,7	3,7	1	1,2	2	7,2	27,3	Positive	
89	23994	M	45	4,8	82	7,2	4,7	1,8	0,8	3,1	12,7	31,2	Positive	
74	34268	M	45	3,6	80	5	6,1	3,7	0,7	3,9	1,7	22	Positive	
87	24000	F	50	4	56	13,7	4,4	2	1	2,5	0,9	29	Positive	
505	23044	F	48	4	38	6,5	4,4	2,3	1,3	2,2	1	23	Positive	
654	23447	M	38	5,8	59	6,7	5,3	2	1,6	2,9	1,4	40,3	Positive	
7	34278	F	46	3	59	5,1	5,7	3,8	1,3	2,8	1,7	24	Positive	
61	34279	M	45	2,3	42	5,1	4,4	1,5	1	2,8	0,5	25	Positive	
102	24005	F	34	3	39	7,2	9,5	1,7	1,3	2,5	0,6	22	Positive	
228	34280	M	43	5,4	62	4,1	5,9	2	1,1	3,9	0,9	21	Positive	
674	34283	M	49	3,9	65	5,2	5,1	1,7	3,9	0,6	1,6	24	Positive	
744	34284	F	49	3,8	56	5,1	6	3,5	1,1	3,5	3,6	25	Positive	
10	34289	F	45	3,1	54	4	5,9	1,8	1,6	3,5	0,8	24	Positive	

Sumber: Penelitian, 2025

Gambar 2. Data Hasil Pemeriksaan Diabetes

Tahap selanjutnya dataset yang diakumulasi akan diklasifikasikan berdasarkan atribut yang telah ditentukan untuk dilakukan pengkajian Algoritma C4.5 dengan menggunakan Rapidminer. Hasil perhitungan dari model yang diusulkan menciptakan pohon keputusan. Berikut atribut serta *range* dan nilai yang dapat dilihat pada tabel 1.

Tabel 1. Atribut dan Nilai Kategori untuk Menentukan Penyakit Diabetes

Jenis Pemeriksaan	Nilai	Range	Sumber
Gender/Jenis Kelamin	F	Perempuan	(Rosita et al., 2022)
	M	Laki-laki	
Age/umur	20-54 Tahun	Risiko rendah	(Milita et al., 2018)
	>55 Tahun	Beresiko	
Creatinine Ratio	61.9 to 114.9 mmol/L	Normal	(Benioff Children's Hospitals, 2021)
	>115 mmol/L	Tidak Normal	
Body Mass Index atau BMI	< 18,5 – 29,9 kg/m ²	Normal	(Kementerian Kesehatan RI, 2019)
	>30 kg/m ²	Tidak Normal	
Chol (Cholesterol)	< 5,17-6,18 mmol/L	Normal	(Rosenson, 2021)
	>6,21 mmol/L	Tidak Normal	
Urea	1,8 - 7,1 mmol/L	Normal	(Liu et al., 2021)
	>7.1 mmol	Tidak Normal	
HbA1c	< 48 mmol/mol	Normal	(Hospitals, 2024)
Trigliserida (TG)	< 1,7 - 5.6mmol/L	Normal	(Kementerian Kesehatan RI, 2018)
	> 5.6 mmol/L atau lebih tinggi	Tidak Normal	
Lipoprotein Densitas Rendah (LDL)	<3,4 mmol/L	Normal	(Cleveland Clinic Abu Dhabi, 2021)
	>3,4 mmol/L	Tidak Normal	
Lipoprotein Densitas Tinggi (HDL)	>1,3 mmol/L	Normal	(Lee & Siddiqui, 2024)
	<1,3 mmol/L	Tidak Normal	
Very Low Density	<0,77 mmol/L	Normal	(Barrell, 2021)

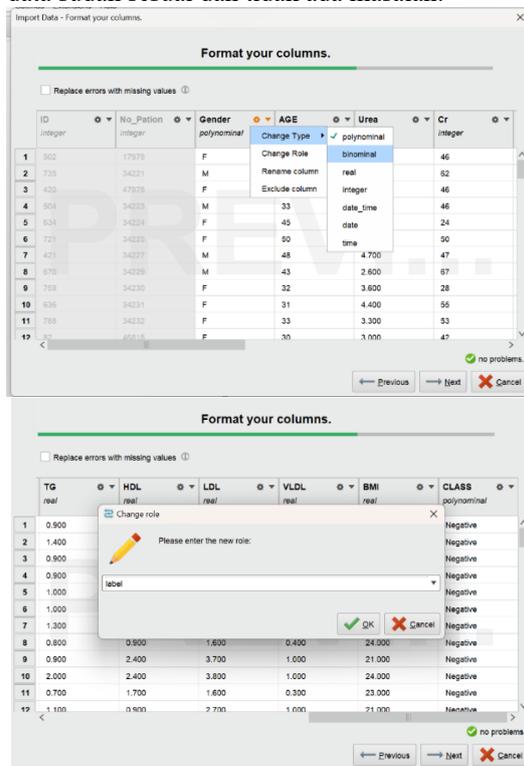
Jenis Pemeriksaan	Nilai	Range	Sumber
Lipoprotein (VLDL)	>0,77 mmol/L).	Tidak Normal	

Sumber: Penelitian, 2025

Analisa Data

Tahap Pre-Processing Data

Tahap proses pengolahan data di aplikasi RapidMiner pada bagian atribut gender untuk type nya diganti ke binominal karna memiliki dua nilai dan bagian class ditambah menjadi label. Pengolahan data akan secara otomatis melakukan pre-processing data yaitu seperti mendeteksi data yang kosong, isi kolom yang berbeda dan akan otomatis memberi centang no problems bila data sudah sesuai dan tidak ada masalah.



Sumber : Penelitian, 2025

Gambar 3. Proses Pengolahan Data Pada RapidMiner

Penetapan Metode

Algoritma C4.5 digunakan dalam penerapan data mining ini dengan tujuan untuk memprediksi pasien penyakit diabetes. Pohon keputusan akan membentuk akar pohon, selanjutnya data dibedakan sesuai dengan atribut yang sama untuk dibentuk daun atau cabang. Proses pemangkasan cabang pohon yang tidak diperlukan oleh pohon yang sudah terbentuk atau melakukan penyederhanaan ukuran pohon karena pohon keputusan yang dibentuk biasanya berukuran besar. Pemangkasan juga berfungsi untuk meminimalisir jumlah *error* pada hasil prediksi.

Berikut adalah cara menghitung nilai *entropy* dan nilai *gain* secara manual dengan menggunakan persamaan berikut :

$$Entropy (total) = ((-844/947)*LOG_2(844/947))+(-103/947)*LOG_2(103/947))$$

$$Entropy (total) = 0,496178092$$

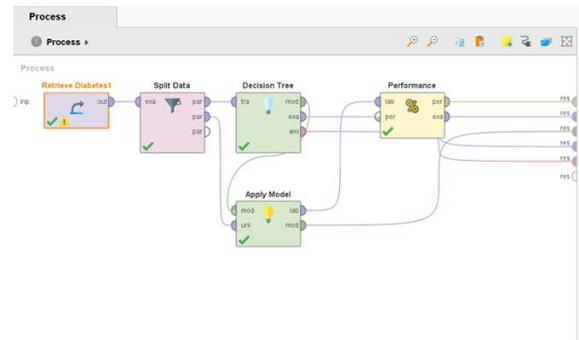
Selanjutnya pada nilai *Gain* pada baris *Gender* dihitung melalui persamaan sebagai berikut:

$$Gain (total, gender) = (0,496178092) - ((529/947) * 0,379669703) - ((418/947) * 0,617568786)$$

$$Gain (total, gender) = 0,011501191$$

Pengolahan Data

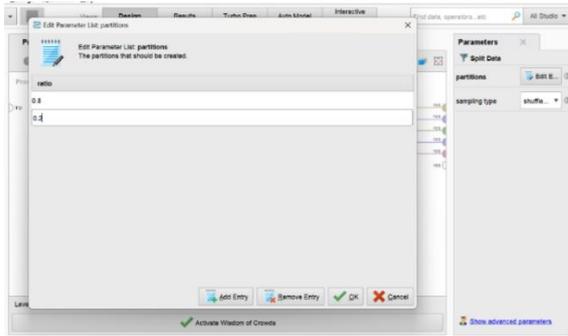
Pengolahan data dilakukan menggunakan aplikasi Rapidminer, pada tahap dataset memakai operator split data yang akan dibagi menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 80:20, 70:30 dan 60:40 serta memakai Shuffled Sampling. Data *training* akan melatih data untuk membangun sebuah model algoritma C4.5 sedangkan data *testing* untuk mengetahui *accuracy*, *precision*, *recall* dan kurva ROC.



Sumber : Penelitian, 2025

Gambar 4. Menggunakan Tools Split Data dan Penggunaan Operator RapidMiner

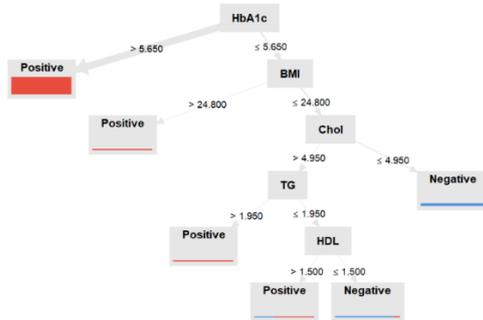
Pada klasifikasi data, RapidMiner akan membangun sebuah model *decision tree* dengan data *training* yaitu dengan membuat pohon keputusan dan *testing* akan digunakan untuk pengujian dan evaluasi data melalui operator *performance*. Selanjutnya ratio untuk subset *training* dan subset *testing* menggunakan perbandingan 0.8 dan 0.2 untuk mengetahui keakuratan data seperti dibawah ini Bentuk gambar mengikuti contoh berikut:



Sumber : Penelitian, 2025
Gambar 5. Subset *Training* dan Subset *Testing*

Pengujian Data

Hasil dari uji coba yang dilakukan dapat menghasilkan *Decision Tree* (pohon keputusan) dan untuk menghasilkan nilai *accuracy*. Berikut untuk melihat *Decision Tree* (Pohon Keputusan)



Sumber : Penelitian, 2025
Gambar 6. Decision Tree

Decision tree (pohon keputusan) adalah cara untuk membangun model dalam penambangan data, itu dapat dipahami sebagai pohon biner terbalik. Dari perhitungan diatas terdapat beberapa rules yang dapat dijadikan sebagai referensi dalam klasifikasi penyakit diabetes. Adapun antara rules diatas berupa empat rules keputusan positif dan dua rules keputusan negatif dapat dijabarkan melalui narasi sebagai berikut:

1. Jika HbA1c = >5.650 maka hasil positif terkena diabetes.
2. Jika HbA1c = < 5.650 cek BMI = > 24.800 maka hasil positif terkena diabetes.
3. Jika HbA1c = < 5.650, BMI = < 24.800, Cholesterol = < 4.950 maka hasil negatif diabetes
4. Jika HbA1c = < 5.650, BMI = < 24.550, Cholesterol = > 4.950, TG = > 1.950 maka hasil positif terkena diabetes.
5. Jika HbA1c = < 5.650, BMI = < 24.800, Cholesterol = > 4.950, TG = < 1.950, HDL = > 1.500 maka hasil positif terkena diabetes

6. Jika HbA1c = < 5.650, BMI = < 24.550, Cholesterol = > 4.950, TG = < 1.950, HDL = > 1.500 maka hasil negatif diabetes

Analisa Akhir

Pengujian pada hasil perhitungan dengan menggunakan aplikasi *RapidMiner* dilakukan dengan beberapa proses sehingga memperoleh hasil akurasi 99.47%. Untuk melihat nilai *accuracy*, *precision* dan *recall* dapat dilihat pada gambar berikut ini:

accuracy: 99.47%

	true Negative	true Positive	class precision
pred. Negative	23	1	95.83%
pred. Positive	0	165	100.00%
class recall	100.00%	99.40%	

precision: 100.00% (positive class: Positive)

	true Negative	true Positive	class precision
pred. Negative	23	1	95.83%
pred. Positive	0	165	100.00%
class recall	100.00%	99.40%	

recall: 99.40% (positive class: Positive)

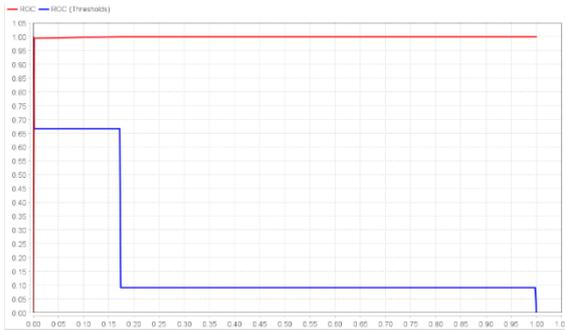
	true Negative	true Positive	class precision
pred. Negative	23	1	95.83%
pred. Positive	0	165	100.00%
class recall	100.00%	99.40%	

Sumber : Penelitian, 2025
Gambar 7. Hasil Nilai *Accuracy*, *Recall* dan *Precision*

Dari gambar diatas telah didapatkan hasil confusion matrix dari pengujian data melalui algoritma C4.5. hasil yang didapatkan adalah *accuracy* 99.47%, *precision* 100.00% dan *recall* 99.40%.

Performance Vector		
accuracy: 99.47%		
ConfusionMatrix:		
True:	Negative	Positive
Negative:	23	1
Positive:	0	165
Precision: 100.00% (positive class: positive)		
ConfusionMatrix:		
True:	Negative	Positive
Negative:	23	1
Positive:	0	165
Recall: 99.40% (positive class: positive)		
ConfusionMatrix:		
True:	Negative	Positive
Negative:	23	1
Positive:	0	165
AUC (optimistic): 1.000 (positive class: Positive)		
AUC: 0.999 (positive class: Positive)		
AUC (pessimistic): 0.999 (positive class: Positive)		

Sumber: Penelitian, 2025
Gambar 8. ConfusionMatrix Algoritma C4.5



Sumber : Penelitian, 2025

Gambar 9. Kurva ROC

Dari gambar diatas didapatkan pada *Area Under Curve* atau *AUC* menghasilkan akurasi yang maksimal, untuk hasil *AUC* yaitu sebesar 0.999% ini menandakan bahwa akurasi tersebut termasuk dalam kategori *Excellent Classification* karena pada perbandingan *true positive rate* dan *false positive rate* didalamnya, *true positive rate* mendapatkan hasil yang lebih besar daripada *false positive rate*.

KESIMPULAN

Klasifikasi data mining dengan menggunakan Algoritma C4.5 dapat membantu peneliti dalam prediksi penyakit diabetes berdasarkan variabel-variabel yang diperlukan yaitu *gender*, *age*, *urea*, *creatinine*, *HbA1c*, *cholesterol*, *trigeliserida*, *HDL*, *LDL*, *VLDL*, dan *BMI*. Pada penelitian ini menghasilkan jumlah prediksi positif sebanyak 165 data dan jumlah prediksi negatif sebanyak 23 data, karena adanya software RapidMiner membantu untuk mempermudah terjadinya keterlambatan diagnosa. Berdasarkan pengolahan data menggunakan software rapidminer menghasilkan nilai akurasi yang ideal, pada penelitian sebelumnya mendapatkan akurasi 74.78% dan 90.00% sedangkan pada penelitian ini didapatkan hasil akurasi sebesar 99.47%, presisi sebesar 100.00%, dan recall sebesar 99.40% dengan pembagian data 80:20. Hasil ini membuktikan bahwa metode tersebut dapat digunakan dengan baik dan artinya bahwa rule yang dihasilkan tingkat kebenaran mendekati 100%.

REFERENSI

- Barrell, A. (2021). *What to know about VLDL cholesterol*. 1–9.
- Dewi, S. P., & Rahayu, E. (2022). *Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor*. 3(4), 639–648. <https://doi.org/10.47065/bits.v3i4.1408>
- Fadhillah, R. P., Rahma, R., Sefharni, A., Mufidah, R., Sari, B. N., & Pangestu, A. (2022). *Klasifikasi penyakit diabetes mellitus berdasarkan faktor-faktor penyebab diabetes menggunakan algoritma c4.5*. 07, 1265–1270.
- Hospitals, S. T. (2024). *Tes HbA1c*. 1–5.
- Lee, Y., & Siddiqui, W. J. (2024). *Cholesterol Levels*. 1–7.
- Liu, Q., Wang, Y., Chen, Z., Guo, X., & Lv, Y. (2021). *Age - and sex - specific reference intervals for blood urea nitrogen in Chinese general population*. *Scientific Reports*, 1–7. <https://doi.org/10.1038/s41598-021-89565-x>
- Milita, F., Handayani, S., & Setiaji, B. (2018). *Kejadian Diabetes Mellitus Tipe II pada Lanjut Usia di Indonesia (Analisis Riskesdas 2018)*.
- Najib, A., Nurcahyono, D., & Setiawan, R. P. P. (2019). *KLASIFIKASI DIAGNOSA PENYAKIT DIABETES MELLITUS (DM)*. *Dm*.
- Robbani, A. A., Siregar, A. M., & Kusumaningrum, D. S. (2022). *Klasifikasi Penderita Penyakit Diabetes Menggunakan*. III, 76–82.
- Rosenson, R. S. (2021). *Edukasi pasien : Kolesterol tinggi dan lipid (Melampaui Dasar)*. 1–10.
- Rosita, R., Kusumaningtiar, D. A., Irfandi, A., Ayu, I. M., Studi, P., Masyarakat, K., Kesehatan, F. I., Esa, U., & Barat, K. J. (2022). *AKTIVITAS FISIK LANSIA DENGAN DIABETES MELLITUS TIPE 2 DI PUSKESMAS BALARAJA KABUPATEN TANGERANG*. 10, 364–371.
- Sudipa, I. G. I., Darmawiguna, I. G. M., Dendi, I. M., & Sanjaya, M. (2024). *Buku ajar data mining* (Issue January).
- Suyanto. (2018). *Machine Learning*. Informatika Bandung.
- Suyanto, D. (2019). *Data Mining untuk Klasifikasi dan Klasterisasi Data* (Revisi). Informatika Bandung.