

Implementasi Pohon Keputusan untuk Prediksi Penyakit Diabetes

Hidayatul Maghfirah¹, Olief Ilmandira Ratu Farisi^{1*}

¹Program Studi Pendidikan Matematika, Universitas Nurul Jadid, Probolinggo, Indonesia

*e-mail: farisi@unuja.ac.id

Diterima: 15 November 2025

Revisi: 29 November 2023

Diterbitkan: 1 Desember 2025

ABSTRAK

Diabetes merupakan salah satu penyakit metabolism kronis yang prevalensinya terus meningkat dan berpotensi menimbulkan berbagai komplikasi serius apabila tidak terdeteksi sejak dini. Pemeriksaan kesehatan secara rutin menjadi upaya penting dalam deteksi dini, namun sering kali terkendala oleh keterbatasan biaya dan waktu. Oleh karena itu, penelitian ini bertujuan untuk memprediksi risiko penyakit diabetes menggunakan algoritma pohon keputusan C4.5 berdasarkan faktor-faktor klinis dan gaya hidup. Data yang digunakan merupakan data sekunder dari Kaggle dengan total 200 data, terdiri atas 100 pasien diabetes dan 100 pasien non-diabetes. Atribut yang digunakan meliputi jenis kelamin, usia, hipertensi, penyakit jantung, riwayat merokok, indeks massa tubuh (BMI), level HbA1c, dan kadar gula darah. Data dinormalisasi ke dalam beberapa kategori sebelum dilakukan proses klasifikasi. Hasil penelitian menunjukkan bahwa atribut kadar gula darah memiliki nilai gain ratio tertinggi sehingga menjadi akar pohon keputusan dan faktor paling dominan dalam penentuan risiko diabetes. Individu dengan kadar gula darah normal dan risiko diklasifikasikan sebagai tidak diabetes, sedangkan kadar gula darah tinggi memerlukan evaluasi lanjutan terutama berdasarkan level HbA1c dan faktor pendukung lainnya. Penelitian ini menunjukkan bahwa algoritma C4.5 efektif digunakan sebagai alat bantu prediksi risiko diabetes dengan akurasi 82% dan berpotensi mendukung proses deteksi dini secara lebih cepat dan efisien.

Kata Kunci: diabetes, klasifikasi, pohon keputusan, algoritma C4.5, prediksi risiko

ABSTRACT

Diabetes is a chronic metabolic disease whose prevalence continues to increase and may lead to serious complications if not detected early. Routine health examinations are essential for early detection; however, they are often constrained by time and cost limitations. Therefore, this study aims to predict the risk of diabetes using the C4.5 decision tree algorithm based on clinical and lifestyle factors. The data used in this study are secondary data obtained from Kaggle, consisting of 200 records, including 100 diabetic patients and 100 non-diabetic patients. The attributes analyzed include gender, age, hypertension, heart disease, smoking history, body mass index (BMI), HbA1c level, and blood glucose level. Data normalization was performed by categorizing several attributes prior to the classification process. The results indicate that blood glucose level has the highest gain ratio and thus becomes the root of the decision tree, making it the most dominant factor in determining diabetes risk. Individuals with normal and at-risk blood glucose levels are classified as non-diabetic, whereas those with high blood glucose levels require further evaluation, particularly based on HbA1c levels and other supporting factors. This study demonstrates that the C4.5 algorithm is effective as a decision-support tool for predicting diabetes risk with 82% accuracy and has the potential to support early detection in a faster and more efficient manner.

Keywords: diabetes, classification, decision tree, C4.5 algorithm, risk prediction

This is an open access article under the [CC BY-SA](#) license.



*Corresponding author

1. PENDAHULUAN

Pemeriksaan kesehatan merupakan salah satu upaya penting untuk mendeteksi penyakit secara dini serta mencegah dan menghindari terjadinya komplikasi. Pemeriksaan kesehatan dapat dilakukan secara rutin, minimal enam bulan sekali, guna menjaga kondisi kesehatan tubuh. Hal ini menjadi penting terutama untuk mendeteksi penyakit yang disebabkan oleh gaya hidup tidak sehat, seperti dislipidemia, diabetes, dan hipertensi[1].

Diabetes merupakan gangguan metabolisme yang ditandai dengan tingginya kadar gula (glukosa) dalam darah. Gula darah berperan penting sebagai sumber energi bagi sel dan jaringan tubuh. Namun, apabila tidak terkontrol, diabetes dapat menimbulkan berbagai komplikasi serius, seperti gangguan pada mata, ginjal, dan saraf, serta meningkatkan risiko obesitas, penyakit jantung koroner, dan stroke[2].

Gula darah dibutuhkan tubuh agar dapat berfungsi dengan baik selama beraktivitas. Akan tetapi, jumlah asupan gula harus seimbang dengan tingkat aktivitas fisik. Terdapat beberapa faktor yang menyebabkan tubuh tidak mampu memetabolisme gula secara optimal, yaitu: (1) kurangnya aktivitas fisik, (2) konsumsi gula berlebihan, (3) resistensi insulin, (4) produksi insulin yang tidak mencukupi oleh pankreas, dan (5) gangguan kerja insulin akibat pengaruh hormon lain[3]. Penumpukan gula dalam darah inilah yang dapat meningkatkan risiko terjadinya diabetes. Oleh karena itu, pemeriksaan gula darah secara rutin sangat diperlukan untuk mendeteksi penyakit diabetes sejak dini.

Untuk menghemat biaya dan waktu, pendekatan penyakit diabetes dapat dilakukan melalui pendekatan prediksi berdasarkan faktor-faktor tertentu yang diklasifikasikan ke dalam beberapa kategori. Klasifikasi merupakan suatu proses pembentukan model atau fungsi untuk menentukan kategori atau kelas dari suatu data atau objek tertentu. Tujuan klasifikasi adalah memprediksi kelas dari objek yang belum memiliki label. Berbagai algoritma dapat digunakan dalam proses klasifikasi, salah satunya adalah algoritma pohon keputusan.

Pohon keputusan merupakan metode visual dalam pengambilan keputusan yang memetakan berbagai pilihan, kemungkinan hasil, serta konsekuensinya dalam bentuk struktur menyerupai pohon[4]. Metode ini membantu menyederhanakan permasalahan yang kompleks menjadi serangkaian langkah logis yang bercabang. Struktur pohon keputusan terdiri atas akar, cabang, dan daun. Akar merepresentasikan atribut atau pertanyaan utama, cabang mewakili pilihan atau kondisi, dan daun menunjukkan hasil keputusan akhir.

Beberapa penelitian terdahulu telah memanfaatkan algoritma C4.5 dalam berbagai permasalahan prediksi. Noviandi[5] memprediksi penyakit diabetes berdasarkan sejumlah faktor, seperti jumlah persalinan pada wanita, kadar gula darah, tekanan darah, kadar insulin, indeks massa tubuh (BMI), dan usia, dengan tingkat akurasi sebesar 70,32%. Hana[6] melakukan penelitian klasifikasi diabetes menggunakan pohon keputusan C4.5 dan memperoleh akurasi sebesar 97,12%, precision 93,02%, serta recall 100,00%. Sementara itu, Fauziah dan Zulfikar[7] menggunakan algoritma Iterative Dichotomiser 3 (ID3) untuk klasifikasi risiko penyakit jantung dan memperoleh akurasi sebesar 85,71%, presisi 84,62%, serta recall 84,62%.

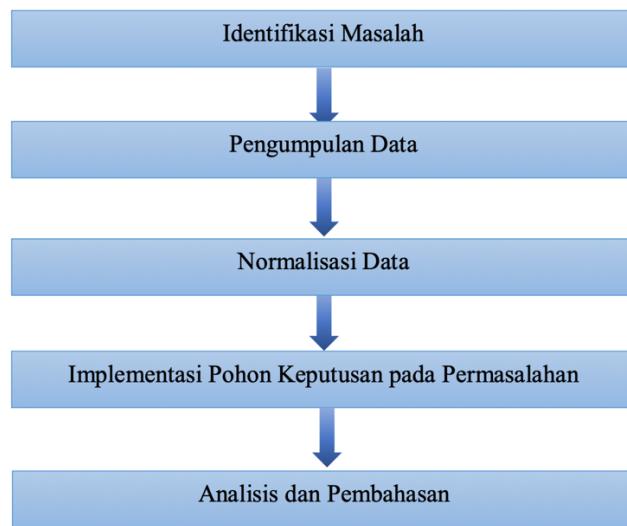
Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memprediksi risiko seseorang terkena penyakit diabetes berdasarkan berbagai faktor, yaitu jenis kelamin, usia, hipertensi, penyakit jantung, riwayat merokok, BMI, level HbA1c, dan kadar gula darah. Dengan adanya prediksi ini, diharapkan dapat dilakukan deteksi awal sebelum seseorang menjalani pemeriksaan kesehatan di fasilitas layanan kesehatan. Hasil penelitian ini diharapkan dapat menjadi alat bantu pendukung pengambilan keputusan bagi tenaga medis maupun masyarakat dalam upaya pencegahan dan pengendalian penyakit diabetes secara lebih efektif dan efisien.

2. METODE

Tahapan yang dilakukan pada penelitian ini ditunjukkan oleh Gambar 1 berikut.

Tahap awal pada penelitian adalah identifikasi masalah. Tahap ini bertujuan menganalisis dan mengevaluasi data dan informasi dilakukan guna peneliti dapat menemukan masalah-masalah terkait penyakit diabetes. Berdasarkan Survei Kesehatan Indonesia (SKI) pada 2023, didapatkan permasalahan diabetes sebagai berikut[8].

- 1) Penyakit diabetes pada penduduk usia di atas 15 tahun meningkat dari 2% ke 2,2% dan untuk semua umur meningkat dari 1,5% ke 1,7% jika dibandingkan dengan SKI 2018.
- 2) Terdapat 1,6% responden usia 18-59 tahun yang terdiagnosis diabetes. Namun hanya 1,46% yang melakukan pengobatan; 1,3% pengobatan sesuai petunjuk; dan 0,9% melakukan kunjungan ulang.
- 3) Terdapat 6,5% responden usia lebih dari 60 tahun yang terdiagnosis. Hanya 6,06% yang melakukan pengobatan; 5,46% pengobatan sesuai petunjuk; dan 4,12% yang melakukan kunjungan ulang.
- 4) Jika tidak ditindaklanjuti, diabetes dapat menyebabkan disabilitas hingga kematian.



Gambar 1. Tahapan Penelitian

Tahap selanjutnya adalah pengumpulan data terkait diabetes. Data yang digunakan merupakan data sekunder yang diperoleh dari situs Kaggle dengan jumlah total 100.000 data. Data tersebut terdiri atas pasien yang terkena dan tidak terkena diabetes dengan atribut jenis kelamin, usia, hipertensi, penyakit jantung, riwayat merokok, BMI, level HbA1c, dan kadar gula darah. Pada penelitian ini digunakan 200 data, yang terdiri atas 100 data pasien diabetes dan 100 data pasien non-diabetes. Setelah data terkumpul, dilakukan normalisasi data. Pada tahap ini, beberapa data perlu disesuaikan untuk memudahkan proses perhitungan. Beberapa data diubah dalam kategori seperti berikut.

- 1) Usia (*sumber: [9]*)
 - Anak-anak = < 14
 - Remaja = 15 – 24
 - Dewasa = 25 – 64
 - Senior = > 65
- 2) BMI (*sumber : [10]*)
 - Underweight = < 18,5
 - Normal Weight = 18,5 – 24,9
 - Overweight = 25,0 – 29,9
 - Obesity I = 30,0 – 34,9
 - Obesity II = 35,0 – 39,9
 - Obesity III = >40,0
- 3) HbA1c_level (*sumber: [11]*)
 - Normal = < 5,7
 - Risiko = 5,7 – 6,4
 - Indikasi = > 6,4
- 4) Level gula darah (*sumber: [12]*)
 - Normal = 80 – 100
 - Risiko = 101 – 125
 - Tinggi = > 125

Data yang telah dinormalisasi dibangun pohon keputusannya menggunakan Algoritma C4.5. Algoritma C4.5 menggunakan teknik membagi dan menaklukkan untuk membangun pohon keputusan. Dengan menghitung dan membandingkan rasio penguatan, algoritma ini memilih opsi terbaik. Teknik membagi dan menaklukkan kemudian dijalankan kembali pada node yang dihasilkan pada level berikutnya. Berikut langkah-langkah dalam mengembangkan pohon keputusan dengan algoritma C4.5[13].

- 1) Menentukan akar dari pohon dengan menghitung nilai gain tertinggi dari masing-masing atribut atau berdasarkan nilai *indeks entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*.

$$\text{Entropy} (S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Keterangan:

S = himpunan kasus

n = jumlah partisi S

p_i = proposisi S_i

- 2) Setelah itu hitung nilai gain dengan metode informasi gain dengan persamaan gain.

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

Keterangan:

A = Fitur

n = Jumlah partisi atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = Jumlah kasus dalam S

- 3) Selanjutnya, hitung nilai dari *SplitInfo* seperti berikut untuk digunakan dalam perhitungan *Gain Ratio*.

$$\text{SplitInfo} (S, A) = -\frac{S_i}{S} \log_2 \left(\frac{S_i}{S} \right)$$

- 4) Langkah selanjutnya adalah menentukan akar dan cabang dari pohon. Akar dipilih dari atribut yang memiliki *Gain Ratio* tertinggi dari atribut yang belum terpilih sebelumnya sedangkan atribut yang memiliki *Gain Ratio* lebih rendah akan menjadi cabang.

$$\text{Gain Ratio} (S, A) = \frac{\text{GainInfo}(S, A)}{\text{SplitInfo}(S, A)}$$

- 5) Ulangi perhitungan *entropy* dan gain pada atribut-atribut yang lain.
 6) Proses perhitungan di atas diulangi sampai semua atribut terpartisi.
 7) Proses akan berhenti apabila semua record dalam sampel N mendapat kelas yang sama, tidak ada atribut yang tidak terpartisi dan tidak ada record dalam cabang yang kosong.

Tahap akhir pada penelitian ini adalah analisis dan pembahasan. Pada tahap ini akan dianalisis hasil dari pohon keputusan yang telah dibangun. Dari tahap ini, didapat faktor-faktor apa saja yang paling mempengaruhi diabetes. Dengan pohon keputusan yang telah didapat, dapat diprediksi lebih awal apakah seseorang terkena diabetes atau tidak.

3. HASIL DAN PEMBAHASAN

Dalam membangun pohon keputusan, ada beberapa kondisi yang digunakan dalam penelitian ini, antara lain sebagai berikut.

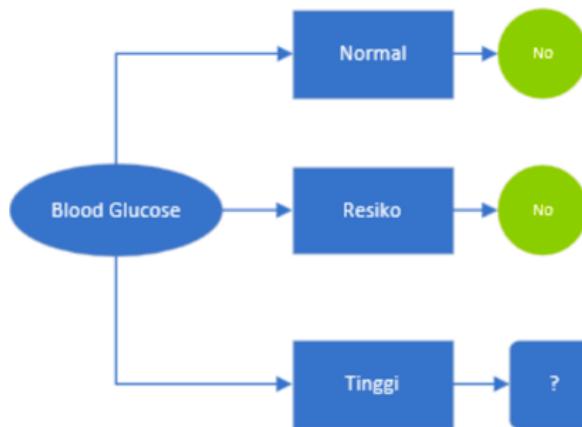
- 1) Jika suatu kategori atau kelas dalam suatu atribut tidak ada dalam perhitungan, dalam artian tidak ada data mengenai kelas tersebut, maka keputusannya adalah 0 (tidak diabetes).
- 2) Jika pada akhir perhitungan atribut, gain rationya didapat 0, yang berarti atribut ini tidak memiliki anak, maka keputusan diambil berdasarkan probabilitas terbesar antara 0 dan 1.
- 3) Jika probabilitas 0 dan 1 pada poin 2 didapat 50:50, maka keputusan yang diambil adalah 0 (tidak diabetes).

Setelah data dinormalisasi, maka dilakukan perhitungan *entropy* dan *gain ratio*. Hasil perhitungan *entropy* dan *gain ratio* awal ditunjukkan oleh Tabel 1. Berdasarkan hasil perhitungan nilai gain ratio tersebut, maka nilai terbesarnya yang digunakan sebagai akar dari pohon keputusan adalah level gula darah, yakni sebesar 0,708529116 sehingga atribut level gula darah digunakan sebagai akar. Kemudian buat cabang sesuai banyaknya partisi pada atribut level gula darah, terdapat tiga partisi yakni normal, risiko, dan tinggi sehingga dari akar level gula darah akan dibentuk tiga cabang yakni normal, risiko, dan tinggi.

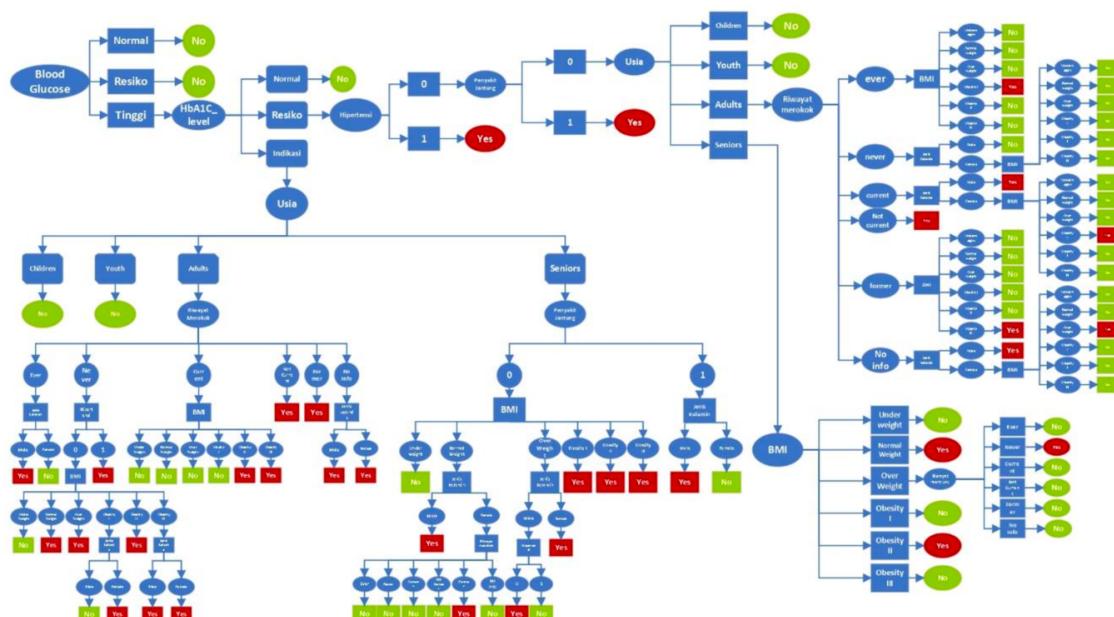
Tabel 1. Perhitungan Entropy dan Gain Ratio

		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
jenis kelamin								0,317883865
	male	81	44	37	0,994605981	0,402815422	-0,528122506	
	female	119	56	63	0,997502546	0,279300713	-0,445677864	
						0,682116135	0,973800369	0,326436377
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
usia								0,515681932
	CHILDREN	13	0	13	0	0	-0,256322071	
	YOUTH	8	0	8	0	0	-0,185754248	
	ADULTS	117	59	58	0,999947304	0,294984455	-0,45249251	
	SENIORS	62	41	21	0,9235786	0,189333613	-0,523794563	
						0,484318068	1,418363391	0,363575326
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
Hipertensi								0,074116186
	0	169	75	94	0,990863141	0,837279354	-0,205315357	
	1	31	25	6	0,708835673	0,088604459	-0,416897281	
						0,925883814	0,622212638	0,119117134
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
penyakit jantung								0,027997413
	0	186	91	95	0,999666365	0,929689719	-0,097368562	
	1	14	9	5	0,940285959	0,042312868	-0,268555089	
						0,972002587	0,365923651	0,076511624
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
riwayat merokok								0,509462347
	ever	10	6	4	0,970950594	0,04854753	-0,216096405	
	never	84	44	40	0,998363673	0,219640008	-0,525646282	
	current	15	8	7	0,996791632	0,039871665	-0,28027242	
	not current	16	12	4	0,811278124	0,048676687	-0,291508495	
	former	28	18	10	0,940285959	0,084625736	-0,397110177	
	no info	47	12	35	0,819600435	0,049176026	-0,490977824	
						0,490537653	2,201611604	0,23140428
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
BMI								0,540163873
	UNDERWEIGHT	5	0	5	0	0	-0,133048202	
	NORMAL WEIGHT	30	11	19	0,948078244	0,052144303	-0,410544839	
	OVERWEIGHT	73	32	41	0,98900768	0,158241229	-0,530721545	
	OBESITY I	27	14	12	1,011279857	0,07078959	-0,390010773	
	OBESITY II	22	19	4	0,629832022	0,059834042	-0,350286703	
	OBESITY III	43	24	19	0,99022469	0,118826963	-0,476782159	
						0,459836127	2,291394221	0,235735897
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
HbA1c_level								0,603478281
	NORMAL	35	0	35	0	0	-0,440050305	
	RESIKO	92	73	19	0,734781352	0,268195194	-0,515335348	
	INDIKASI	73	27	46	0,950566853	0,128326525	-0,530721545	
						0,396521719	1,486107198	0,406079913
Total		Jumlah (S)	Ya (S1)	Tidak (S2)	Entropy	Info	Splitinfo	Gain
Total		200	100	100	1			
Blood_glucose_level								0,510423639
	NORMAL	21	0	21	0	0	-0,341411571	
	RESIKO	8	0	8	0	0	-0,185754248	
	TINGGI	171	100	71	0,979152723	0,489576361	-0,193233142	
						0,489576361	0,72039896	0,708529116

Untuk partisi normal dan risiko memiliki nilai 0 pada kasus Ya sehingga dapat dipastikan untuk level gula darah normal dan resiko semuanya akan menghasilkan label Tidak. Sedangkan untuk partisi tinggi, masih terdapat dua keputusan antara Ya dan Tidak, maka perlu melihat atribut yang lain. Adapun atribut yang perlu dipertimbangkan lainnya adalah jenis kelamin, usia, hipertensi, penyakit jantung, riwayat merokok, BMI, dan level gula darah sementara untuk atribut level gula darah tidak perlu dihitung kembali karena telah menjadi simpul. Hitung *entropy* dan *gain ratio* dari masing-masing atribut dengan jumlah kasus yang baru (yang telah direduksi dari sebelumnya). Gambar 2 menunjukkan percabangan dari gula darah dengan partisi Normal dan Risiko menghasilkan keputusan Tidak Diabetes.



Gambar 2. Percabangan Gula Darah dengan Tiga partisi: Normal, Risiko, dan Tinggi



Gambar 3. Hasil Pohon Keputusan

Gambar 3 menunjukkan hasil pohon keputusan yang dibangun. Pada Gambar 3, terlihat bahwa faktor utama yang mempengaruhi diabetes adalah gula darah. Hasil pohon keputusan menunjukkan bahwa kadar gula darah merupakan faktor utama dalam penentuan risiko diabetes. Seseorang dengan kadar gula darah normal (80–100) atau risiko (101–125) dapat langsung diklasifikasikan sebagai tidak diabetes. Namun, apabila kadar gula darah lebih dari 125, maka diperlukan pemeriksaan lanjutan menggunakan atribut lain, terutama level HbA1c. Jika level HbA1c berada pada kategori normal, maka individu tersebut diklasifikasikan sebagai tidak diabetes. Sebaliknya, jika level HbA1c berada pada kategori risiko atau indikasi, maka diperlukan evaluasi lanjutan terhadap faktor lain seperti hipertensi, penyakit jantung, usia, dan riwayat merokok untuk menentukan keputusan akhir.

Setelah diperoleh pohon keputusan, dilakukan pengujian terhadap hasil prediksi yang dihasilkan. Pengujian ini menggunakan 50 data uji sebagaimana ditunjukkan pada Tabel 2. Dari total 50 data uji tersebut, sebanyak 41 data berhasil diprediksi dengan benar, sedangkan 9 data lainnya tidak sesuai dengan data aktual. Dengan demikian, tingkat akurasi model yang diperoleh adalah sebesar **82%**.

Tabel 2. Hasil pengujian dengan 50 data uji

jenis kelamin	usia		hipertensi	penyakit jantung	riwayat merokok	BMI	HbA1c_level		Blood_glucose_level		diabetes	Predksi
Male	38.0	38	ADULTS	0	0	never	27,63	OVERWEIGHT	4,00	NORMAL	145	TINGGI 0 NO
Female	51.0	51	ADULTS	1	0	No Info	28,67	OVERWEIGHT	6,10	RESIKO	145	TINGGI 0 NO
Male	38.0	38	ADULTS	0	0	not current	25,34	OVERWEIGHT	6,50	INDIKASI	85	NORMAL 0 NO
Male	15.0	15	YOUTH	0	0	No Info	18,25	UNDERWEIGHT	6,60	INDIKASI	100	RESIKO 0 NO
Male	54.0	54	ADULTS	0	0	not current	27,32	OVERWEIGHT	6,10	RESIKO	100	RESIKO 0 NO
Female	45.0	45	ADULTS	0	0	former	32,81	OBESEITY I	4,50	NORMAL	145	TINGGI 0 NO
Female	51.0	51	ADULTS	0	0	never	21,49	NORMAL WEIGHT	6,20	RESIKO	155	TINGGI 0 NO
Female	30.0	30	ADULTS	0	0	never	22,88	NORMAL WEIGHT	4,50	NORMAL	80	NORMAL 0 NO
Male	32.0	32	ADULTS	0	0	ever	37,71	OBESEITY II	6,10	RESIKO	126	TINGGI 0 NO
Female	57.0	57	ADULTS	0	0	not current	31,78	OBESEITY I	5,00	NORMAL	126	TINGGI 0 NO
Female	23.0	23	YOUTH	0	0	never	21,30	NORMAL WEIGHT	6,20	RESIKO	126	TINGGI 0 NO
Female	51.0	51	ADULTS	0	0	never	27,32	OVERWEIGHT	4,00	NORMAL	140	TINGGI 0 NO
Female	51.0	51	ADULTS	0	0	No Info	27,32	OVERWEIGHT	6,00	RESIKO	85	NORMAL 0 NO
Male	13.0	13	CHILDREN	0	0	never	21,01	NORMAL WEIGHT	4,80	NORMAL	140	TINGGI 0 NO
Female	22.0	22	YOUTH	0	0	No Info	24,46	NORMAL WEIGHT	6,20	RESIKO	160	TINGGI 0 NO
Male	14.0	14	CHILDREN	0	0	never	23,67	NORMAL WEIGHT	5,80	RESIKO	145	TINGGI 0 NO
Female	52.0	52	ADULTS	0	0	never	27,32	OVERWEIGHT	6,10	RESIKO	145	TINGGI 0 NO
Male	26.0	26	ADULTS	0	0	No Info	24,08	NORMAL WEIGHT	5,80	RESIKO	90	NORMAL 0 NO
Female	80.0	80	SENIORS	0	0	never	30,00	OBESEITY I	6,00	RESIKO	126	TINGGI 0 NO
Female	21.0	21	YOUTH	0	0	No Info	27,32	OVERWEIGHT	5,70	RESIKO	126	TINGGI 0 YES
Female	56.0	56	ADULTS	0	0	former	42,03	OBESEITY II	6,50	INDIKASI	130	TINGGI 0 YES
Female	21.0	21	YOUTH	0	0	No Info	45,92	OBESEITY II	5,70	RESIKO	130	TINGGI 0 NO
Female	26.0	26	ADULTS	0	0	ever	45,90	OBESEITY II	5,00	NORMAL	160	TINGGI 0 NO
Female	53.0	53	ADULTS	0	0	No Info	49,79	OBESEITY II	5,70	RESIKO	159	TINGGI 0 NO
Female	58.0	58	ADULTS	1	0	never	38,31	OBESEITY II	7,00	INDIKASI	200	TINGGI 1 YES
Female	61.0	61	ADULTS	0	0	current	30,11	OBESEITY I	6,20	RESIKO	240	TINGGI 1 YES
Female	61.0	61	ADULTS	0	0	No Info	34,45	OBESEITY I	6,50	INDIKASI	280	TINGGI 1 YES
Male	55.0	55	ADULTS	0	1	former	30,42	OBESEITY I	6,20	RESIKO	300	TINGGI 1 YES
Female	65.0	65	SENIORS	1	1	never	33,55	OBESEITY I	8,20	INDIKASI	140	TINGGI 1 NO
Female	37.0	37	ADULTS	0	0	No Info	36,87	OBESEITY II	8,80	INDIKASI	160	TINGGI 1 YES
Female	74.0	74	SENIORS	0	0	current	29,51	OVERWEIGHT	8,20	INDIKASI	130	TINGGI 1 YES
Male	67.0	67	SENIORS	0	0	current	27,32	OVERWEIGHT	6,80	INDIKASI	220	TINGGI 1 YES
Male	64.0	64	ADULTS	1	0	former	33,12	OBESEITY I	5,70	RESIKO	300	TINGGI 1 YES
Female	71.0	71	SENIORS	0	0	never	32,00	OBESEITY I	5,80	RESIKO	155	TINGGI 1 NO
Female	43.0	43	ADULTS	1	0	never	34,21	OBESEITY I	6,50	INDIKASI	160	TINGGI 1 YES
Male	58.0	58	ADULTS	0	0	former	32,38	OBESEITY I	6,60	INDIKASI	159	TINGGI 1 YES
Male	63.0	63	ADULTS	0	1	No Info	27,32	OVERWEIGHT	6,60	INDIKASI	300	TINGGI 1 YES
Female	80.0	80	SENIORS	0	0	never	22,04	NORMAL WEIGHT	8,20	INDIKASI	130	TINGGI 1 NO
Female	80.0	80	SENIORS	1	0	never	21,33	NORMAL WEIGHT	6,80	INDIKASI	240	TINGGI 1 NO
Female	70.0	70	SENIORS	0	0	current	22,15	NORMAL WEIGHT	6,00	RESIKO	240	TINGGI 1 YES
Female	50.0	50	ADULTS	0	0	never	35,91	OBESEITY II	6,00	RESIKO	280	TINGGI 1 NO
Female	65.0	65	SENIORS	1	0	never	35,00	OBESEITY II	7,00	INDIKASI	145	TINGGI 1 YES
Male	80.0	80	SENIORS	0	1	never	25,76	OVERWEIGHT	9,00	INDIKASI	220	TINGGI 1 YES
Male	61.0	61	ADULTS	1	1	former	35,78	OBESEITY II	6,00	RESIKO	145	TINGGI 1 YES
Male	58.0	58	ADULTS	0	0	never	49,25	OBESEITY II	8,20	INDIKASI	126	TINGGI 1 YES
Male	62.0	62	ADULTS	0	0	not current	27,32	OVERWEIGHT	6,00	RESIKO	300	TINGGI 1 YES
Female	71.0	71	SENIORS	0	0	No Info	64,80	OBESEITY II	9,00	INDIKASI	280	TINGGI 1 YES
Female	66.0	66	SENIORS	0	0	former	44,64	OBESEITY II	6,00	RESIKO	155	TINGGI 1 NO
Female	54.0	54	ADULTS	1	1	never	41,81	OBESEITY II	6,50	INDIKASI	300	TINGGI 1 YES

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma pohon keputusan C4.5 dapat digunakan untuk memprediksi risiko penyakit diabetes secara efektif berdasarkan faktor-faktor tertentu. Dari hasil perhitungan gain ratio, atribut kadar gula darah menjadi faktor paling dominan dalam penentuan risiko diabetes dan berperan sebagai akar pada pohon keputusan yang dibangun. Hasil klasifikasi menunjukkan bahwa individu dengan kadar gula darah normal dan risiko dapat langsung diklasifikasikan sebagai tidak diabetes. Sementara itu, individu dengan kadar gula darah tinggi memerlukan pemeriksaan lanjutan menggunakan atribut lain, terutama level HbA1c, serta faktor pendukung seperti usia, hipertensi, penyakit jantung, dan riwayat merokok. Hal ini menunjukkan bahwa kombinasi faktor klinis dan gaya hidup berpengaruh dalam menentukan risiko diabetes. Penelitian ini diharapkan dapat menjadi alat bantu pendukung pengambilan keputusan bagi tenaga medis maupun masyarakat dalam melakukan deteksi dini penyakit diabetes. Dengan adanya sistem prediksi ini, proses skrining awal dapat dilakukan secara lebih cepat, efisien, serta berpotensi mengurangi biaya dan waktu pemeriksaan kesehatan. Untuk penelitian selanjutnya, disarankan menggunakan jumlah data yang lebih besar serta membandingkan performa algoritma C4.5 dengan metode klasifikasi lainnya guna meningkatkan akurasi dan keandalan hasil prediksi.

REFERENSI

- [1] “Direktorat Jenderal Kesehatan Lanjutan.” Accessed: Jan. 09, 2026. [Online]. Available: https://keslan.kemkes.go.id/view_artikel/1839
- [2] N. Nur Rosyidah, P. Studi Kebidanan, S. Tinggi Ilmu Kesehatan Dian Husada Mojokerto, E. Agus Cahyono, P. Studi Keperawatan, and A. Keperawatan Dian Husada Mojokerto, “DIABETES MELITUS TIPE 2 ; ARTIKEL REVIEW,” *Enfermeria Ciencia*, vol. 3, no. 1, pp. 44–63, Feb. 2025, doi: 10.56586/EC.V3I1.74.
- [3] “Direktorat Jenderal Kesehatan Lanjutan.” Accessed: Jan. 09, 2026. [Online]. Available: https://keslan.kemkes.go.id/view_artikel/3230/mengenal-komplikasi-diabetes-melitus
- [4] S. N. Sari, P. Annisa, A. Nisa, D. Rahma, R. Prameswara Ritonga, and D. P. Utomo, “Teknik Data Mining Dengan Menggunakan Algoritma Decision Tree Untuk Mengetahui Pola Pemahaman Mahasiswa Pada Mata Kuliah Pemrograman,” *Bulletin of Information Technology (BIT)*, vol. 6, no. 4, pp. 417–431, Dec. 2025, doi: 10.47065/BIT.V6I4.2339.

- [5] N. Program, S. R. Medis, I. Kesehatan, and I.-I. Kesehatan, "Implementasi Algoritma Decision Tree C4.5 Untuk Prediksi Penyakit Diabetes," *Indonesian of Health Information Management Journal (INOHIM)*, vol. 6, no. 1, pp. 1–5, 2018, Accessed: Jan. 09, 2026. [Online]. Available: <https://inohim.esaunggul.ac.id/index.php/INO/article/view/142>
- [6] U. M. Kudus, J. Ganesha, and P. Kudus, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, Oct. 2020, doi: 10.47970/SISKOM-KB.V4I1.173.
- [7] E. Fauziah and A. F. Zulfikar, "Penerapan Metode Decision Tree Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) Untuk Klasifikasi Resiko Penyakit Jantung," *OKTAL : Jurnal Ilmu Komputer dan Sains*, vol. 2, no. 04, pp. 1207–1219, Apr. 2023, Accessed: Jan. 09, 2026. [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal/article/view/1302>
- [8] "Potret Sehat Indonesia dari Kacamata SKI 2023 - Badan Kebijakan Pembangunan Kesehatan | BKPK Kemenkes." Accessed: Jan. 09, 2026. [Online]. Available: <https://www.badankebijakan.kemkes.go.id/potret-sehat-indonesia-dari-kacamata-ski-2023/>
- [9] "Previous standard - age." Accessed: Jan. 09, 2026. [Online]. Available: <https://www.statcan.gc.ca/en/concepts/definitions/previous/age1a>
- [10] K. Coulman and S. Toran, "Body mass index may not be the best indicator of our health – how can we improve it?," Jul. 2020, doi: 10.64628/AB.VQE5GG996.
- [11] "HbA1C Test - Procedure, Normal Range, and Benefits." Accessed: Jan. 09, 2026. [Online]. Available: <https://www.apollohospitals.com/diagnostics-investigations/hba1c-test>
- [12] "Normal Blood Sugar Levels Chart: Sugar Level Chart." Accessed: Jan. 09, 2026. [Online]. Available: <https://www.medkart.in/blog/normal-blood-sugar-levels-chart>
- [13] A. Afifuddin and L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5," *Krisnadana Journal*, vol. 3, no. 1, pp. 25–33, Sep. 2023, doi: 10.58982/KRISNADANA.V3I1.470.