

Klasifikasi Diabetes Melitus Menggunakan Metode Decision Tree Dan C4.5 Di Kabupaten Grobogan

Zahra Fatma Saniya^{*1}, Eko Supriyadi², Kartika Imam Santoso³

^{1,2,3}Program Studi Ilmu Komputer, Universitas An Nuur, Grobogan, Indonesia

e-mail: ^{*1}zahrasaniya0702@gmail.com, ²ekalaya56@gmail.com,

³kartikaimams@gmail.com

Abstrak

Penelitian ini muncul karena tingginya kasus diabetes yang jadi masalah kesehatan penting di masyarakat. Deteksi awal dengan metode klasifikasi penting agar penanganan medis bisa lebih cepat dan tepat. Penelitian ini bertujuan untuk tahu seberapa akurat algoritma Decision Tree dan C4.5 dalam mengklasifikasi diabetes, cari faktor utama yang pengaruhi klasifikasi, serta analisis potensi bias dalam data. Data riset didapat dari posyandu dan kuesioner masyarakat di Kecamatan Penawangan, Grobogan. Proses klasifikasi dilakukan pakai aplikasi RapidMiner dengan tahapan persiapan data, buat model, hingga evaluasi pakai *confusion matrix*. Hasil riset menunjukkan algoritma Decision Tree C4.5 hasilkan akurasi 96,49%. Faktor terkuat yang pengaruhi status diabetes yaitu umur, tinggi badan, dan riwayat keluarga. Hasil riset ini tunjukkan algoritma C4.5 bisa dipakai jadi metode yang baik untuk bantu klasifikasi diabetes. Dengan begitu, riset ini harapannya bisa sumbangsih dalam pengembangan sistem deteksi awal dan upaya pencegahan serta edukasi masyarakat tentang bahaya diabetes.

Kata kunci — C4.5, Decision Tree, Diabetes Melitus, Klasifikasi, RapidMiner,

Abstract

This research was conducted due to the high number of diabetes cases, which is a significant health problem in society. Early detection using classification methods is important so that medical treatment can be provided more quickly and accurately. This study aims to determine the accuracy of the Decision Tree and C4.5 algorithms in classifying diabetes, identify the main factors that influence classification, and analyze potential bias in the data. The research data was obtained from health centers and community questionnaires in Penawangan District, Grobogan. The classification process was carried out using the RapidMiner application with stages of data preparation, model creation, and evaluation using a confusion matrix. The results showed that the Decision Tree C4.5 algorithm produced an accuracy of 96.49%. The strongest factors influencing diabetes status were age, height, and family history. These research results show that the C4.5 algorithm can be used as a good method to assist in diabetes classification. Thus, this research is expected to contribute to the development of early detection systems and prevention efforts, as well as public education about the dangers of diabetes.

Keywords — C4.5, Classification, Decision Tree, Diabetes Mellitus, RapidMiner

1. PENDAHULUAN

Diabetes Militus menjadi salah satu penyakit kronis dengan prevelensi tinggi yang dihadapi oleh masyarakat global. Penyakit ini timbul akibat meningkatnya kadar gula darah yang tidak terkontrol, yang jika diabaikan dapat memicu berbagai komplikasi serius. Oleh sebab itu. Deteksi sejak dini serta penanganan yang sesuai menjadi hal yang sesuai menjadi sangat penting untuk dilakukan. Dalam konteks ini, diperlukan metode diagnosis yang akurat dan efisien, sehingga upaya penanganan medis dapat diberikan sedini mungkin.[1]

Penelitian ini mengadopsi pendekatan kecerdasan buatan yaitu melalui metode klasifikasi yang digunakan dalam penelitian ini adalah algoritma decision tree dan C4.5, dengan tujuan untuk menentukan status pasien diabetes secara lebih terstruktur dan akurat, berdasarkan data yang tersedia berdasarkan sejumlah fitur klinis seperti jenis kelamin, usia, riwayat keturunan, dan aktifitas fisik[2].

Berdasarkan latarbelakang tersebut, rumusan masalah dan tujuan penelitian ini disusun untuk memberikan arah yang jelas dalam proses pengumpulan data, analisis, serta penarikan kesimpulan dari hasil penelitian, penelitian ini berfokus pada tiga pokok pertanyaan, yaitu: (1) Sejauh mana tingkat akurasi algoritma Decision Tree dan C4.5 dalam melakukan klasifikasi penyakit diabetes di wilayah Kecamatan Penawangan, Grobogan (2) Faktor-faktor yang memiliki pengaruh paling dominan perlu diidentifikasi agar dapat dipahami variabel mana yang memberikan kontribusi terbesar terhadap hasil analisis dan (3) Apakah terdapat bias dalam dataset yang diperoleh yang diperoleh dari data posyandu dan kuisioner masyarakat.

Tujuan dari penelitian ini adalah menjawab ketiga pertanyaan tersebut, yakni mengukur tingkat akurasi model, mengidentifikasi, faktor dominan, dan menganalisis potensi bias pada data, sehingga mampu menghasilkan metode klasifikasi yang akurat untuk mendukung diagnosis diabetes.

Tinjauan pustaka telah memuat kajian yang relevan mengenai konsep data mining, klasifikasi, serta landasan teori mendalam tentang Algoritma Decision Tree dan pengembangannya, yaitu Algoritma C4.5 secara spesifik, Decision Tree dan C4.5 dipilih karena kemampuannya mempresentasikan proses pengambilan keputusan disajikan dalam bentuk struktur pohon bersifat sederhana sehingga dapat mudah di pahami serta dapat diinterpretasikan dengan jelas, sementara C4.5 yang menggunakan gain ratio dijadikan landasan untuk memilih atribut paling berpengaruh dalam membangun model klasifikasi.[3] Kajian ini juga didukung oleh penelitian-penelitian terkait yang menunjukkan efektivitas algoritma tersebut dalam konteks klasifikasi penyakit.

2. METODE PENELITIAN

Penelitian ini memakai pendekatan mixed method, yaitu gabungan antara metode kuantitatif dan kualitatif agar hasil yang diperoleh lebih menyeluruh dan komperhensif. Desain penelitian yang dipilih berbentuk eksploratori sekensual, dimana tahap awal dilakukan dengan metode kuantitatif untuk memperoleh gambaran analisis berbasis angka. Setelah itu, penelitian dilanjutkan dengan metode kualitatif yang berfungsi memperdalam serta menjelaskan makna dari temuan yang telah diperoleh sebelumnya. Pendekatan kuantitatif digunakan untuk menguji nilai serta hasil perhitungan sesuai prosedur yang berlaku, sedangkan pendekatan kuantitatif membantu memberikan hasil perhitungan sesuai prosedur yang berlaku, sedangkan pendekatan kualitatif membantu memberikan penjelasan yang lebih luas dan mendalam mengenai fenomena yang diteliti.[4]

Metode penelitian menjabarkan rencana kegiatan, batasan atau objek studi, material dan alat utama, lokasi riset, cara pengumpulan data, definisi operasional variabel, hingga metode analisis yang digunakan. Semua ini dirancang untuk menjelaskan tahapan riset secara sistematis supaya tujuan riset tercapai. Alur langkah-langkah ini juga divisualisasikan dalam flowchart agar lebih mudah diikuti..[5]

2.1. Jenis Penelitian

Penelitian ini menerapkan pendekatan mixed method, yakni menggabungkan metode kuantitatif dan kualitatif. Desain yang digunakan yaitu eksploratori sekuenasial, dimana penelitian diawali dengan tahap kuantitatif untuk memperoleh hasil analisis numerik, kemudian dilanjutkan dengan tahap kualitatif sebagai pendalaman makna dari hasil yang diperoleh. Metode kuantitatif bertujuan untuk menguji nilai serta hasil analisis berdasarkan rumus dan

prosedur yang ada, sedangkan metode kualitatif dimanfaatkan untuk memberikan gambaran yang lebih menyeluruh terkait fenomena yang menjadi fokus penelitian.

2.2. Fokus Penelitian

Fokus penelitian diarahkan pada pertanyaan-pertanyaan utama yang relevan dengan topik kajian, sehingga proses pengamatan maupun analisis dapat berjalan secara sistematis dan terarah. Fokus ini berfungsi sebagai pedoman agar pembahasan tidak melebar ke luar konteks, namun keseluruhan proses tetap di arahkan agar sejalan dengan judul serta tujuan penelitian yang telah di tetapkan. Dalam penelitian ini diterapkan metode Decision tree dengan algoritma C4.5 dipilih sebagai alat analisis utama, karena dianggap mampu memberikan gambaran yang jelas mengenai pola data serta mendukung proses klasifikasi yang akurat.

2.3. Sumber Data

Penelitian ini menggunakan data sekunder sebagai dasar dalam proses analisis, yaitu informasi diperoleh secara tidak langsung melalui pihak lain. Data tersebut dikumpulkan melalui arsip-arsip yang tersedia di posyandu, dan mengambil data dari responden langsung melalui kuisioner. Dalam penelitian ini, data yang digunakan memiliki beberapa atribut penting yang berperan sebagai variabel analisis antarlain: Nama, Jenis Kelamin, Usia, Tinggi Badan (TB), Berat Badan(BB), Riwayat Keturunan Diabetes, Aktivitas Fisik, Serta Status Diabetes. Adapun kelas yang menjadi target klasifikasi terdiri dari dua katagori, yaitu positif diabetes dan negatif diabetes.

2.4. Desain Sistem

Berikut merupakan seluruh rangkaian kegiatan penelitian yang dilakukan



Gambar 1 Desain Sistem

2.5. Pengumpulan Data

Metode penelitian menjabarkan rencana kegiatan, batasan atau objek studi, material dan alat utama, lokasi riset, cara pengumpulan data, definisi operasional variabel, hingga metode analisis yang digunakan. Semua ini dirancang untuk menjelaskan tahapan riset secara sistematis

supaya tujuan riset tercapai. Alur langkah-langkah ini juga divisualisasikan dalam flowchart agar lebih mudah diikuti.[6]

2.6. *Pra-Pemrosesan Data*

Tahap selanjutnya setelah data terkumpul adalah pra-pemrosesan data, yang menjadi langkah penting untuk memastikan data berada dalam kondisi baik dan siap digunakan pada proses analisis. Proses ini meliputi penanganan nilai-nilai yang hilang agar tidak menimbulkan bias atau kesalahan interpretasi dalam analisis. Data yang tidak lengkap ditangani melalui pengisian data yang dilakukan dengan metode yang tepat, misalnya menggunakan nilai rata-rata atau pendekatan lain yang dianggap relevan sesuai dengan karakteristik. Selain itu, dilakukan pula normalisasi data serta penghapusan data pencilon (outlier) apabila diperlukan, sehingga kualitas data lebih terjaga.[7]

2.7. *Pembentukan Model Decision Tree*

Setelah tahap pra-pemrosesan selesai, langkah berikutnya adalah menentukan model klasifikasi dengan metode Decision Tree. Model ini dibangun dengan memilih atribut yang paling berpengaruh dalam memisahkan data ke dalam kelompok yang lebih homogen. Setiap percabangan pohon ditentukan berdasarkan kriteria tertentu, seperti ukuran ketidakpastian atau tingkat kemurnian data. Struktur pohon yang dihasilkan kemudian digunakan untuk memprediksi status diabetes pada pasien berdasarkan atribut-atribut yang dimiliki. Dengan demikian, Decision Tree menjadi alat yang efektif dalam membantu pasien dan mendukung analisis lebih lanjut terkait faktor-faktor yang berhubungan dengan diabetes.[8]

2.8. *Evaluasi*

Tahap berikutnya setelah model Decision Tree terbentuk adalah evaluasi untuk menilai kinerja model. Proses evaluasi dilakukan dengan menggunakan data uji yang sebelumnya tidak dilibatkan dalam proses pelatihan, sehingga hasil pengujian lebih objektif. Beberapa metrix digunakan dalam tahap ini diantaranya akurasi, presisi, dan recall. Akurasi menggambarkan tingkat ketepatan model dalam mengklasifikasi pasien secara keseluruhan, presisi menunjukkan sejauh mana model bisa memprediksi kasus positif dengan benar, sementara itu, recall digunakan untuk menilai sejauh mana model mampu mengenali seluruh pasien yang benar-benar teridentifikasi positif diabetes. Dengan evaluasi ini, keandalan model dalam mendiagnosis dapat diketahui dan menjadi dasar untuk perbaikan. [9]

3. HASIL DAN PEMBAHASAN

Dataset ini terdiri dari 331 data posyandu dan responden di kabupaten Grobogan khususnya di kecamatan Penawangan yang mencakup informasi dasar kesehatan dan gaya hidup. Atribut yang digunakan untuk klasifikasi yaitu: Nama, Jenis kelamin, Usia, TB, BB Aktifitas Fisik, Keturunan Diabetes, Status Diabetes.

3.1 *Pengumpulan Data*

Setelah data klinis di kumpulkan, langkah selanjutnya pra-pemrosesan data, yang berfungsi untuk membersihkan dan menyiapkan data agar lebih rapi serta terstruktur sehingga siap digunakan dalam tahap analisis. misalnya dengan mengatasi data kosong, duplikasi, atau transformasi nilai. Selanjutnya, dilakukan proses split data dilakukan dengan membagi dataset menjadi dua bagian, yaitu data training dan data testing untuk menguji kinerjanya. Langkah ini bertujuan agar diperoleh nilai akurasi serta gambaran kemampuan model dalam memprediksi data baru.

Tabel 1 Sampel Data Penyakit Diabetes Melitus

NO	NAMA	JK	USIA	TB	BB	AKTIFITAS FISIK	KETURUNAN DIABETES	STATUS DIABETES
1	MFK	P	21	165	46	Sedang	Tidak	Tidak
2	KH	P	24	170	60	Sedang	Tidak	Tidak
3	HAD	P	22	155	54	Sedang	Ya	Tidak
4	D	P	21	160	58	Sedang	Tidak	Tidak
-----	-----	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----
327	N	L	54	165	61	Ringan	Ya	Tidak
328	S	P	50	154	57	Ringan	Ya	Tidak
329	A	L	63	170	57	Sedang	Tidak	Ya
330	N	P	22	150	67	Ringan	Tidak	Tidak
331	NW	P	22	150	67	Ringan	Tidak	Tidak

3.2 Pra-Pemrosesan Data

Setelah pengumpulan data klinis, langkah selanjutnya adalah pra-pemrosesan data. Tahap ini melibatkan pembersihan dan penyiapan data agar lebih terstruktur, misalnya dengan menangani nilai yang hilang, duplikasi data, atau transformasi nilai. Kemudian, dilakukan pembagian data menjadi dua bagian: data pelatihan yang dipakai untuk melatih model, dan data pengujian yang dipakai untuk mengukur kinerja model. Hal ini bertujuan untuk mendapatkan nilai akurasi dan menilai kemampuan model dalam memprediksi data baru.

Tabel 2 Data Training

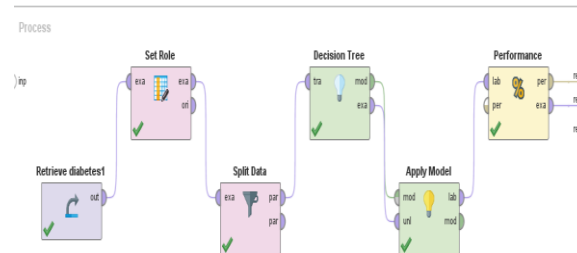
NO	STATUS DIABETES	NAMA	JK	USIA	TB	BB	AKTIFITAS FISIK	KETURUNAN DIABETES
1	Tidak	MFK	P	21	165	46	Tidak	Tidak
2	Tidak	KH	P	24	170	60	Tidak	Tidak
3	Tidak	HAD	P	22	155	54	Ya	Ya
4	Tidak	D	P	21	160	58	Tidak	Tidak
-----	-----	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----
-----	-----	-----	-----	-----	-----	-----	-----	-----
229	Tidak	VAY	P	25	150	67	Sedang	Tidak
230	Tidak	ST	P	25	158	58	Ringan	Tidak
231	Tidak	HK	P	31	146	52	Sedang	Tidak
232	Tidak	R	P	52	154	55	Sedang	Ya

Tabel 3 Data Testing

NO	STATUS DIABETES	NAMA	JK	USIA	TB	BB	AKTIVITAS FISIK	KETURUNAN DIABETES
1	Tidak	MFK	P	21	165	48	Sedang	Tidak
2	Tidak	KH	P	24	170	60	Sedang	Tidak
3	Tidak	HAD	P	22	155	54	Sedang	Ya
4	Tidak	D	P	21	160	58	Sedang	Tidak
-----	-----	-----	-----	-----	-----	-----	-----	-----
97	Ya	IN	P	67	150	46	Ringan	Ya
98	Tidak	M	P	23	160	48	Berat	Tidak
99	Ya	ISU	P	53	152	48	Sedang	Tidak

3.3 Proses Klasifikasi

Proses klasifikasi pada penelitian ini dilakukan dengan menggunakan metode Decision Tree melalui penerapan algoritma C4.5 menggunakan aplikasi RapidMiner diawali dengan mengimpor dataset ke aplikasi, kemudian dilakukan pra-pemrosesan data seperti pemilihan atribut, penanganan data kosong. Model klasifikasi dibangun dengan metode Decision Tree dan C4.5, dimana Decision Tree memilih atribut berdasarkan information gain, sedangkan C4.5 menggunakan gain ratio dan melakukan pruning untuk mengurangi overfitting. Setelah model terbentuk, tahap pengujian dilakukan dengan apply model dan hasil dievaluasi menggunakan performance (*classification*) untuk mendapatkan nilai akurasi, presisi, recall, serta confusion matrix.[10]



Gambar 2 Pengolahan data dengan algoritma Decision Tree C4.5

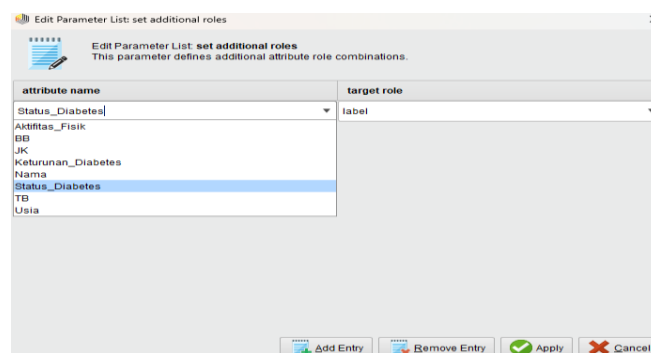
Gambar 2 merupakan alur proses pada aplikasi Rapidminer, prosesnya adalah sebagai berikut : Retrieve data, Split Role, Split Data, Apply Model, Performance, hasil dari proses tersebut akan dijelaskan sebagai berikut.

1. Retrieve Data

Retrieve data adalah proses pengambilan data dan memilih data dari berbagai sumber lalu diinput di aplikasi Rapidminer untuk analisis dan proses klasifikasi dilakukan menggunakan algoritma Decision Tree dan C4.5 agar dapat menghasilkan klasifikasi yang optimal dan akurat.

2. Set Role

Select Atribut merupakan tahap yang dilakukan untuk memilih atribut atau variable yang akan diolah, dimana pada penelitian ini digunakan metode decision tree dan C4.5, dengan dataset yang umumnya terdiri dari 8 atribut sebagai dasar untuk proses analisis data.

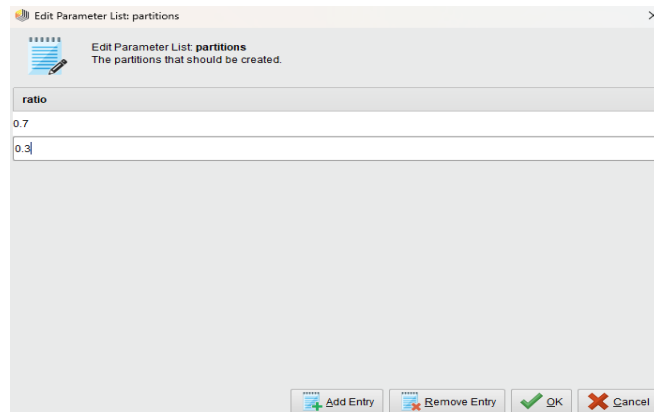


Gambar 3 Tahap Pemilihan Atribut

3. Split data

Split data adalah proses membagi dataset menjadi dua atau lebih bagian dengan tujuan memisahkan data untuk melatih (training) model dan data yang digunakan untuk menguji (testing) model. Teknik ini penting dalam machine learning agar model yang dibangun tidak hanya menghaafal data tetapi juga dapat memprediksi dengan baik. Pada aplikasi rapidminer,

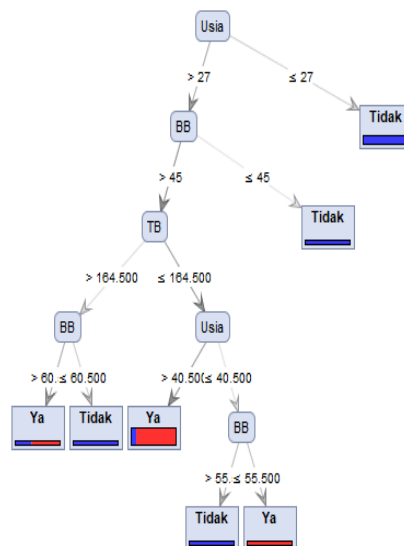
yang digunakan saat menjalankan operator split data. Dalam pengaturan tersebut, terdapat dua rasio yaitu 0.7 dan 0.3.



Gambar 4. Rasio Operator Split

4. Penerapan Decesion tree C4.5

Pada tahap ini pemrosesan data dengan algoritma decision tree C4.5



Gambar 5 Penerapan Decision Tree dan C4.5

Proses klasifikasi menggunakan metode decision tree

1. Root node (Akar pohon): usia

- A. Jika usia $\leq 27 \rightarrow$ langsung di klasifikasikan sebagai **Tidak** diabetas.
- B. Jika usia $\geq 27 \rightarrow$ evaluasi dilanjutkan ke atribut **BB**.

2. Node BB (Jika usia > 27)

- A. Jika BB $\leq 45 \rightarrow$ langsung diklasifikasi sebagai **Tidak** diabetes.
- B. Jika BB $> 45 \rightarrow$ lanjut ke evaluasi **TB** (Tinggi Badan)

3. Node TB (Jika BB > 45)

- A. Jika TB $> 184.5 \rightarrow$ lanjut evaluasi BB.
 - a) Jika BB $\geq 80 \rightarrow$ klasifikasi: **Ya**
 - b) Jika BB $\leq 80.5 \rightarrow$ klasifikasi: **Tidak**

B. Jika $TB \leq 184$ → lanjut ke evaluasi Usia

- a) Jika usia > 40.5 → klasifikasi: **Ya**
- b) Jika usia ≤ 40.5 → lanjut evaluasi: **BB**
- c) Jika $BB \leq 55.5$ → klasifikasi: **Tidak**
- d) Jika $BB > 55.5$ → klasifikasi: **Ya**

5. Apply Model

Apply model adalah proses dalam data mining di mana model yang telah dilatih (*trained model*) pada data latih (*training data*) digunakan untuk memprediksi label atau kelas pada data baru (testing data). Dalam hal ini, model berfungsi untuk memperkirakan status diabetes berdasarkan atribut-atribut seperti usia, jenis kelamin, tinggi badan, berat badan, tinggi badan, aktivitas fisik, dan keturunan diabetes.

Row No.	Status_Diab...	prediction(S...	confidence[...	confidence[...	Nama	JK	Usia	TB	BB	Aktivitas_Fisik	Ketur...
1	Tidak	Tidak	1	0	Maulida filat...	P	21	165	46	Sedang	Tidak
2	Tidak	Tidak	1	0	Kristin handa...	P	24	170	60	Sedang	Tidak
3	Tidak	Tidak	1	0	Hellena Aura ...	P	22	155	54	Sedang	Ya
4	Tidak	Tidak	1	0	dila	P	21	160	58	Sedang	Tidak
5	Tidak	Tidak	1	0	Septiafuti nu...	P	21	158	47	Sedang	Ya
6	Tidak	Tidak	1	0	Nur sholeihah	P	46	150	50	Sedang	Tidak
7	Tidak	Tidak	1	0	Agesti Multi	P	21	155	45	Berat	Tidak
8	Tidak	Tidak	1	0	Saadatun nisa	P	21	155	56	Sedang	Tidak
9	Tidak	Tidak	1	0	Herinda Hu...	P	22	160	58	Ringan	Tidak
10	Ya	Ya	0	1	Maryem	P	62	160	52	Ringan	Tidak
11	Ya	Ya	0	1	Maryem	P	62	160	52	Ringan	Tidak
12	Tidak	Tidak	0.750	0.250	Amri	L	65	165	60	Sedang	Tidak
13	Tidak	Tidak	1	0	Dinda Febily...	P	22	160	45	Berat	Tidak
14	Tidak	Tidak	1	0	Avril Salmani...	P	20	160	43	Sedang	Tidak
15	Tidak	Tidak	1	0	Parjiem	P	42	160	44	Sedang	Tidak
16	Tidak	Tidak	0.750	0.250	Suwami	P	61	166	57	Ringan	Tidak

Gambar 6 Hasil Apply Model

6. Performance

1. Hasil akurasi menggunakan confusion matrix

Pada tahap ini akan dihitung dari algoritma decision tree C4.5 menggunakan confusion matrix

accuracy: 96.49%

	true Tidak	true Ya	class precision
pred. Tidak	34	2	94.44%
pred. Ya	0	21	100.00%
class recall	100.00%	91.30%	

Gambar 7 Hasil Akurasi Menggunakan Confusion Matrix

Berdasarkan gambar 7, dapat disimpulkan bahwa penerapan algoritma decision tree C4.5 untuk mengklasifikasi penyakit diabetes menghasilkan akurasi sebesar 96.49%. Hasil ini menunjukkan bahwa algoritma Decision Tree C4.5 mampu melakukan klasifikasi penyakit diabetes dengan tingkat keakuratan yang tinggi.

2. Hasil decision tree

Tree

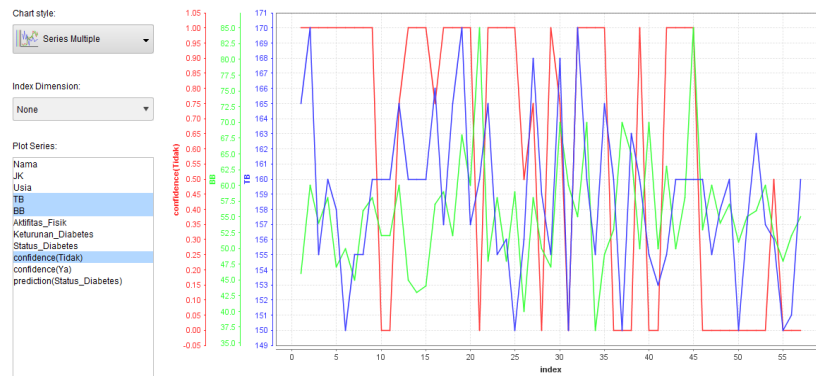
```

Usia > 42.500
|   TB > 164: Tidak {Tidak=3, Ya=1}
|   TB ≤ 164
|   |   TB > 156.500: Ya {Tidak=0, Ya=13}
|   |   TB ≤ 156.500
|   |   |   Keturunan_Diabetes = Tidak
|   |   |   |   Usia > 48.500
|   |   |   |   |   TB > 155.500: Tidak {Tidak=1, Ya=1}
|   |   |   |   |   TB ≤ 155.500: Ya {Tidak=0, Ya=6}
|   |   |   |   |   |   Usia ≤ 48.500: Tidak {Tidak=2, Ya=0}
|   |   |   |   |   |   |   Keturunan_Diabetes = Ya: Ya {Tidak=0, Ya=2}
Usia ≤ 42.500: Tidak {Tidak=28, Ya=0}

```

Gambar 8 Hasil Decision Tree

3. Hasil Grafik



Gambar 9 Hasil Grafik

7. Evaluasi

Setelah proses klasifikasi menggunakan metode decision tree dan algoritma C4.5 pada dataset, di peroleh hasil evaluasi model dengan tingkat akurasi sebesar 96,49%.

accuracy: 96.49%

	true Tidak	true Ya	class precision
pred. Tidak	34	2	94.44%
pred. Ya	0	21	100.00%
class recall	100.00%	91.30%	

Gambar 10 Evaluasi Confusion matrix

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan untuk klasifikasi penyakit diabetes melitus dengan algoritma Decesion Tree C4.5 pada data responden di kabupaten Grobogan, Kesimpulan dari penelitian ini menunjukkan bahwa metode yang digunakan mampu menjelaskan faktor-faktor yang mempengaruhi terhadap status diabetes melitus :

1. Dataset yang digunakan terdiri dari 331 data responden dengan atribut seperti jenis kelamin, usia, berat badan, aktivitas fisik, tinggi badan, dan keturunan diabetes.
2. Pengolahan data dilakukan menggunakan aplikasi Rapidminer dengan tahapan mulai Retrieve data, Set role, Split data, Apply model, hingga Performance.

3. Hasil evaluasi menunjukkan bahwa algoritma decision tree C4.5 menghasilkan akurasi sebesar 96,49%, yang menandakan bahwa model memiliki kinerja yang baik dalam mengklasifikasi status diabetes responden.

5. SARAN

Berdasarkan kesimpulan yang diperoleh terdapat beberapa saran untuk penelitian selanjutnya antarlain :

1. Disarankan agar penelitian mendatang mempeerluas kajian dengan menambah lebih banyak sumber maupun referensi yang relevan guna memperkuat landasan teori dan hasil analisis.
2. Peneliti berikutnya diharapkan dapat melakukan perbandingan hasil klasifikasi dengan algoritma lain sehingga dapat diketahui model yang paling optimal dalam mengklasifikasikan penyakit diabetes.
3. Pemerintah daerah dan dinas Kesehatan dapat menggunakan hasil klasifikasi ini untuk edukasi masyarakat tentang pentingnya menjaga gaya hidup sehat.

DAFTAR PUSTAKA

- [1] C. A. Shawputri, A. R. Lutfi, N. A. Fauziyyah, W. N. Ramadani, and D. S. S. Rejeki, "Literature Review : Faktor Risiko Diabetes Melitus Tipe Ii Di Dunia," *J. Kesehat. Masy.*, vol. 12, no. 4, pp. 247–259, 2024, [Online]. Available: <http://ejournal3.undip.ac.id/index.php/jkm>
- [2] M. Safitri and A. D. Praba, "Prediksi Penyakit Diabetes Dengan Menggunakan Algoritma C4.5," *JIKA (Jurnal Inform.)*, vol. 8, no. 1, p. 74, 2024, doi: 10.31000/jika.v8i1.9998.
- [3] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [4] Mustaqim, "Metode Penelitian Gabungan Kuantitatif Kualitatif/Mixed Methods Suatu Pendekatan Alternatif," *J. Intelegensia*, vol. 04, no. 1, pp. 1–9, 2016, [Online]. Available: <https://ejournal.unisnu.ac.id/JI/article/view/1351>
- [5] I. Widyastuti and A. C. Wijayanti, "Hubungan Pengetahuan dan Sikap dengan Kualitas Hidup Penderita Diabetes Melitus Tipe 2 di Surakarta," *J. Kesehat. Masy. Indones.*, vol. 16, no. 3, p. 136, 2021, doi: 10.26714/jkmi.16.3.2021.136-147.
- [6] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," 2020-10-07.
- [7] N. Nurussakinah and M. Faisal, "Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree," *J. Inform.*, vol. 10, no. 2, pp. 143–149, Oct. 2023, doi: 10.31294/inf.v10i2.15989.
- [8] R. Putri Fadhillah *et al.*, "KLASIFIKASI PENYAKIT DIABETES MELLITUS BERDASARKAN FAKTOR-FAKTOR PENYEBAB DIABETES MENGGUNAKAN ALGORITMA C4.5." [Online]. Available: www.kaggle.com
- [9] A. F. Fadhlullah and T. Widiyaningtyas, "Comparative Analysis of Decision Tree and Random Forest Algorithms for Diabetes Prediction," *JTAM (Jurnal Teor. dan Apl. Mat.)*, vol. 8, no. 4, p. 1121, 2024, doi: 10.31764/jtam.v8i4.24388.
- [10] N. N. U. R. Habibah, F. Sains, D. A. N. Teknologi, U. Islam, N. Sultan, and S. Kasim, "Penyakit Diabetes Dengan Faktor Penyebab Menggunakan Decision Tree C4 . 5 Tugas Akhir," 2023.