


Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
<https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/index>
 Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
 Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
 e-ISSN: 2685-0893

Deteksi Dini Diabetes Mellitus Tipe 2 Pada Usia Dewasa Muda Menggunakan Algoritma *Decision Tree* - C4.5

DOI: <http://dx.doi.org/10.35889/jutisi.v15i2.3578>

Creative Commons License 4.0 (CC BY – NC) 

Sofia Darfinda Dalimajun¹, Azwar Riza Habibi^{2*}

Teknik Informatika, Institut Teknologi dan Bisnis Asia Malang, Malang, Indonesia

*e-mail *Corresponding Author*: riza.bj@gmail.com

Abstract

The shift in the epidemiological pattern of metabolic diseases is indicated by the increasing incidence of Type 2 Diabetes Mellitus (T2DM) in young individuals. The increasing incidence of glucose metabolism disorders is largely due to changes in modern lifestyles, such as high calorie intake and lack of physical activity. Early detection during productive age is crucial to prevent long-term problems. Based on clinical data, this study created a model for early detection of type 2 diabetes using the Decision Tree C4.5 algorithm. The dataset was filtered for people aged between 20 and 40 years after being obtained from Kaggle. The study phase included data pre-processing, data splitting for testing and training, and model development using entropy and information gain. Accuracy, precision, recall, F1-score, and ROC-AUC were used to evaluate the model. The accuracy was 92.4% and the ROC-AUC was 0.938. The completed model can be used as a data-driven interpretive health screening tool.

Keywords: *Early diagnosis; Decision Tree algorithm; C4.5; Classification model*

Abstrak

Pergeseran pola epidemiologi penyakit metabolik ditunjukkan oleh meningkatnya kejadian Diabetes Melitus Tipe 2 (T2DM) pada individu muda. Meningkatnya kejadian gangguan metabolisme glukosa sebagian besar disebabkan oleh perubahan gaya hidup modern, seperti pola asupan kalori tinggi dan kurangnya aktivitas fisik. Deteksi dini pada usia produktif sangat penting untuk mencegah masalah jangka panjang. Berdasarkan data klinis, penelitian ini menciptakan model deteksi dini diabetes tipe 2 menggunakan algoritma *Decision Tree* C4.5. Dataset difilter untuk orang berusia antara 20 dan 40 tahun setelah diperoleh dari Kaggle. Fase studi meliputi pra-pemrosesan data, pemisahan data untuk pengujian dan pelatihan, serta pengembangan model menggunakan entropi dan perolehan informasi. Akurasi, presisi, recall, F1-score, dan ROC-AUC digunakan untuk menilai model tersebut. Akurasinya adalah 92,4% dan ROC-AUC adalah 0,938. Model yang telah selesai dapat digunakan sebagai alat skrining kesehatan interpretatif berbasis data.

Kata kunci: *Diabetes Mellitus Tipe 2; Deteksi dini; Decision Tree; C4.5; klasifikasi*

1. Pendahuluan

Diabetes melitus tipe 2 (T2DM) adalah salah satu penyakit metabolik kronis yang saat ini menjadi perhatian utama kesehatan masyarakat global. Gangguan ini ditandai dengan ketidakmampuan tubuh untuk mengontrol kadar glukosa darah dan disebabkan oleh resistensi insulin serta penurunan fungsi sel beta pankreas. Dalam beberapa dekade terakhir, prevalensi DMT2 menunjukkan peningkatan yang signifikan seiring dengan perubahan gaya hidup masyarakat modern, seperti meningkatnya konsumsi makanan berkalori tinggi, pola makan tidak seimbang, serta rendahnya aktivitas fisik sehari-hari [1], [2]. Dampaknya, DMT2 tidak hanya menjadi masalah individu, tetapi juga berkembang menjadi beban kesehatan masyarakat yang memerlukan penanganan serius. Untuk menurunkan risiko konsekuensi jangka panjang, inisiatif pencegahan dan deteksi dini sangat penting.

Perkembangan terbaru menunjukkan adanya peningkatan kasus Diabetes Mellitus Tipe 2 pada kelompok usia dewasa muda, yang sebelumnya relatif jarang ditemukan pada kelompok

usia tersebut. Individu dengan rentang usia 20–40 tahun kini mulai menunjukkan kecenderungan peningkatan risiko terhadap penyakit metabolik ini [3], [4]. Berbagai faktor berkontribusi terhadap kondisi tersebut, antara lain kurangnya aktivitas fisik, pola konsumsi makanan tidak sehat, serta minimnya kesadaran untuk melakukan pemeriksaan kesehatan secara rutin. Akibatnya, diagnosis sering kali dilakukan pada tahap lanjut ketika kondisi penyakit sudah memburuk. Selain berdampak pada kesehatan individu, tingginya angka kejadian diabetes pada usia produktif juga menimbulkan konsekuensi sosial dan ekonomi, seperti meningkatnya biaya pengobatan serta menurunnya produktivitas kerja. Oleh sebab itu, diperlukan metode deteksi dini yang cepat dan akurat untuk mendukung proses skrining pada kelompok usia ini.

Sejumlah penelitian terdahulu telah mengembangkan metode untuk mendeteksi dan memprediksi risiko diabetes dengan memanfaatkan teknik komputasi berbasis data. Penelitian oleh [5] mengkaji penggunaan algoritma *Decision Tree* C4.5 dalam proses klasifikasi penyakit diabetes dengan memanfaatkan data kesehatan pasien, yang menghasilkan model klasifikasi yang relatif mudah dipahami melalui aturan keputusan yang terbentuk. Penelitian lain oleh [6], [7] menerapkan metode *machine learning* untuk memprediksi kemungkinan terjadinya diabetes berdasarkan beberapa parameter medis pasien melalui proses pelatihan dan pengujian model menggunakan dataset klinis. Selain itu, penelitian oleh [8], [9] juga menunjukkan bahwa algoritma klasifikasi berbasis pohon keputusan dapat digunakan untuk membantu proses identifikasi risiko penyakit melalui pembentukan struktur keputusan yang sistematis. Penelitian lain yang dilakukan oleh [10] menunjukkan bahwa algoritma *Random Forest* dan *Support Vector Machine* mampu memberikan performa yang tinggi dalam klasifikasi diabetes dengan memanfaatkan kombinasi atribut klinis dan gaya hidup pasien. Sementara itu, studi oleh [11] mengungkapkan bahwa pemanfaatan teknik *data mining* pada dataset kesehatan berskala besar dapat meningkatkan akurasi dalam prediksi penyakit kronis, termasuk diabetes, terutama ketika dilakukan proses seleksi fitur yang optimal. Meskipun berbagai penelitian tersebut telah memberikan kontribusi dalam pengembangan sistem prediksi diabetes, sebagian besar penelitian masih menggunakan data populasi umum tanpa melakukan segmentasi khusus berdasarkan kelompok usia. Namun, individu muda mungkin memiliki pola metabolisme dan faktor risiko yang berbeda dibandingkan kelompok usia yang lebih tua. Hal ini menunjukkan adanya kekurangan penelitian dalam pembuatan model deteksi dini Diabetes Mellitus Tipe 2 yang secara khusus ditargetkan pada individu muda.

Berdasarkan permasalahan tersebut, penelitian ini menyarankan penggunaan algoritma *Decision Tree* C4.5 untuk membuat model deteksi dini Diabetes Mellitus Tipe 2 pada dewasa muda (usia 20 hingga 40 tahun). Metode ini dipilih karena dapat menciptakan struktur pohon keputusan yang mudah dipahami, sehingga pengguna dan tenaga medis dapat memahami hasil kategorisasi [1], [12] [13]. Lebih lanjut, teknik *Decision Tree* C4.5 dapat memilih karakteristik yang paling relevan untuk membuat aturan keputusan untuk proses kategorisasi. Agar model akhir dapat menggambarkan fitur risiko penyakit dengan lebih tepat, penelitian ini sangat menekankan segmentasi dewasa muda sebagai fokus utama analisis. Untuk mendukung proses skrining kesehatan pada kelompok usia produktif, diharapkan penelitian ini dapat membantu pengembangan sistem deteksi dini berbasis data yang lebih efisien.

2. Metodologi

Penelitian ini dilakukan dalam beberapa fase berikut:

- 1) Pengumpulan Data
Dalam penelitian ini digunakan dataset kesehatan dari platform Kaggle. Untuk tujuan mengklasifikasikan Diabetes Mellitus Tipe 2, dataset tersebut terdiri dari 100.000 titik data dengan sembilan variabel kesehatan.
- 2) Seleksi Data Usia 20–40 Tahun
Dataset disaring berdasarkan kelompok usia dewasa muda, yaitu 20–40 tahun. Setelah proses seleksi, diperoleh sebanyak 35.794 data dengan proporsi penderita diabetes sebesar 2%.
- 3) Preprocessing Data
Sebelum proses klasifikasi, langkah persiapan meliputi pembersihan data dari nilai yang tidak valid atau kosong, pengkodean variabel kategorikal, dan normalisasi fitur numerik untuk meningkatkan kualitas data.
- 4) Pembagian Data

Data pelatihan dan pengujian dipisahkan dari dataset yang telah diproses sebelumnya dengan rasio 80:20. Sebanyak 7.159 titik data digunakan untuk pengujian dan 28.635 titik data untuk pelatihan.

5) Penerapan Algoritma *Decision Tree* C4.5

Model klasifikasi Diabetes Melitus Tipe 2 berdasarkan perhitungan entropi dan perolehan informasi dibangun menggunakan metode *Decision Tree* C4.5 dan data pelatihan.

a. Entropy

Tingkat ketidakpastian suatu dataset dapat ditentukan menggunakan entropi.

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

Keterangan:

S : himpunan data

p_i : proporsi data pada kelas ke- i

n : jumlah kelas

Nilai entropi berkisar antara 0 dan 1. Skor entropi yang lebih rendah menunjukkan homogenitas data yang lebih besar [14].

b. Information Gain

Jumlah *entropi* yang berkurang setelah data dipisahkan menurut suatu atribut diukur menggunakan perolehan informasi.

$$Gain(S,A) = Entropy(S) - \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

Keterangan:

A : atribut yang diuji

S_i : subset data ke- i

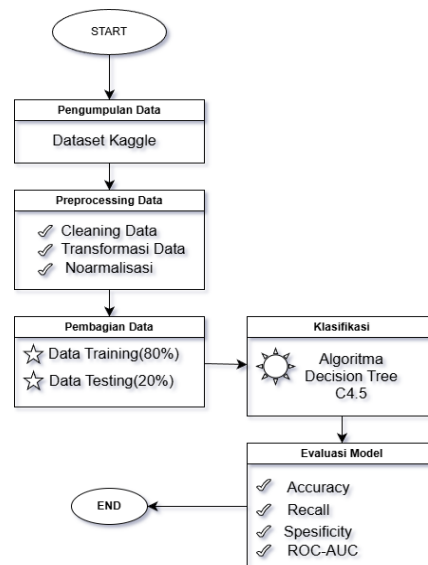
$|S|$: jumlah data pada dataset awal

$|S_i|$: jumlah data pada subset ke- i

Atribut dengan nilai Gain terbesar memiliki kemampuan terbaik dalam memisahkan data.

6) Evaluasi Model

Efektivitas model dalam mengidentifikasi Diabetes Melitus Tipe 2 pada kelompok usia dewasa muda dievaluasi menggunakan data uji dengan menggunakan metrik akurasi, presisi, recall, F1-score, matriks kebingungan, dan ROC-AUC. [15].



Gambar 1. Flowchart Metodologi Penelitian Deteksi Dini Diabetes Mellitus Tipe 2

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari platform Kaggle, yang terdiri dari 100.000 catatan pasien dengan sembilan fitur yang terkait dengan faktor risiko diabetes melitus.

Rentang usia 20 hingga 40 tahun digunakan untuk memilih data yang sesuai dengan tujuan penelitian. Sebanyak 35.794 titik data, atau sekitar 35,8% dari dataset asli, digunakan dalam penelitian ini, berdasarkan hasil seleksi. Dataset ini memiliki fitur ketidakseimbangan kelas yang cukup tinggi, dengan 2,0% data diklasifikasikan sebagai penderita diabetes dan sisanya 98,0% sebagai bukan penderita diabetes, menurut distribusi kelas data. Beberapa data yang dipilih dari sampel penelitian yang terdiri dari orang berusia 20 hingga 40 tahun ditunjukkan pada **Tabel 1**. Usia, jenis kelamin, riwayat medis, BMI, kadar HbA1c, kadar glukosa darah, dan label kategorisasi diabetes termasuk di antara parameter yang disertakan dalam data.

Tabel 1. Contoh Struktur Data Usia 20–40 Tahun

No.	Gender	age	Hyper tension	Heart disease	smoking history	bmi	HbA1c level	Blood Glucose Level (mm/dL)	diabetes
1.	1	28	0	0	0	27.32	5.7	158	0
2.	0	34	0	0	1	25.91	9.0	160	1
3.	0	21	0	0	1	26.01	5.8	140	0
4.	0	36	0	0	2	32.27	6.2	220	1
.....
35.794	1	29	0	0	2	25.41	6.1	130	1

Tabel 1 menyajikan sebagian data hasil seleksi berdasarkan rentang usia 20–40 tahun yang digunakan dalam penelitian ini. Sejumlah karakteristik penting disertakan dalam data, termasuk usia, jenis kelamin, riwayat penyakit jantung, hipertensi, merokok, indeks massa tubuh (BMI), kadar HbA1c, kadar glukosa darah, dan label klasifikasi diabetes. Tabel ini ditampilkan sebagai representasi struktur dataset yang digunakan dalam proses analisis.

3.2 Preprocessing Data

1) Cleaning Data

Untuk memastikan bahwa data yang digunakan tidak mengandung nilai yang tidak akurat atau tidak konsisten, tahap pembersihan melibatkan penyaringan data sesuai dengan kriteria usia penelitian. Sebelum digunakan dalam langkah pemodelan, prosedur ini bertujuan untuk meningkatkan kualitas data.

2) Transformasi Data

Untuk menangani atribut kategorikal yang digunakan dalam proses klasifikasi, transformasi data dilakukan dengan mengubah tipe data. Agar lebih mudah menafsirkan hasil analisis, ciri-ciri BMI kemudian dibagi menjadi empat kategori: kekurangan berat badan, normal, kelebihan berat badan, dan obesitas.

3) Normalisasi Data

Karakteristik numerik seperti kadar glukosa darah, BMI, dan HbA1c dinormalisasi. Untuk menghindari beberapa atribut mendominasi proses pembelajaran model, prosedur ini berupaya menyeimbangkan skala antar variabel.

3.3 Pembagian Data

Selanjutnya, metode hold-out 80:20 digunakan untuk memisahkan dataset yang telah diproses menjadi data pelatihan dan pengujian. Sebanyak 7.159 titik data digunakan untuk pengujian, sedangkan 28.635 titik data digunakan untuk pelatihan. Kelas non-diabetes, dengan persentase 98,0%, dan kelas diabetes, dengan persentase 2,0%, terus mendominasi distribusi kelas di kedua subset data.

3.4 Klasifikasi Menggunakan Decision Tree C4.5

Proses klasifikasi dilakukan dengan menggunakan algoritma *Decision Tree* C4.5 dan data pelatihan. Untuk membuat struktur pohon keputusan, metode ini memilih fitur optimal berdasarkan nilai perolehan informasi. Kemampuan model untuk secara sistematis mengkarakterisasi hubungan antara indikator kesehatan dan risiko diabetes melitus ditunjukkan oleh hasil pelatihan.

3.5 Evaluasi Model

3.5.1 Confusion Matrix Hasil Klasifikasi

Berdasarkan hasil pengujian model menggunakan matriks kebingungan, nilai akurasi sebesar 92,4% menunjukkan bahwa algoritma *Decision Tree C4.5* dapat melakukan klasifikasi dengan tingkat akurasi yang tinggi.

3.5.2 Classification Report

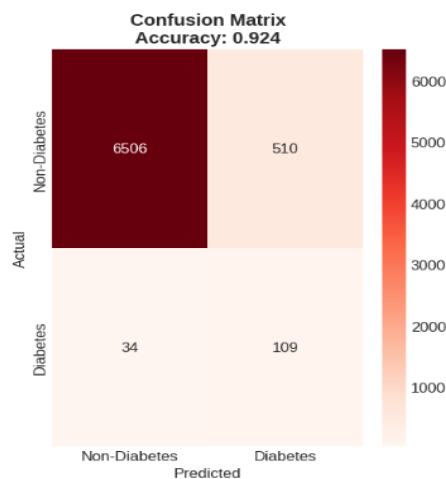
Evaluasi kinerja model *Decision Tree C4.5* menggunakan data uji dengan metrik presisi, recall, F1-score, dan akurasi ditampilkan pada **Tabel 2** di bawah ini.

Tabel 2. Hasil Classification Decision Tree C4.5

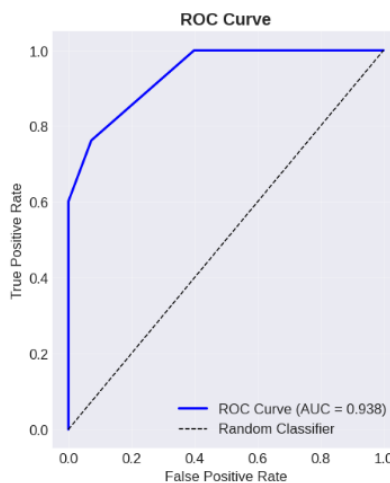
	precision	recall	F1-score	support
Non-Diabetes	0.99	0.93	0.96	7016
Diabetes	0.18	0.76	0.29	143
Accuracy			0.92	7159
Macro avg	0.59	0.84	0.62	7159
Weighted avg	0.98	0.92	0.95	7159

3.5.3 Hasil ROC Model Decision Tree C4.5

Selanjutnya, analisis kurva ROC menghasilkan nilai Area Under Curve (AUC) sebesar 0,938, yang menunjukkan bahwa model tersebut sangat efektif dalam membedakan antara individu yang berisiko terkena diabetes dan individu yang tidak berisiko.



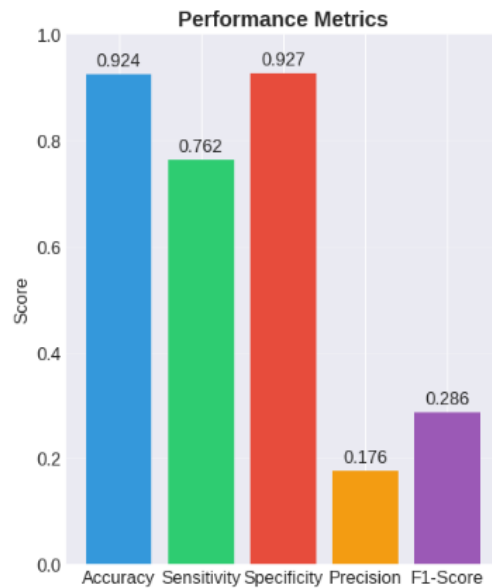
Gambar 2. Confusion Matrix Hasil Klasifikasi



Gambar 3. Kurva ROC Model Decision Tree C4.5

3.5.4 Hasil Evaluasi Performa Model

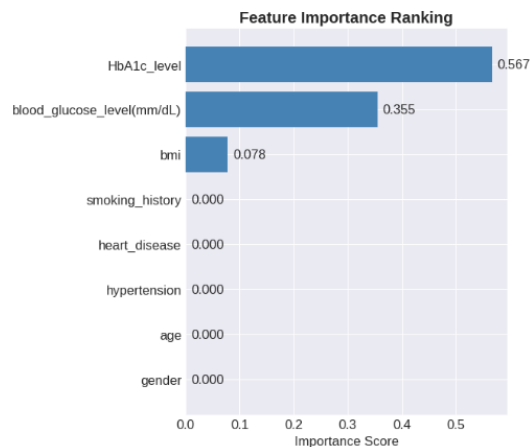
Selain itu, hasil evaluasi metrik klasifikasi menunjukkan nilai sensitivity sebesar 0.762, specificity sebesar 0.927, serta nilai F1-score yang menggambarkan keseimbangan antara ketepatan dan kemampuan model dalam mendeteksi kasus diabetes.



Gambar 4. Hasil Evaluasi Performa Model

3.5.5 Analisis Feature Importance

Berdasarkan analisis feature importance yang ditampilkan pada Gambar 5, atribut HbA1c_level memiliki nilai kepentingan tertinggi dengan skor sebesar 0,567, sehingga menjadi faktor paling dominan dalam proses klasifikasi. Atribut blood_glucose_level menempati peringkat kedua dengan nilai 0,355, diikuti oleh BMI dengan skor 0,078. Sementara itu, atribut lain seperti riwayat merokok, penyakit jantung, hipertensi, usia, dan jenis kelamin memiliki nilai kepentingan yang sangat rendah atau mendekati nol.

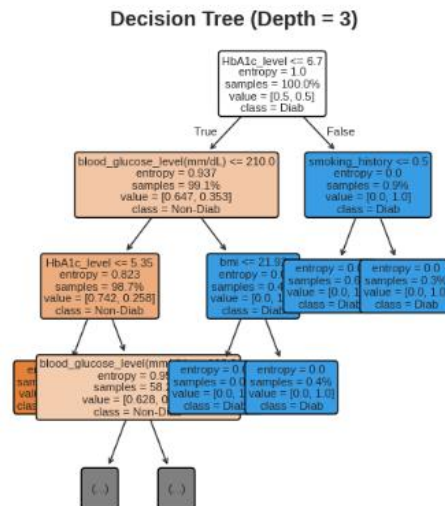


Gambar 5. Hasil Ranking Feature Importance

3.5.6 Pohon Keputusan

Berdasarkan visualisasi pohon keputusan pada Gambar 6, atribut HbA1c_level muncul sebagai simpul akar (root node), yang menandakan bahwa atribut tersebut memiliki peran paling dominan dalam proses klasifikasi Diabetes Mellitus Tipe 2. Percabangan selanjutnya menunjukkan bahwa kadar glukosa darah dan BMI digunakan sebagai atribut lanjutan untuk membedakan kelas diabetes dan non-diabetes pada kondisi tertentu. Kemampuan model untuk membuat aturan keputusan yang tepat dengan sejumlah kecil atribut ditunjukkan oleh struktur pohonnya yang relatif sederhana. Hal ini menunjukkan kapasitas algoritma *Decision Tree* C4.5

untuk menyediakan model yang mudah dipahami dan dapat berfungsi sebagai alat interpretasi untuk deteksi dini. Struktur pohon keputusan menunjukkan bahwa pada kelompok usia dewasa muda, klasifikasi diabetes lebih ditentukan oleh indikator metabolik dibandingkan faktor demografis, yang membedakan penelitian ini dari studi sebelumnya yang mencakup seluruh kelompok usia.



Gambar 6. Hasil pohon keputusan

4. Pembahasan

Menurut temuan penelitian, algoritma *Decision Tree* C4.5 dapat memberikan model klasifikasi yang secara akurat memprediksi risiko Diabetes Melitus Tipe 2 pada orang muda. Nilai ROC-AUC sebesar 0,9381 dan nilai akurasi sebesar 92,4% menunjukkan bahwa model tersebut cukup baik dalam membedakan antara mereka yang berisiko terkena diabetes dan mereka yang tidak. Selain itu, nilai spesifisitas model sebesar 92,73% menunjukkan bahwa model tersebut sangat baik dalam mengidentifikasi orang yang tidak menderita diabetes, sedangkan nilai sensitivitasnya sebesar 76,22% menunjukkan bahwa model tersebut cukup berhasil dalam mengenali kasus kondisi tersebut. Meskipun demikian, nilai presisi yang relatif rendah sebesar 17,61% menunjukkan bahwa jumlah positif palsu masih cukup besar. Hal ini dipengaruhi oleh distribusi kelas yang tidak seimbang pada dataset, dengan hanya 2% data yang berkaitan dengan diabetes.

Hasil penelitian ini juga menunjukkan seberapa baik algoritma berbasis *Decision Tree* C4.5 dapat memproses data medis untuk klasifikasi. Model akhir menunjukkan bahwa karakteristik indeks massa tubuh (BMI), kadar glukosa darah, dan HbA1c merupakan penentu utama klasifikasi. Indikator metabolik memainkan peran utama dalam mengidentifikasi risiko Diabetes Melitus Tipe 2, khususnya pada kelompok usia dewasa muda, sesuai dengan nilai penting yang tinggi dari atribut HbA1c dan kadar glukosa darah. Temuan ini konsisten dengan penelitian lain yang menemukan bahwa dengan menggunakan karakteristik klinis pasien, algoritma *Decision Tree* C4.5 dapat menghasilkan model klasifikasi dengan tingkat akurasi tinggi dalam memprediksi diabetes [16], [4], [13], [20]. Selain itu, penelitian lain juga menunjukkan bahwa parameter kesehatan seperti kadar glukosa darah, tekanan darah, serta indikator fisiologis lainnya memiliki kontribusi yang besar dalam meningkatkan performa model klasifikasi untuk diagnosis diabetes [17], [7], [9], [18]. Dengan demikian, temuan dalam penelitian ini memperkuat bahwa penggunaan atribut klinis yang relevan menjadi faktor penting dalam menghasilkan model prediksi yang akurat.

Selain itu, dengan berfokus pada identifikasi dini diabetes pada rentang usia 20-40 tahun, penelitian ini menambahkan sesuatu yang baru. Strategi penelitian ini, berbeda dengan penelitian sebelumnya yang biasanya menggunakan populasi secara keseluruhan, berkonsentrasi pada segmen usia dewasa muda, sehingga model yang dihasilkan dapat menggambarkan pola risiko secara lebih akurat pada kelompok usia produktif. Pendekatan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa segmentasi data berdasarkan karakteristik demografis dapat meningkatkan performa model prediksi, terutama dalam bidang kesehatan [19].

Dengan demikian, penelitian ini tidak hanya menunjukkan efektivitas algoritma *Decision Tree* C4.5 dalam melakukan klasifikasi risiko diabetes, tetapi juga menegaskan pentingnya pemanfaatan atribut klinis serta pendekatan segmentasi data dalam meningkatkan akurasi dan relevansi model prediksi pada kasus kesehatan.

5. Simpulan

Penelitian ini berhasil mengembangkan model deteksi dini Diabetes Melitus Tipe 2 pada orang muda (20–40 tahun) menggunakan dataset kesehatan dari Kaggle dan algoritma *Decision Tree* C4.5. Hasil pengujian menunjukkan bahwa model tersebut memiliki tingkat akurasi 92,4% dan nilai ROC-AUC 0,938, yang menunjukkan kemampuan luar biasa untuk membedakan antara mereka yang berisiko terkena diabetes dan mereka yang tidak. Karena nilai sensitivitasnya yang relatif tinggi dalam mendeteksi diabetes, model ini berpotensi untuk digunakan sebagai alat skrining dini pada kelompok usia produktif yang seringkali kurang informasi tentang bahaya penyakit tersebut, membuktikan betapa bermanfaatnya paradigma ini dalam mengidentifikasi kasus positif.

Investigasi selanjutnya mengungkapkan bahwa variabel terpenting dalam prosedur klasifikasi adalah indeks massa tubuh (BMI), kadar glukosa darah, dan HbA1c. Karena rentang usia penelitian yang sempit, faktor demografis termasuk jenis kelamin dan usia, di sisi lain, kurang berkontribusi. Manfaat algoritma *Decision Tree* C4.5, terutama dalam meningkatkan transparansi pengambilan keputusan, lebih lanjut ditunjukkan oleh struktur pohon keputusan yang dihasilkan, yang relatif mudah dan sederhana untuk dipahami. Hasil ini juga mendukung gagasan bahwa penanda metabolik adalah cara utama untuk mengidentifikasi risiko Diabetes Melitus Tipe 2 pada individu muda. Studi ini dapat diperluas untuk mencakup fluktuasi variabel klinis, menerapkan teknik manajemen ketidakseimbangan data, dan membandingkan model dengan algoritma klasifikasi lainnya untuk meningkatkan akurasi dan kinerjanya.

Daftar Referensi

- [1] A. K. Jaggi, A. Sharma, N. Sharma, R. Singh, and P. S. Chakraborty, "Diabetes Prediction Using Machine Learning," *Lecture Notes in Networks and Systems*, vol. 185 LNNS, no. 09, pp. 383–392, 2021, doi: 10.1007/978-981-33-6081-5_34.
- [2] W. H. O. G. Report, *Global Report on Diabetes*, vol. 978. 2016. [Online]. Available: http://www.who.int/about/licensing/copyright_form/index.html%0Ahttp://www.who.int/about/licensing/copyright_form/index.html%0Ahttps://apps.who.int/iris/handle/10665/204871%0Ahttp://www.who.int/about/licensing/
- [3] IDF, "IDF Diabetes Atlas. In IDF Diabetes Atlas," *International Diabetes Federation*, vol. 11th editi, p. 131, 2025, [Online]. Available: <https://www.idf.org/aboutdiabetes/type-2-diabetes.html>
- [4] F. A. Ahda and M. Zainuddin, "Prediksi Kepuasan Pelayanan Perpustakaan Menggunakan Algoritma Decision Tree (C4.5)," *Jurnal Teknologi Informasi*, vol. 10, pp. 143–150, 2019, doi: 10.36382/jti-tki.v10i2.368.
- [5] A. Iyer, J. S., and R. Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 1, pp. 01–14, 2015, doi: 10.5121/ijdkp.2015.5101.
- [6] Y. Nuryamin and F. Risyda, "Analisis Prediksi Penyakit Diabetes Menggunakan Metode Decision Tree C4.5 Dan Naive Bayes," *Jurnal Sistem Informasi Universitas Suryadarma*, vol. 12, no. 2, pp. 234–242, 2014, doi: 10.35968/jsi.v12i2.1549.
- [7] D. A. Sulisty, D. D. Prasetya, F. A. Ahda, and A. P. Wibawa, "Pivoted Low Resource Multilingual Translation with NER Optimization," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 24, no. 5, 2025, doi: 10.1145/3727876.
- [8] Daw Khin Po, "Simulation of Process Scheduling Algorithms," *International Journal of Trend in Scientific Research and Development*, vol. 3, no. 4, pp. 1629–1632, 2019, doi: <https://doi.org/10.31142/ijtsrd25124>.
- [9] F. A. Ahda, A. P. Wibawa, D. D. Prasetya, D. A. Sulisty, and A. Nafalski, "Minangkabau Language Stemming: A New Approach with Modified Enhanced Confix Stripping," *Jurnal RESTI*, vol. 9, no. 3, pp. 677–687, 2025, doi: 10.29207/resti.v9i3.6511.
- [10] S. Liu, "Diabetes Prediction by KNN, SVM, Random Forest and XGBoost," *Highlights in Science, Engineering and Technology*, vol. 72, pp. 1113–1120, 2023, doi: 10.54097/8h8dff76.

-
- [11] Y. Tian, "Machine Learning Models for Diabetes Prediction: Logistic Regression, SVM, Random Forest, and Neural Networks," *Applied and Computational Engineering*, vol. 211, no. 1, pp. 174–179, 2025, doi: 10.54254/2755-2721/2026.tj30648.
- [12] H. Setiani, M. N. Arridho, and S. Supriyanto, "Early Detection of Type 2 Diabetes Using C4.5 Decision Tree Algorithm on Clinical Health Records," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1663–1669, 2025, doi: 10.30871/jaic.v9i4.10190.
- [13] F. Almu'iini Ahda, A. Prasetya Wibawa, D. Prasetya, and A. Sulisty, "International Journal On Informatics Visualization journal homepage : www.joiv.org/index.php/joiv International Journal On Informatics Visualization Comparison of Adam Optimization and RMSprop in Minangkabau-Indonesian Bidirectional Translation with Neura," vol. 8, no. March, pp. 231–238, 2024, [Online]. Available: www.joiv.org/index.php/joiv
- [14] A. W. Wicaksono and T. Setiadi, "Penerapan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar di Kecamatan Juai," *Format: Jurnal Ilmiah Teknik Informatika*, vol. 12, no. 2, p. 151, 2023, doi: 10.22441/format.2023.v12.i2.008.
- [15] T. H. Sinaga, A. Wanto, I. Gunawan, and Z. Masruro, "Journal of Computer Networks , Architecture and High Performance Computing Implementation of Data Mining Using C4 . 5 Algorithm on Customer Satisfaction in Tirta Lihou PDAM Journal of Computer Networks , Architecture and High Performance Computing," vol. 3, no. 1, pp. 9–20, 2021.
- [16] Fatmawati, "Perbandingan Algoritma Klasifikasi Data Mining Model C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *Jurnal Techno Nusa Mandiri*, vol. XIII, no. 1, p. 50, 2016.
- [17] P. B. Khokhar, V. Pentangelo, F. Palomba, and C. Gravino, "Towards Transparent and Accurate Diabetes Prediction Using Machine Learning and Explainable Artificial Intelligence," 2025, [Online]. Available: <http://arxiv.org/abs/2501.18071>
- [18] H. Afandi, D. A. Sulisty, and S. A. Malang, "Sistem Pakar Untuk Diagnosa Hama dan Penyakit Pada Bunga Krisan Menggunakan Forward Chaining," vol. 13, no. 2, pp. 101–114, 2019.
- [19] H. K. Ashwin, K. K. Poojari, S. R. Rahul, and S. C. I. L., "Design of an automobile instrument cluster using CAN protocol," no. June, pp. 5722–5725, 2020.
- [20] V. Venecia, G. Hoendarto, & T. Darmanto, "Design of Diabetes Prediction Interface Using E-ss and Classification Tree Algorithm. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 14, no. 3, pp. 1856-1867, 2025.