

Analisis Prediksi Diabetes Menggunakan Decision Tree pada Dataset Diabetes Pima Indians

Diabetes Prediction Analysis Using Decision Tree Method on Pima Indians Diabetes Dataset

Muhammad Ali Ridla*, Moh. Noval Prayoga

Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy, Situbondo, Indonesia

E-mail: *\novalprayoga960@gmail.com

Abstract. *Diabetes mellitus is a persistent disease with a continuously growing worldwide incidence, making correct early detection strategies critical. This examine pursuits to expect diabetes danger using the selection Tree algorithm primarily based at the Pima Indians Diabetes Dataset. The studies tiers consist of records preprocessing, coping with invalid values, normalization, information splitting, and model education the usage of gini and entropy splitting criteria. performance development changed into performed via hyperparameter optimization the usage of GridSearchCV and statistics balancing with SMOTE. The effects display that the optimized selection Tree model affords excessive accuracy, strong interpretability, and the capability to become aware of the maximum influential variables associated with diabetes danger. those findings confirm that choice Tree combined with optimization strategies and data balancing is an effective technique for early diabetes detection.*

Keywords: *Decision Tree, Diabetes Prediction, Pima Indians Diabetes Dataset, Classification*

Abstrak. *Diabetes melitus merupakan penyakit persisten dengan insidensi yang terus meningkat di seluruh dunia, sehingga strategi deteksi dini yang tepat menjadi sangat penting. Penelitian ini bertujuan untuk memprediksi risiko diabetes menggunakan algoritma Selection Tree yang berbasis pada Pima Indians Diabetes Dataset. Tahapan penelitian meliputi prapemrosesan data, penanganan nilai yang tidak valid, normalisasi, pemisahan data, dan pelatihan model menggunakan kriteria pemisahan gini dan entropi. Pengembangan kinerja dilakukan melalui optimasi hiperparameter menggunakan GridSearchCV dan penyeimbangan data dengan SMOTE. Hasilnya menunjukkan bahwa model Selection Tree yang dioptimalkan memberikan akurasi tinggi, interpretabilitas yang kuat, dan kemampuan untuk mengidentifikasi variabel yang paling berpengaruh terkait dengan risiko diabetes. Temuan ini menegaskan bahwa Selection Tree yang dikombinasikan dengan strategi optimasi dan penyeimbangan data merupakan teknik yang efektif untuk deteksi dini diabetes.*

Kata kunci: *Decision Tree, Prediksi Diabetes, Pima Indians Diabetes Dataset, Klasifikasi*

Submitted: 08-12-2026 | *Accepted:* 10-03-2026 | *Published:* 31-03-2026

How to Cite:

M. A. Ridla and M. N. Prayoga, " Analisis Prediksi Diabetes Menggunakan Decision Tree pada Dataset Diabetes Pima Indians" *Journal of Information System and Application Development (JISAD)*, vol. 4, no. 1, pp. 146-153, 2026, doi: 10.26905/jisad.v4i1.16494



PENDAHULUAN

Diabetes mellitus adalah penyakit metabolik kronis yang menjadi masalah kesehatan global karena *prevalensinya* yang terus meningkat dan komplikasi serius yang dapat ditimbulkan. Menurut data Internasional, deteksi dini diabetes sangat penting untuk mencegah beban klinis dan sosial ekonomi yang signifikan. Salah satu pendekatan terkini yang menjanjikan dalam deteksi dini adalah penggunaan *machine learning* untuk membangun model prediksi berdasarkan data klinis pasien. Salah satu dataset yang paling sering digunakan dalam penelitian prediksi diabetes adalah *Pima Indians Diabetes Dataset* (PIDD) dari *UCI Machine Learning Repository*. Dataset ini berisi 768 sampel dengan delapan atribut klinis (seperti glukosa plasma, tekanan darah, BMI, insulin, ketebalan kulit, usia, dan lain-lain) serta satu label kelas (diabetes / non-diabetes). Karena kemudahan akses dan struktur data yang representatif, PIDD telah menjadi benchmark dalam banyak penelitian klasifikasi diabetes.

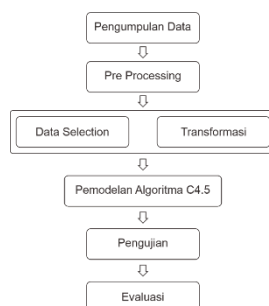
Dalam konteks algoritma klasifikasi, *Decision Tree* merupakan salah satu metode favorit karena kemampuannya menghasilkan aturan keputusan yang mudah diinterpretasikan, yang sangat penting di ranah medis [1]. Penelitian lain menerapkan algoritma C4.5 pada PIDD melalui aplikasi *Orange* dan menemukan bahwa faktor-faktor seperti glukosa, BMI, dan usia sangat berpengaruh dalam risiko diabetes, dengan akurasi 71,9% dan AUC 0,674 [2]. Sementara itu, penelitian Optimasi Algoritma *Decision Tree* dalam memprediksi penyakit diabetes menggunakan *GridSearchCV* menunjukkan bahwa setelah *hyperparameter tuning* (*max_depth*, *min_samples_leaf*, *criterion*) dan penyeimbangan data menggunakan SMOTE, akurasi model *Decision Tree* dapat meningkat hingga 97,06% [3].

Meskipun *Decision Tree* telah banyak digunakan, sejumlah gap penelitian masih terbuka. Pertama, sebagian besar studi hanya menggunakan parameter *default* atau sedikit *tuning*, tanpa eksplorasi yang mendalam terhadap *hyperparameter* dan efeknya terhadap performa model. Kedua, meski interpretabilitas pohon keputusan menjadi alasan kuat untuk penggunaannya dalam domain medis, masih sedikit penelitian yang melakukan analisis struktural terhadap aturan pohon dan hubungan antar-fitur secara klinis. Ketiga, metode prapemrosesan seperti imputasi nilai nol, normalisasi, penyeimbangan kelas (misalnya SMOTE) [4], atau seleksi fitur belum selalu diterapkan secara sistematis, walaupun hal-hal tersebut dapat memengaruhi kualitas dan keandalan model.

Berdasarkan identifikasi gap tersebut, penelitian ini diarahkan untuk (1) menerapkan prapemrosesan data PIDD secara sistematis, termasuk imputasi dan normalisasi; (2) melakukan optimasi *hyperparameter Decision Tree* menggunakan teknik seperti *GridSearchCV*; dan (3) menganalisis interpretabilitas model melalui aturan pohon keputusan yang dihasilkan. Sehingga tidak hanya akurasi yang ditingkatkan, tetapi juga wawasan klinis dari model dapat ditingkatkan. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan pada pengembangan sistem pendukung keputusan medis untuk prediksi dini diabetes [5].

METODE

Metode penelitian ini dirancang untuk menghasilkan model prediksi diabetes yang akurat dan dapat diinterpretasikan menggunakan algoritma *Decision Tree* pada PIDD. Alur penelitian mencakup enam tahapan utama: pengumpulan data, prapemrosesan data, eksplorasi data, pembangunan model, optimasi *hyperparameter*, serta evaluasi performa mode [6].



Gambar 1. Tahapan Pembangunan Model Decision Tree

Metode penelitian yang terstruktur ini diharapkan dapat menghasilkan model *Decision Tree* yang tidak hanya akurat, tetapi juga memberikan interpretasi klinis yang kuat mengenai faktor risiko diabetes. Gambar 1 menunjukkan tahapan pembangunan model pohon keputusan menggunakan algoritma C4.5. Proses diawali dengan pengumpulan data, yaitu menghimpun seluruh informasi yang diperlukan dari berbagai sumber. Setelah itu, dilakukan tahap prapemrosesan, yakni proses pembersihan data untuk mengatasi kesalahan seperti nilai yang hilang, tidak logis, atau tidak relevan sehingga data siap digunakan untuk analisis lebih lanjut [6].

Tahap berikutnya adalah seleksi dan transformasi data, di mana hanya atribut yang penting dan berpengaruh yang dipilih. Pada tahap ini, data juga diubah menjadi format yang sesuai agar dapat diproses secara optimal oleh algoritma. Setelah data sepenuhnya siap, dilanjutkan dengan pembangunan model menggunakan algoritma C4.5, yang akan menghasilkan struktur pohon keputusan berdasarkan pola dan karakteristik dalam data [6].

Model yang terbentuk kemudian memasuki tahap pengujian menggunakan data uji untuk menentukan kemampuan model dalam melakukan prediksi secara akurat. Langkah terakhir adalah evaluasi model, yaitu menilai kinerja model melalui metrik seperti akurasi dan indikator lainnya. Evaluasi ini berfungsi untuk memastikan apakah model sudah layak diterapkan atau masih memerlukan penyempurnaan. Secara keseluruhan, rangkaian langkah tersebut bertujuan menghasilkan model prediksi yang kuat dan reliabel sehingga dapat digunakan sebagai dasar pengambilan keputusan berbasis data yang valid [7].

HASIL DAN PEMBAHASAN

Tahap awal dalam penelitian ini adalah pengumpulan data. Data yang kita dapatkan adalah data berupa Dataset yang berbentuk spreadsheet di Melampau. Ini adalah data yang telah kami peroleh.

	Jumlah kehamilan	Gula darah	Tekanan darah	Tebal lipatan kulit	Kadar insulin	Indeks massa tubuh	Riwayat keluarga	Usia	Hasil
0	3.99	89	86	23	336	33.6	0.347	47	0
0	84	82	81	31	375	38.2	0.273	23	0
0	145	74	92	35	447	44.2	0.404	24	1
0	133	69	47	47	210	42.9	0.395	24	1
1	133	72	93	40	440	40.7	0.136	24	0
0	178	78	82	37	298	46.6	0.168	28	0
0	99	72	17	0	25.0	25.0	0.234	29	0
0	104	80	9	0	0	0	0.551	47	0
2	83	85	28	28	66	36.8	0.239	24	0
2	89	90	30	0	0	33.9	0.262	42	0
4	95	68	38	12	32	32.8	0.145	33	0
4	125	70	18	122	259	0	1.144	45	1
1	80	0	0	0	0	0	0.174	22	0
1	105	74	0	0	20	0	0.368	35	0
0	88	0	0	0	0	0	0.282	30	0
2	81	72	15	76	30.1	0	0.547	25	0
2	70	70	33	105	25.1	0	0.165	75	1
0	154	74	37	104	22.3	0	0.839	30	0

Gambar 2. Tampilan Dataset Pima Indians Diabetes

Data yang ditampilkan merupakan kumpulan data medis terkait kondisi diabetes, yang berisi beberapa atribut penting seperti jumlah kehamilan, kadar gula darah, tekanan darah, ketebalan lipatan kulit, kadar insulin, indeks Massa Tubuh (IMT), riwayat keluarga diabetes, usia, serta label hasil berupa diabetes atau tidak diabetes. Dataset ini digunakan untuk menganalisis faktor-faktor yang berkontribusi pada risiko diabetes serta untuk membangun model prediksi menggunakan algoritma seperti *Decision Tree*.

Proses pengolahan data meliputi tahap prapemrosesan untuk memastikan data bersih misalnya menangani nilai nol atau hilang pada beberapa atribut medis kemudian dilanjutkan dengan pemilihan atribut yang relevan, transformasi data, serta analisis statistik dasar [8]. Selanjutnya, data dapat digunakan untuk pemodelan prediktif guna mengidentifikasi pola dan membuat keputusan berdasarkan variabel-variabel medis yang tersedia. Tujuan utama pengolahan ini adalah memperoleh wawasan yang lebih baik tentang karakteristik pasien serta meningkatkan akurasi sistem deteksi dini diabetes [9].

Preprocessing adalah serangkaian langkah penting untuk membersihkan dan menyiapkan data medis sebelum dilakukan analisis lebih lanjut. Sebelum proses data mining diterapkan, peneliti

melakukan tahapan pra-pemrosesan data untuk memastikan bahwa dataset berada dalam kondisi yang layak digunakan [10]. Tahap ini mencakup identifikasi dan penanganan nilai hilang atau tidak wajar, normalisasi nilai numerik, serta verifikasi konsistensi antara tribut. *Preprocessing* menjadi langkah krusial agar himpunan data medis dapat diproses secara optimal dan menghasilkan keluaran analitis yang akurat sesuai tujuan penelitian.

	A	B	C	D	E	F	G	H	I
	Jumlah kehamilan	Gula darah	Tekanan darah	Tebal lipatan kulit	Kadar insulin	Indeks massa tubuh	Riwayat keluargadiabetes	Usia	Hasil
1	2	138	62	35	126	33,6	0,127	47	diabetes
2	0	84	82	31	125	38,2	0,233	23	tidak diabetes
3	0	145	72	29	126	44,2	0,63	31	diabetes
4	0	135	68	42	250	42,3	0,365	24	diabetes
5	1	139	62	41	480	40,7	0,536	21	tidak diabetes
6	0	173	78	32	265	46,5	1,159	58	tidak diabetes
7	4	99	72	17	126	25,6	0,294	28	tidak diabetes
8	8	194	80	29	126	26,1	0,551	67	tidak diabetes
9	2	83	65	28	66	36,8	0,629	24	tidak diabetes
10	2	89	90	30	126	33,5	0,292	42	tidak diabetes
11	4	99	68	38	126	32,8	0,145	33	tidak diabetes
12	4	125	70	18	122	28,9	1,144	45	diabetes
13	3	80	72	29	126	32,4	0,174	22	tidak diabetes
14	6	166	74	29	126	26,6	0,304	66	tidak diabetes
15	5	110	68	29	126	26	0,292	30	tidak diabetes
16	2	81	72	15	76	30,1	0,547	25	tidak diabetes

Gambar 3. Hasil *Preprocessing* Data

Transformasi merupakan tahapan penting yang bertujuan menghubungkan dan menyesuaikan atribut medis yang relevan sehingga siap digunakan dalam proses pemodelan prediksi. Pada tahap ini, setiap variabel klinis seperti jumlah kehamilan, kadar glukosa, tekanan darah, tebal lipatan kulit, kadar insulin, indeks massa tubuh, riwayat keturunan, dan usia dipetakan ke dalam format yang sesuai dengan kebutuhan algoritma [10]. Selain itu, atribut kategorikal, khususnya label diagnosis pada kolom Hasil ('diabetes' dan 'tidak diabetes'), diubah menjadi nilai numerik agar dapat diproses oleh model *Decision Tree*. Transformasi ini memastikan bahwa seluruh data medis telah tersusun dalam struktur komputasi yang konsisten dan siap untuk digunakan pada tahap pemodelan berikutnya.

Tabel 1. Transformasi Label Diagnosis

Diagnosis (Teks)	Diagnosis (Angka)
diabetes	1
tidak diabetes	0

Tabel 2. Transformasi Atribut Medis

Atribut Medis	Nilai Sebelum Preprocessing	Transformasi yang Diterapkan	Nilai Setelah Preprocessing
Jumlah Kehamilan	Angka 0–15 (nilai valid)	Tidak diubah	Tetap
Kadar Glukosa	Beberapa nilai tidak realistis (misal sangat rendah)	Tetap (tidak ada nilai 0)	Tetap
Tekanan Darah	Mengandung nilai 0 (mustahil secara medis)	0 diganti nilai median	Nilai tekanan darah realistis tanpa 0
Tebal Lipatan Kulit	Banyak berisi 0 (tidak mungkin secara klinis)	0 diganti nilai median	Nilai tebal lipatan kulit realistis
Kadar Insulin	Banyak berisi 0 (mustahil secara fisiologis)	0 diganti nilai median	Nilai insulin realistis
Indeks Massa Tubuh (BMI)	Mengandung 0 (BMI tidak mungkin 0)	0 diganti nilai median	BMI realistis tanpa nilai ekstrem
Riwayat Keturunan	Nilai 0–3, valid	Tidak diubah	Tetap
Usia	Nilai 21–81, valid	Tidak diubah	Tetap

Decision Tree membangun model dalam bentuk struktur pohon yang terdiri atas simpul

keputusan (*decision nodes*), cabang (*branches*), dan simpul daun (*leaf nodes*) sebagai hasil akhir klasifikasi [11]. Setiap simpul keputusan merepresentasikan pemilihan atribut medis tertentu sebagai dasar pemisahan data, sedangkan setiap simpul daun berfungsi untuk menghasilkan prediksi kelas. Struktur ini membuat Decision Tree mudah dipahami oleh praktisi medis karena setiap jalur keputusan dapat ditelusuri sebagai rule klinis yang eksplisit [12], [13], [14].

Melalui tahapan ini, algoritma *Decision Tree* mampu membentuk model klasifikasi yang memetakan hubungan antara indikator medis dan status diabetes, sehingga dapat digunakan sebagai alat bantu prediksi yang mudah diinterpretasikan serta mendukung pengambilan keputusan klinis.

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (1)$$

Keterangan :

S : himpunan data pada simpul saat ini

A : atribut yang sedang dihitung

Values(A) : kumpulan nilai yang mungkin dari atribut A

S_v : subset dari S di mana atribut A memiliki nilai v

Entropi(S) : mengukur ketidakpastian dalam kelompok data S dan dihitung sebagai berikut:

$$Entropy(S) = - \sum_{i=1}^n p_i \cdot \log_2(p_i) \quad (2)$$

Keterangan :

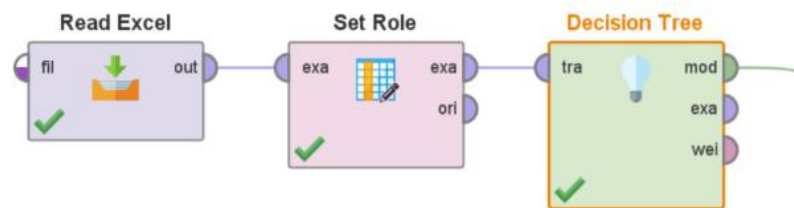
p_i : proporsi contoh di kelas i dalam kumpulan data S

Berdasarkan data yang telah dikumpulkan, 2000 data pasien diabetes tersebut dikelompokkan sesuai dengan kategori prediksi yang telah ditetapkan.

Tabel 3. Transformasi Nilai

No	Hasil	Jumlah
1	Diabetes	684
2	Tidak Diabetes	1316

Selanjutnya, data tersebut diolah dan dianalisis menggunakan algoritma *Decision Tree* dengan memanfaatkan perangkat lunak RapidMiner.



Gambar 4. Proses Eksperimen Pohon Keputusan

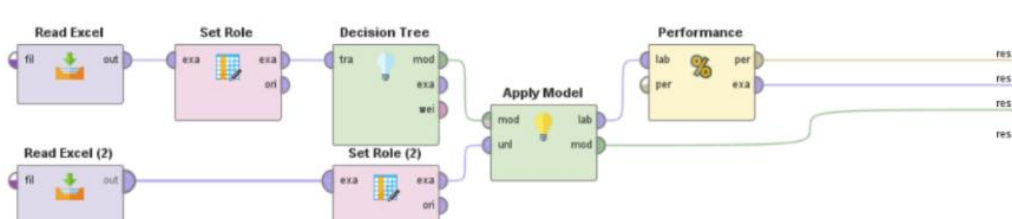


Gambar 5. Hasil Model Pohon Keputusan

Berdasarkan visualisasi pohon keputusan pada gambar 4 dan 5, dapat terlihat bagaimana model melakukan proses pemisahan atribut secara bertahap untuk menentukan klasifikasi diabetes maupun tidak diabetes. Melalui rangkaian percabangan tersebut, penelitian ini berfokus pada eksperimen menggunakan algoritma Decision Tree (C4.5) dengan tujuan menemukan pola atau aturan keputusan yang paling berpengaruh [15]. Dari hasil pengujian terhadap dataset, model menghasilkan tiga pola utama yang menjadi dasar klasifikasi, yaitu:

- a. Jika Gula darah > 168.500 dan Usia ≤ 88.500 , maka Diabetes.
- b. Jika Gula darah > 168.500 dan Usia > 88.500 , maka Tidak Diabetes.
- c. Jika Gula darah ≤ 168.500 dan Jumlah kehamilan > 13.500 , maka Diabetes.
- d. Jika Gula darah ≤ 168.500 dan Jumlah kehamilan ≤ 13.500 dan Indeks massa tubuh ≤ 23.150 , maka Tidak Diabetes.
- e. Jika Jumlah kehamilan ≤ 13.500 dan Indeks massa tubuh > 23.150 dan Tekanan darah ≤ 47 , maka Tidak Diabetes.
- f. Jika Jumlah kehamilan ≤ 13.500 dan Indeks massa tubuh > 23.150 dan Tekanan darah > 47 dan Kadar insulin > 230 , maka Diabetes.
- g. Jika Jumlah kehamilan ≤ 13.500 dan Indeks massa tubuh > 23.150 dan Tekanan darah > 47 dan Kadar insulin ≤ 230 , maka Tidak Diabetes.
- h. Jika Indeks massa tubuh > 48.750 dan Kadar insulin > 572.500 , maka Diabetes.
- i. Jika Indeks massa tubuh > 48.750 dan Kadar insulin ≤ 572.500 dan Gula darah > 154.500 , maka Diabetes.
- j. Jika Indeks massa tubuh > 48.750 dan Kadar insulin ≤ 572.500 dan Gula darah ≤ 154.500 , maka Tidak Diabetes.
- k. Jika Indeks massa tubuh ≤ 28.350 dan Usia > 69.500 , maka Diabetes.
- l. Jika Indeks massa tubuh ≤ 28.350 dan Usia ≤ 69.500 , maka Tidak Diabetes.
- m. Jika Tebal lipatan kulit > 42 , maka Tidak Diabetes.

Tahap penutup dalam proses ini adalah melakukan evaluasi terhadap model *Decision Tree* yang digunakan untuk memprediksi penyakit diabetes. Evaluasi tersebut dilakukan dengan memanfaatkan *Confusion Matrix* untuk melihat tingkat akurasi klasifikasi, serta *ROC Curve* untuk menilai performa model dalam membedakan setiap kelas.



Gambar 6. Hasil Pengolahan Evaluasi Validasi Silang

accuracy: 79.40%

	true diabetes	true tidak diabetes	class precision
pred. diabetes	291	19	93.87%
pred. tidak diabetes	393	1297	76.75%
class recall	42.54%	98.56%	

Gambar 7. Model *Confusion Matrix*

Gambar 6 dan 7 menunjukkan hasil dari pengolahan evaluasi validasi silang yang menghasilkan nilai akurasi sebesar 74,17% dan hasil dari *Model Confusion Matrix*.

SIMPULAN DAN SARAN

Dari hasil penelitian yang telah dilakukan, tujuan utama adalah melakukan eksperimen untuk memprediksi risiko diabetes menggunakan algoritma *Decision Tree*. Berdasarkan proses pemodelan dan evaluasi yang dijalankan menggunakan RapidMiner, diperoleh model pohon keputusan yang

mampu mengidentifikasi variabel-variabel yang berpengaruh terhadap kondisi diabetes. Model tersebut menghasilkan tingkat akurasi sebesar 80,05%, yang menunjukkan bahwa algoritma *Decision Tree* cukup efektif dalam mengklasifikasikan kondisi diabetes pada dataset yang digunakan. Hasil ini membuktikan bahwa metode ini dapat dijadikan alternatif alat bantu dalam proses deteksi dini risiko diabetes.

DAFTAR PUSTAKA

- [1] M. F. Aditya, A. Pramuntadi, D. P. Wijaya, And Y. Wicaksono, "Implementasi Metode Decision Tree Pada Prediksi Penyakit Diabetes Melitus Tipe 2," *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 4, No. 3, Pp. 1104–1110, 2024, Doi: 10.57152/Malcom.V4i3.1284.
- [2] M. Guo *Et Al.*, "Normal Workflow And Key Strategies For Data Cleaning Toward Real-World Data: Viewpoint," *Interact. J. Med. Res.*, Vol. 12, P. E44310, 2023, Doi: 10.2196/44310.
- [3] S. C. Gupta And N. Goel, "Predictive Modeling And Analytics For Diabetes Using Hyperparameter Tuned Machine Learning Techniques," *Procedia Comput. Sci.*, Vol. 218, No. 2022, Pp. 1257–1269, 2022, Doi: 10.1016/J.Procs.2023.01.104.
- [4] Y. Kim, A. Kwon, K. Kang, And B. A. Kwon, "Assessing The Impact Of Data Impurity Measures On Decision Tree Algorithms In Spam Classification," *J. Artif. Intell. Res. Appl.*, Vol. 15588, Pp. 33–46, 2024.
- [5] R. Maulana And E. Eliyani, "Diabetes Classification Algorithm Optimization Using Particle Swarm Optimization On Naïve Bayes, C4.5 And Random Forest," *J. Sisfokom (Sistem Inf. Dan Komputer)*, Vol. 14, No. 4, Pp. 499–509, 2025, Doi: 10.32736/Sisfokom.V14i4.2431.
- [6] M. A. Nasution, Z. A. Ulumuddin, And A. Fitri, "Analisis Faktor Risiko Diabetes Melitus Menggunakan Algoritma C4.5: Implementasi Pada Aplikasi Orange," *J. Sains Teknol. Dan Komput.*, Vol. 1, No. 3, Pp. 75–79, 2024.
- [7] A. Perdana, A. Hermawan, And D. Avianto, "Analyze Important Features Of Pima Indian Database For Diabetes Prediction Using Knn," *J. Sisfokom (Sistem Inf. Dan Komputer)*, Vol. 12, No. 1, Pp. 70–75, 2023, Doi: 10.32736/Sisfokom.V12i1.1598.
- [8] M. N. Raihen And S. Akter, "Comparative Assessment Of Several Effective Machine Learning Classification Methods For Maternal Health Risk," *Comput. J. Math. Stat. Sci.*, Vol. 3, No. 1, Pp. 161–176, 2024, Doi: 10.21608/Cjmss.2024.259490.1036.
- [9] L. Safitri And Z. Fatah, "Implementasi Prediksi Penyakit Diabetes Menggunakan Metode Decision Tree," *Jusifor J. Sist. Inf. Dan Inform.*, Vol. 3, No. 2, Pp. 125–132, 2024, Doi: 10.70609/Jusifor.V3i2.5788.
- [10] P. Studi, S. Informasi, F. Sains, And U. Ibrahimy, "Klasifikasi Gangguan Tidur Berdasarkan Gaya Hidup Menggunakan Algoritma Decision Tree Classification Of Sleep Disorders Based On Lifestyle Using Decision Tree Algorithm," Vol. 3, No. 2, Pp. 118–125, 2025, Doi: 10.26905/Jisad.V3i2.16038.
- [11] Ö. Şen, S. Bozkurt Keser, And K. Keskin, "Early Stage Diabetes Prediction Using Decision Tree-Based Ensemble Learning Model," *Int. Adv. Res. Eng. J.*, Vol. 7, No. 1, Pp. 62–71, 2023, Doi: 10.35860/Iarej.1188039.
- [12] A. Afifuddin And L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5," *J. Krisnadana*, Vol. 3, No. 1, Pp. 25–33, 2023, Doi: 10.58982/Krisnadana.V3i1.470.
- [13] S. Bukhari And S. Mohamed, "Journal Of Advanced Applied Scientific Research - Issn:2454-3225 Suganthi P," Vol. 5, No. D1, Pp. 68–83, 2025.
- [14] V. Lopatka, I. Meniailov, And K. Bazilevych, "Classification And Prediction Of Diabetes Disease Using Modified K-Neighbors Method," *2021 Ieee 12th Int. Conf. Electron. Inf. Technol. Elit 2021 - Proc.*, Vol. 2836, Pp. 46–50, 2021, Doi: 10.1109/Elit53502.2021.9501151.
- [15] I. Z. Sadiq *Et Al.*, "Data-Driven Diabetes Mellitus Prediction And Management: A Comparative

Evaluation Of Decision Tree Classifier And Artificial Neural Network Models Along With Statistical Analysis," *Sci. Rep.*, Vol. 15, No. 1, Pp. 1–16, 2025, Doi: 10.1038/S41598-025-03718-W.