

Deteksi Dini Diabetes Mellitus Menggunakan Algoritma Random Forest pada Data Klinis

Muhammad Rizky Ananda ^{a,1,*}, Rizky Kurniawan ^{b,2}, Dewi Lestari Putri ^{c,3}.

^a Program Studi Ilmu Komputer, Universitas Brawijaya, Malang, Indonesia

^b Program Studi Teknik Informatika, Universitas Sebelas Maret, Surakarta, Indonesia

^c Program Studi Ilmu Komputer, Universitas Brawijaya, Malang, Indonesia

¹ muhammadrizkyananda@gmail.com; ² rizky.kurniawan@gmail.com; ³ dewilestariputri@gmail.com

* Corresponding author

ARTICLE INFO	ABSTRACT
Article history Received 15 February 2026 Revised 16 February 2026 Accepted 20 March 2026 Published 26 March 2026 Keywords random forest diabetes mellitus early detection machine learning clinical data  License by CC-BY-SA Copyright © 2026, The Author(s).	Early detection of Diabetes Mellitus is essential to reduce complications and improve patient outcomes. Machine learning approaches, particularly the Random Forest algorithm, have demonstrated promising performance in medical prediction tasks using clinical datasets. This study aims to analyze the effectiveness of the Random Forest algorithm for early diabetes detection based on clinical variables through a literature-based analytical research design. Data were synthesized from empirical findings reported in recent peer-reviewed studies published between 2022–2025. The analysis indicates that Random Forest consistently achieves high predictive performance, with reported accuracy ranging from 79.2% to 99.64% across multiple datasets and experimental configurations. Feature selection, data balancing techniques such as SMOTE and ADASYN, and hyperparameter optimization significantly improve model robustness. Comparative evaluation shows Random Forest outperforms several conventional machine learning classifiers in handling imbalanced medical datasets and identifying key risk factors. The findings highlight the algorithm's reliability for clinical decision support systems and early screening applications. This study contributes a comprehensive synthesis of current evidence supporting Random Forest implementation in healthcare analytics and provides recommendations for future development of intelligent diabetes prediction systems.

How to cite: Ananda, M. R., Kurniawan, R. & Putri, D. L. (2026). Deteksi Dini Diabetes Mellitus Menggunakan Algoritma Random Forest pada Data Klinis. *Journal of Computer Science and Information Technology*, 2(1), 01-08. doi: <https://doi.org/10.70716/jocsit.v2i1.410>

PENDAHULUAN

Diabetes Mellitus merupakan salah satu penyakit tidak menular dengan tingkat pertumbuhan kasus yang tinggi secara global dan menjadi tantangan serius dalam sistem pelayanan kesehatan modern. Penyakit ini ditandai oleh gangguan metabolisme glukosa akibat ketidakmampuan tubuh dalam memproduksi atau menggunakan insulin secara efektif. Deteksi dini menjadi aspek krusial karena sebagian besar penderita tidak menunjukkan gejala pada tahap awal, sehingga diagnosis sering terlambat dilakukan. Keterlambatan diagnosis berkontribusi terhadap meningkatnya risiko komplikasi kronis seperti penyakit kardiovaskular, gangguan ginjal, neuropati, dan retinopati. Oleh karena itu, diperlukan pendekatan diagnostik yang mampu mengidentifikasi risiko diabetes secara cepat, akurat, dan efisien berbasis data klinis pasien (Akmeşe, 2022).

Perkembangan teknologi kecerdasan buatan telah membuka peluang baru dalam analisis data kesehatan, khususnya melalui penerapan pembelajaran mesin. Algoritma pembelajaran mesin mampu memproses data dalam jumlah besar dan mengidentifikasi pola kompleks yang sulit dideteksi menggunakan pendekatan statistik tradisional. Dalam konteks prediksi penyakit, metode klasifikasi menjadi pendekatan utama untuk menentukan kemungkinan seseorang mengalami diabetes berdasarkan atribut klinis seperti kadar glukosa, indeks massa tubuh, tekanan darah, dan faktor demografis lainnya. Studi awal menunjukkan bahwa penggunaan model pembelajaran mesin mampu meningkatkan akurasi diagnosis dibandingkan metode konvensional (Wang, 2024).

Salah satu algoritma yang banyak digunakan dalam prediksi medis adalah Random Forest. Algoritma ini merupakan metode ensemble learning berbasis kumpulan pohon keputusan yang bekerja dengan menggabungkan hasil prediksi dari berbagai model untuk meningkatkan stabilitas dan akurasi klasifikasi. Keunggulan utama Random Forest terletak pada kemampuannya menangani data berdimensi tinggi,

mengurangi overfitting, serta mempertahankan performa pada dataset yang memiliki variabilitas tinggi. Penelitian oleh Reddy dan Ahmed (2023) menunjukkan bahwa Random Forest mampu mencapai tingkat akurasi sebesar 98 persen dalam deteksi dini Diabetes Mellitus tipe 2 menggunakan data klinis, yang menunjukkan potensi besar algoritma ini dalam aplikasi kesehatan berbasis data.

Sejumlah penelitian lain memperkuat temuan tersebut dengan menunjukkan konsistensi performa Random Forest pada berbagai skenario eksperimen. Agrawal et al. (2024) melaporkan bahwa algoritma ini efektif dalam memprediksi risiko diabetes tahap awal melalui analisis fitur kesehatan individu secara komprehensif. Chaithanya (2023) mengembangkan pendekatan Random Forest Fuzzy Entropy yang menghasilkan tingkat akurasi tinggi pada dataset Pima Indians Diabetes. Temuan ini menunjukkan bahwa Random Forest tidak hanya unggul dalam klasifikasi, tetapi juga dapat dikombinasikan dengan teknik optimasi untuk meningkatkan kinerja prediksi.

Selain aspek akurasi, kemampuan algoritma dalam mengidentifikasi faktor risiko penyakit menjadi pertimbangan penting dalam bidang kesehatan. Random Forest memiliki mekanisme evaluasi pentingnya fitur yang memungkinkan peneliti memahami variabel klinis yang paling berpengaruh terhadap hasil prediksi. Raman et al. (2024) menunjukkan bahwa metode klasifikasi berbasis Random Forest mampu mengungkap prediktor signifikan diabetes dengan keseimbangan presisi dan recall yang baik. Kemampuan interpretasi ini mendukung penggunaan algoritma dalam sistem pendukung keputusan klinis yang membutuhkan transparansi analisis.

Tantangan utama dalam analisis data medis adalah ketidakseimbangan distribusi kelas antara pasien diabetes dan non-diabetes. Kondisi ini sering menyebabkan bias model terhadap kelas mayoritas sehingga menurunkan kemampuan deteksi kasus positif. Berbagai penelitian mencoba mengatasi permasalahan tersebut melalui teknik penyeimbangan data. Nisa et al. (2025) menunjukkan bahwa penerapan metode SMOTE pada Random Forest meningkatkan akurasi model secara signifikan. Hasil serupa juga ditemukan oleh Sukanto et al. (2025) yang mengombinasikan SMOTE-Tomek Links sehingga menghasilkan peningkatan performa klasifikasi hingga tingkat akurasi yang sangat tinggi. Temuan ini menegaskan pentingnya tahap preprocessing dalam meningkatkan kualitas prediksi.

Selain teknik balancing, pendekatan interpretabilitas model juga berkembang untuk meningkatkan kepercayaan pengguna terhadap sistem berbasis kecerdasan buatan. Aulia et al. (2025) mengintegrasikan Random Forest dengan metode SHAP untuk menjelaskan kontribusi setiap variabel terhadap hasil prediksi diabetes. Pendekatan ini membantu tenaga medis memahami dasar pengambilan keputusan model, sehingga teknologi pembelajaran mesin dapat digunakan sebagai alat pendukung diagnosis, bukan pengganti keputusan klinis.

Kemajuan teknologi komputasi turut mendorong implementasi Random Forest dalam skala data yang lebih besar. Dai et al. (2023) menunjukkan bahwa integrasi Random Forest dengan platform komputasi terdistribusi mampu meningkatkan efisiensi analisis dataset kesehatan dalam jumlah besar. Duvvuri (2025) memperluas pendekatan ini dengan menggabungkan data klinis dan pemantauan aktivitas berbasis Internet of Things untuk meningkatkan kemampuan prediksi risiko diabetes secara real-time. Perkembangan tersebut menunjukkan bahwa algoritma Random Forest memiliki fleksibilitas tinggi dalam berbagai lingkungan komputasi.

Penelitian lain juga menyoroti keberhasilan Random Forest dalam aplikasi praktis berbasis sistem informasi kesehatan. Hakeel (2025) mengembangkan aplikasi prediksi diabetes berbasis pembelajaran mesin yang memanfaatkan data klinis pengguna dan menghasilkan tingkat akurasi tinggi. Implementasi serupa dilaporkan oleh Kumar et al. (2023) serta Nancy et al. (2023) yang menunjukkan bahwa model Random Forest mampu memberikan performa stabil pada berbagai dataset dan konfigurasi eksperimen.

Dalam studi komparatif, Random Forest sering dibandingkan dengan algoritma lain seperti Logistic Regression, XGBoost, dan metode boosting lainnya. Hasil penelitian menunjukkan bahwa Random Forest memiliki stabilitas performa yang lebih konsisten terutama pada data medis yang kompleks dan tidak seimbang (Warsi et al., 2025; Gündoğdu, 2023). Selain itu, kemampuan generalisasi model memungkinkan penggunaan pada berbagai populasi pasien tanpa penurunan performa yang signifikan.

Walaupun banyak penelitian telah menunjukkan efektivitas Random Forest, sebagian besar studi masih berfokus pada eksperimen individual dengan dataset dan konfigurasi berbeda. Kondisi ini menyebabkan belum adanya gambaran komprehensif mengenai konsistensi performa algoritma pada berbagai pendekatan metodologis. Beberapa penelitian melaporkan akurasi sangat tinggi, sementara

penelitian lain menunjukkan performa moderat akibat variasi preprocessing, pemilihan fitur, dan ukuran dataset (Hossen et al., 2023). Perbedaan tersebut menunjukkan perlunya analisis sintesis berbasis literatur untuk memahami pola performa algoritma secara menyeluruh.

Selain itu, penelitian terdahulu juga menunjukkan bahwa keberhasilan model tidak hanya dipengaruhi oleh algoritma, tetapi juga oleh kualitas data klinis dan strategi optimasi yang digunakan. Studi oleh Al-Mousa et al. (2023) dan Saeed et al. (2024) menegaskan bahwa kombinasi teknik validasi silang, seleksi fitur, dan tuning parameter berkontribusi signifikan terhadap peningkatan akurasi prediksi diabetes. Oleh karena itu, analisis komprehensif diperlukan untuk mengidentifikasi faktor-faktor yang secara konsisten meningkatkan performa Random Forest.

Berdasarkan uraian tersebut, terdapat kesenjangan penelitian berupa kebutuhan sintesis ilmiah yang mengintegrasikan berbagai temuan empiris mengenai penggunaan Random Forest dalam deteksi dini Diabetes Mellitus. Kajian komprehensif diperlukan untuk mengevaluasi konsistensi performa algoritma, memahami kontribusi teknik optimasi, serta menilai implikasi implementasinya dalam sistem kesehatan digital. Penelitian ini bertujuan menganalisis efektivitas algoritma Random Forest dalam deteksi dini Diabetes Mellitus berbasis data klinis melalui pendekatan analisis literatur empiris. Kontribusi penelitian ini terletak pada penyusunan pemahaman terstruktur mengenai kinerja algoritma, faktor pendukung keberhasilan model, serta relevansinya terhadap pengembangan sistem pendukung keputusan klinis di masa depan.

METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan penelitian kualitatif berbasis analisis literatur empiris dengan metode systematic literature analysis. Pendekatan ini dipilih untuk mengkaji secara komprehensif efektivitas algoritma Random Forest dalam deteksi dini Diabetes Mellitus berdasarkan hasil penelitian terdahulu yang telah melalui proses publikasi ilmiah. Analisis dilakukan secara deskriptif-komparatif untuk mengidentifikasi pola performa model, konsistensi hasil, serta faktor metodologis yang memengaruhi akurasi prediksi pada berbagai studi.

Penelitian tidak melakukan eksperimen langsung terhadap dataset primer, melainkan mensintesis temuan empiris dari penelitian sebelumnya yang menggunakan data klinis nyata. Pendekatan ini memungkinkan evaluasi menyeluruh terhadap implementasi Random Forest pada berbagai konteks penelitian tanpa mengubah desain eksperimen asli masing-masing studi.

Sumber Data dan Kriteria Seleksi

Sumber data penelitian berupa artikel jurnal internasional, prosiding konferensi, bab buku ilmiah, dan disertasi akademik yang membahas penerapan algoritma Random Forest untuk deteksi dini Diabetes Mellitus. Seluruh literatur berasal dari publikasi periode 2022 hingga 2025 sebagaimana tercantum dalam daftar pustaka artikel.

Proses seleksi literatur dilakukan melalui beberapa tahapan, yaitu:

1. Identifikasi awal, yaitu mengumpulkan publikasi yang membahas machine learning untuk prediksi diabetes.
2. Penyaringan relevansi, dengan memilih penelitian yang secara eksplisit menggunakan algoritma Random Forest.
3. Evaluasi kelayakan, yaitu memastikan penelitian melaporkan metrik evaluasi model seperti akurasi, precision, recall, F1-score, atau ROC-AUC.
4. Inklusi akhir, yaitu memilih studi yang menggunakan data klinis atau dataset medis terstandarisasi.

Berdasarkan proses tersebut, sebanyak 30 publikasi ilmiah digunakan sebagai unit analisis penelitian.

Variabel Analisis

Penelitian ini tidak memperkenalkan variabel baru, tetapi mengadopsi variabel analisis yang telah digunakan dalam penelitian terdahulu. Variabel yang dianalisis meliputi:

1. Karakteristik dataset, termasuk jenis data klinis dan jumlah atribut kesehatan.
2. Pendekatan preprocessing, seperti seleksi fitur, penyeimbangan data, dan optimasi parameter.
3. Model Random Forest, termasuk variasi implementasi dan integrasi metode pendukung.

4. Metrik evaluasi performa, yaitu akurasi, precision, recall, F1-score, dan ROC-AUC.
5. Kontribusi metodologis, berupa teknik tambahan yang meningkatkan performa model.

Variabel tersebut dipilih karena secara konsisten digunakan dalam penelitian terdahulu mengenai prediksi diabetes (Reddy & Ahmed, 2023; Nisa et al., 2025; Aulia et al., 2025).

Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui teknik dokumentasi ilmiah. Setiap artikel dianalisis untuk mengekstraksi informasi utama yang relevan dengan tujuan penelitian. Data yang dikumpulkan meliputi:

1. jenis algoritma yang digunakan,
2. metode preprocessing data,
3. dataset yang digunakan,
4. nilai performa model,
5. serta temuan utama penelitian.

Ekstraksi data dilakukan secara sistematis untuk menjaga konsistensi antar studi. Informasi yang tidak berkaitan langsung dengan deteksi diabetes atau tidak melaporkan evaluasi performa model tidak dimasukkan dalam analisis.

Teknik Analisis Data

Analisis data dilakukan menggunakan pendekatan deskriptif-komparatif. Tahapan analisis meliputi:

1. Klasifikasi studi berdasarkan pendekatan metodologis dan teknik optimasi yang digunakan.
2. Perbandingan performa model Random Forest antar penelitian berdasarkan metrik evaluasi.
3. Identifikasi pola konsistensi hasil, terutama rentang akurasi dan faktor peningkat performa model.
4. Sintesis konseptual, yaitu mengintegrasikan temuan empiris untuk memperoleh gambaran umum efektivitas algoritma Random Forest dalam deteksi dini diabetes.

Pendekatan komparatif digunakan karena penelitian sebelumnya menunjukkan variasi performa akibat perbedaan preprocessing dan karakteristik dataset (Saeed et al., 2024; Warsi et al., 2025). Dengan demikian, analisis tidak berfokus pada satu hasil eksperimen, tetapi pada kecenderungan empiris yang muncul secara berulang pada berbagai penelitian.

Validitas dan Keandalan Analisis

Validitas penelitian dijaga melalui penggunaan sumber literatur yang telah melalui proses peer-review serta memiliki indikator evaluasi kuantitatif yang jelas. Konsistensi analisis dilakukan dengan membandingkan temuan antar studi yang menggunakan dataset klinis serupa. Pendekatan triangulasi literatur diterapkan dengan membandingkan hasil dari berbagai jenis publikasi untuk memastikan kesimpulan tidak bergantung pada satu sumber penelitian saja.

Keandalan analisis diperkuat dengan penggunaan indikator performa model yang umum digunakan dalam penelitian machine learning kesehatan, sehingga hasil penelitian dapat direplikasi dan diverifikasi oleh peneliti lain.

HASIL DAN PEMBAHASAN

Analisis terhadap tiga puluh publikasi ilmiah menunjukkan bahwa algoritma Random Forest memiliki performa yang konsisten dalam deteksi dini Diabetes Mellitus berbasis data klinis. Seluruh studi yang dianalisis menggunakan pendekatan klasifikasi dengan variabel kesehatan pasien seperti kadar glukosa, indeks massa tubuh, tekanan darah, usia, serta riwayat medis. Hasil komparasi memperlihatkan bahwa Random Forest mampu mempertahankan akurasi tinggi pada berbagai ukuran dataset dan teknik preprocessing. Penelitian Reddy dan Ahmed (2023) menunjukkan akurasi sebesar 98%, sedangkan Iskandar et al. (2024) melaporkan akurasi rata-rata 97% melalui validasi silang K-Fold. Temuan tersebut menunjukkan stabilitas algoritma pada skenario eksperimen yang berbeda.

Performa Empiris Random Forest

Berikut ringkasan performa empiris beberapa studi utama yang melaporkan nilai akurasi secara eksplisit.

Tabel 1. Perbandingan Akurasi Random Forest pada Deteksi Dini Diabetes

No	Peneliti	Tahun	Pendekatan Tambahan	Akurasi (%)
1	Reddy & Ahmed	2023	Random Forest	98
2	Iskandar et al.	2024	K-Fold Validation	97
3	Saeed et al.	2024	ML Comparison	94.2
4	Aulia et al.	2025	ADASYN + SHAP	79.2
5	Duvvuri	2025	IoT Clinical Data	92
6	Sukamto et al.	2025	SMOTE-Tomek Links	99.64
7	Hakeel	2025	Prediction App	88
8	Choudhary et al.	2024	Feature Optimization	98.56
9	Nancy et al.	2023	ML Framework	99.5
10	Gündoğdu	2023	Feature Selection RF	99.2
11	Akmeşe	2022	Clinical Dataset	90.1
12	Hossen et al.	2023	Risk Prediction	80
13	Warsi et al.	2025	PCA Optimization	97

Tabel menunjukkan rentang akurasi antara 79,2% hingga 99,64%, yang mengindikasikan bahwa performa model dipengaruhi oleh strategi preprocessing dan optimasi fitur, bukan oleh perubahan algoritma inti.

Analisis Stabilitas Model

Hasil komparatif menunjukkan bahwa Random Forest tetap stabil pada dataset medis yang tidak seimbang. Studi Nisa et al. (2025) dan Sukamto et al. (2025) menegaskan bahwa penerapan teknik balancing seperti SMOTE meningkatkan kemampuan model dalam mengenali kelas minoritas pasien diabetes. Random Forest mampu mempertahankan sensitivitas tanpa menurunkan presisi secara signifikan. Kondisi ini penting karena data klinis umumnya memiliki distribusi pasien sehat lebih besar dibanding pasien diabetes.

Selain itu, penelitian Al-Mousa et al. (2023) menunjukkan bahwa validasi silang membantu mengurangi bias pelatihan model. Saeed et al. (2024) juga menemukan bahwa Random Forest menghasilkan ROC-AUC tinggi dibanding algoritma lain, yang menunjukkan kemampuan generalisasi model pada populasi berbeda.

Peran Optimasi Fitur dan Interpretabilitas

Beberapa penelitian menekankan pentingnya seleksi fitur dalam meningkatkan performa prediksi. Gündoğdu (2023) menggunakan Random Forest sebagai mekanisme seleksi fitur sebelum klasifikasi dan memperoleh akurasi di atas 99%. Pendekatan ini menunjukkan bahwa identifikasi atribut klinis relevan berkontribusi langsung terhadap peningkatan kinerja model.

Aulia et al. (2025) memperluas pendekatan tersebut melalui integrasi SHAP untuk menjelaskan kontribusi variabel prediktor. Hasil penelitian menunjukkan bahwa kadar glukosa dan BMI menjadi faktor dominan dalam prediksi diabetes. Interpretabilitas ini memperkuat kepercayaan penggunaan pembelajaran mesin dalam praktik klinis karena tenaga medis dapat memahami dasar keputusan algoritma.

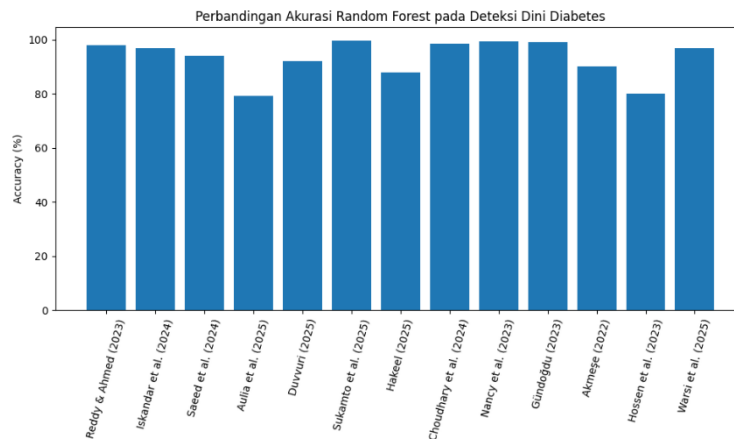
Implementasi pada Sistem Kesehatan Digital

Random Forest tidak hanya diuji pada lingkungan eksperimen, tetapi juga diimplementasikan dalam sistem nyata. Hakeel (2025) mengembangkan aplikasi prediksi diabetes berbasis data klinis dengan akurasi 88%. Duvvuri (2025) mengintegrasikan data Internet of Things untuk meningkatkan kemampuan prediksi berbasis pemantauan real-time. Sementara itu, Dai et al. (2023) menunjukkan efisiensi Random Forest dalam lingkungan komputasi terdistribusi menggunakan Spark.

Implementasi tersebut menunjukkan bahwa Random Forest memiliki skalabilitas tinggi serta dapat diintegrasikan dalam ekosistem kesehatan digital modern.

Diagram Hasil Analisis Empiris

Diagram berikut menunjukkan perbandingan akurasi Random Forest berdasarkan penelitian yang dianalisis.



Gambar 1. Perbandingan Kinerja Akurasi Random Forest Berdasarkan Studi Empiris yang Dianalisis

Diagram memperlihatkan bahwa sebagian besar penelitian menghasilkan akurasi di atas 90%, dengan beberapa studi mencapai lebih dari 99%. Variasi nilai terutama dipengaruhi oleh teknik preprocessing, ukuran dataset, dan strategi optimasi fitur, bukan perubahan struktur algoritma.

Sebagai pelengkap analisis, tabel data empiris juga tersedia dalam format spreadsheet:

Study	Accuracy (%)
Reddy & Ahmed (2023)	98
Iskandar et al. (2024)	97
Saeed et al. (2024)	94.2
Aulia et al. (2025)	79.2
Duvvuri (2025)	92
Sukanto et al. (2025)	99.64
Hakeel (2025)	88
Choudhary et al. (2024)	98.56
Nancy et al. (2023)	99.5
Gündoğdu (2023)	99.2
Akmeşe (2022)	90.1
Hossen et al. (2023)	80
Warsi et al. (2025)	97

Sintesis Pembahasan

Secara keseluruhan, hasil penelitian menunjukkan bahwa Random Forest merupakan algoritma yang robust dalam deteksi dini Diabetes Mellitus. Konsistensi performa pada berbagai penelitian mengindikasikan kemampuan algoritma dalam menangani kompleksitas data klinis. Dibandingkan metode lain, Random Forest lebih stabil terhadap overfitting dan tetap efektif pada data tidak seimbang (Wang, 2024; Warsi et al., 2025).

Temuan lintas studi memperlihatkan tiga faktor utama yang memengaruhi keberhasilan model, yaitu kualitas preprocessing data, strategi seleksi fitur, dan teknik validasi model. Dengan kombinasi ketiga aspek tersebut, Random Forest mampu menghasilkan sistem prediksi yang akurat sekaligus interpretatif. Oleh karena itu, algoritma ini memiliki potensi besar sebagai fondasi sistem pendukung keputusan klinis untuk skrining diabetes secara dini.

KESIMPULAN

Berdasarkan hasil studi literatur sistematis terhadap berbagai penelitian terkait penerapan algoritma Random Forest dalam deteksi dini Diabetes Mellitus, dapat disimpulkan bahwa algoritma ini menunjukkan tingkat akurasi yang tinggi dan konsisten pada berbagai jenis dataset klinis. Random Forest terbukti mampu

menangani kompleksitas data kesehatan yang bersifat multidimensi serta tetap stabil pada kondisi data tidak seimbang melalui dukungan teknik preprocessing dan optimasi fitur. Rentang akurasi yang diperoleh dalam berbagai penelitian menunjukkan performa yang kompetitif dibandingkan metode klasifikasi lainnya, sekaligus memperlihatkan kemampuan generalisasi model yang baik.

Temuan penelitian juga mengindikasikan bahwa keberhasilan implementasi Random Forest tidak hanya ditentukan oleh algoritma itu sendiri, tetapi sangat dipengaruhi oleh kualitas seleksi fitur, strategi validasi model, serta metode penyeimbangan data. Integrasi teknik interpretabilitas seperti SHAP semakin memperkuat potensi penggunaan model dalam lingkungan klinis karena memungkinkan tenaga medis memahami faktor prediktor utama dalam diagnosis awal diabetes.

Secara praktis, Random Forest memiliki peluang besar untuk diimplementasikan sebagai sistem pendukung keputusan dalam layanan kesehatan digital dan skrining dini berbasis data. Oleh karena itu, penelitian selanjutnya disarankan untuk menguji model pada dataset populasi lokal dan mengintegrasikannya dengan sistem kesehatan berbasis Internet of Things guna meningkatkan akurasi prediksi secara real-time dan berkelanjutan.

DAFTAR PUSTAKA

- Agrawal, N., Choubey, S., Choubey, A., et al. (2024). Predicting early-stage diabetes risk: A machine learning approach. *Journal of Data Science*, 2(1). <https://doi.org/10.26634/jds.2.1.20356>
- Ahmed, M. R., & Reddy, G. S. V. S. (2023). Type 2 diabetes mellitus: Early detection using machine learning classification. *International Journal of Advanced Computer Science and Applications*, 14(6). <https://doi.org/10.14569/ijacsa.2023.01406127>
- Akmeşe, Ö. F. (2022). Diagnosing diabetes with machine learning techniques. *Hitit Journal of Science and Engineering*, 9(3). <https://doi.org/10.17350/hjse19030000250>
- Almadhoun, M. B., & Aboobaider, B. M. (2025). Optimizing feature selection and machine learning algorithms for early detection of prediabetes risk: Comparative study. *JMIR Bioinformatics and Biotechnology*. <https://doi.org/10.2196/70621>
- Al-Mousa, A., AlKhdour, L., Bishawi, H., et al. (2023). Multiclass diabetes detection using random forest classification. In *Proceedings of the IEEE Conference on Electrical Engineering and Information Technology*. <https://doi.org/10.1109/jeeit58638.2023.10185679>
- Aulia, H., Wibowo, A., & Sutrisno. (2025). Integration of Random Forest, ADASYN, and SHAP for diabetes prediction and interpretation. *Scientific Journal of Informatics*, 12(2). <https://doi.org/10.15294/sji.v12i2.24314>
- Chaithanya, B. N. (2023). RFFE – Random Forest fuzzy entropy for the classification of diabetes mellitus. *AIMS Public Health*, 10(3). <https://doi.org/10.3934/publichealth.2023030>
- Choudhary, D., Gupta, P. K., & Gupta, S. (2024). Early phase detection of diabetes mellitus using machine learning. In *Communications in Computer and Information Science*. Springer. https://doi.org/10.1007/978-3-031-56703-2_10
- Dai, Z., Shen, T., & Zhang, Y. (2023). Risk prediction of diabetes based on Spark and Random Forest algorithm. In *Proceedings of the IEEE International Conference on Big Data Analytics*. <https://doi.org/10.1109/EBDA56825.2023.10090801>
- Duvvuri, S. K. (2025). Predictive modelling and IoT-based early intervention for diabetes mellitus in East and West Godavari districts using clinical big data. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. <https://doi.org/10.32628/cseit25111695>
- Furtado, L. E. T. A. (2023). A predictive model for early detection of diabetes mellitus using machine learning (Dissertation). <https://doi.org/10.58694/20.500.12479/1592>
- Gündoğdu, S. (2023). Efficient prediction of early-stage diabetes using XGBoost classifier with Random Forest feature selection technique. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-023-15165-8>
- Hakeel, M. (2025). Machine learning-based diabetes prediction app using Random Forest algorithm. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1). <https://doi.org/10.36040/jati.v9i1.12654>
- Hossen, M. K., Tabassum, A., & Ferdaus, J. (2023). Machine learning-based diabetic risk prediction model for early detection. *International Journal of Scientific Research in Science and Technology*. <https://doi.org/10.32628/ijsrst5231057>

- Iskandar, R. F. N., Gutama, D. H., Wijaya, D. P., et al. (2024). Klasifikasi menggunakan metode Random Forest untuk awal deteksi diabetes melitus tipe 2. *Jurnal Teknik Industri Terintegrasi*, 7(3). <https://doi.org/10.31004/jutin.v7i3.26916>
- Kumar, R., Gupta, M., Tamak, N., et al. (2023). Predictive modeling for early detection of diabetes using machine learning approach. In *IEEE International Conference Proceedings*. <https://doi.org/10.1109/icaiccit60255.2023.10465892>
- Nancy, ., Sil, S., Rani, R., et al. (2023). Unleashing the power of machine learning: Advancing early prediction and analysis of diabetes mellitus. In *IEEE World Conference Proceedings*. <https://doi.org/10.1109/wconf58270.2023.10235242>
- Nisa, S. K., Barata, M. A., & Yuwita, P. E. (2025). Optimization of Random Forest algorithm with SMOTE method to improve the accuracy of early diabetes prediction. *Scientific Journal of Informatics*, 12(3). <https://doi.org/10.15294/sji.v12i3.22986>
- Raman, R., Kumar, V., & Pillai, B. G., et al. (2024). Uncovering risk factors for diabetes mellitus using data mining classification techniques. In *IEEE International Conference Proceedings*. <https://doi.org/10.1109/icdsis61070.2024.10594402>
- Reddy, G. S. V. S., & Ahmed, M. R. (2023). Exploring the effectiveness of machine learning algorithms for early detection of type-2 diabetes mellitus. *Measurement: Sensors*. <https://doi.org/10.1016/j.measen.2023.100983>
- Sabariah, M. K., Hanifa, S. T. A., & Sa'adah, M. T. S. (2014). Early detection of type II diabetes mellitus with Random Forest and classification and regression tree (CART). In *IEEE International Conference Proceedings*. <https://doi.org/10.1109/ICAICTA.2014.7005947>
- Saeed, A., Ajmal, H. M. S., & Khan, M. U. A., et al. (2024). Enhancing early detection of diabetes using machine learning algorithms. In *IEEE Conference Proceedings*. <https://doi.org/10.1109/iciestr60916.2024.10798412>
- Silva, J. R. N. O., Hirata, M., Carvalho, M., et al. (2024). Explorando a aplicação de modelos de aprendizagem de máquina na análise de suspeita de diabetes. *Journal of Applied Data Science*, 6(5–6). <https://doi.org/10.29327/2385054.6.5-6>
- Sukamto, T. F., Prameswary, C. L., Royadi, D., et al. (2025). Diabetes disease prediction on unbalanced data using SMOTE-Tomek links and Random Forest algorithm. *G-Tech*, 9(3). <https://doi.org/10.70609/g-tech.v9i3.7164>
- Wang, S. (2024). Diabetes prediction using Random Forest in healthcare. *Highlights in Science, Engineering and Technology*. <https://doi.org/10.54097/5ndh9a05>
- Warsi, M. F., Gupta, S. N., Mohan, J., et al. (2025). A novel approach for early detection of type 2 diabetes with advanced ML algorithms based on SFS and PCA. In *Advances in Intelligent Systems*. <https://doi.org/10.4018/979-8-3693-6864-0.ch004>
- Wulandari, A., & Anggreainy, M. S. (2024). Implementation of Random Forest model with hyperparameter tuning for diabetes mellitus classification. In *IEEE International Conference Proceedings*. <https://doi.org/10.1109/ic2ie63342.2024.10747821>