JISTech Journal of Information Systems and Technology
DOI : 10.71234/jistech.v2i2.104
e-ISSN : 3064-0172
Vol. 2, No. 2, Desember 2025, pp 106-111

Analisis Kinerja Algoritma Random Forest untuk Klasifikasi Penyakit Diabetes di Puskesmas Wundulako

Nurul Aisyah Fitri[1*], Kharis Syaban[2], Nisa Miftachurohmah
[123]*Ilmu Komputer, Universitas Sembilanbelas November Kolaka, Indonesia*
[1]nafitri721@gmail.com, [2]k.syaban@gmail.com, [3]nisa.informatics@gmail.com

Abstract

Diabetes is a chronic disease with a continuously increasing prevalence and requires early detection to prevent more serious complications. Primary healthcare centers play an important role in the early identification of patients at risk of diabetes. However, manual diabetes classification processes may lead to delays and inaccuracies in decision-making. Therefore, this study aims to analyze the performance of the Random Forest algorithm in classifying diabetes using patients' medical record data at Wundulako Primary Healthcare Center. The dataset consists of 227 patient records with several predictor variables, including age, body mass index, random blood glucose, blood pressure, and other health-related attributes. The classification process was conducted using the Random Forest algorithm with a 5-Fold Cross Validation scheme to ensure model stability. The results show that the Random Forest algorithm achieved an average accuracy of 99.57%, with precision of 100%, recall of 98.75%, and an F1-score of 99.35%, as well as a very low error rate. Feature importance analysis indicates that random blood glucose is the most influential variable in determining diabetes classification. Based on these results, the Random Forest algorithm demonstrates excellent performance and has strong potential to be implemented as a decision support system for early diabetes detection in primary healthcare services.

Keywords: *Random Forest; Diabetes Classification; Data Mining; Decision Support System*

Abstrak

Penyakit diabetes merupakan salah satu penyakit kronis dengan prevalensi yang terus meningkat dan memerlukan deteksi dini untuk mencegah komplikasi yang lebih serius. Puskesmas sebagai fasilitas layanan kesehatan tingkat pertama memiliki peran penting dalam melakukan identifikasi awal terhadap pasien berisiko diabetes. Namun, proses klasifikasi penyakit diabetes yang masih dilakukan secara manual berpotensi menimbulkan keterlambatan dan ketidaktepatan dalam pengambilan keputusan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kinerja algoritma Random Forest dalam mengklasifikasikan penyakit diabetes menggunakan data rekam medis pasien di Puskesmas Wundulako. Data yang digunakan berjumlah 227 data pasien dengan beberapa variabel prediktor, antara lain usia, indeks massa tubuh, gula darah sewaktu, tekanan darah, dan variabel kesehatan lainnya. Proses klasifikasi dilakukan menggunakan algoritma Random Forest dengan skema pengujian 5-Fold Cross Validation untuk memastikan kestabilan model. Hasil penelitian menunjukkan bahwa algoritma Random Forest menghasilkan akurasi rata-rata sebesar 99.57% dengan nilai precision 100%, recall 98.75%, dan F1-score 99.35%, serta error rate yang sangat rendah. Analisis feature importance menunjukkan bahwa gula darah sewaktu merupakan variabel paling dominan dalam menentukan klasifikasi diabetes. Berdasarkan hasil tersebut, algoritma Random Forest terbukti memiliki kinerja yang sangat baik dan berpotensi digunakan sebagai sistem pendukung keputusan dalam deteksi dini penyakit diabetes di tingkat pelayanan kesehatan dasar.

Kata Kunci: Random Forest; Klasifikasi Diabetes; Data Mining; Sistem Pendukung Keputusan

Published Online 31-12-2025

I. PENDAHULUAN

Penyakit diabetes merupakan salah satu penyakit kronis yang prevalensinya terus meningkat secara global maupun nasional. Di Indonesia, diabetes menjadi penyebab utama berbagai komplikasi kesehatan serius, seperti penyakit kardiovaskular, gagal ginjal, dan gangguan penglihatan. Salah satu faktor utama yang memperparah dampak diabetes adalah keterlambatan dalam proses diagnosis, sehingga pasien tidak memperoleh penanganan sejak dini.

Puskesmas Wundulako sebagai fasilitas layanan kesehatan tingkat pertama di Kabupaten Kolaka menghadapi peningkatan jumlah pasien dengan risiko diabetes dalam beberapa tahun terakhir. Proses klasifikasi status diabetes yang masih bergantung pada pemeriksaan manual berpotensi menimbulkan keterlambatan dan ketidaktepatan dalam pengambilan keputusan medis. Oleh karena itu, diperlukan pendekatan berbasis teknologi yang mampu membantu tenaga kesehatan dalam mengklasifikasikan status diabetes secara cepat dan akurat.

Data mining, khususnya teknik klasifikasi dalam machine learning, telah banyak diterapkan dalam bidang kesehatan untuk mendukung pengambilan keputusan klinis. Berbagai algoritma seperti C4.5, Naïve Bayes, dan K-Nearest Neighbor telah digunakan dalam penelitian sebelumnya, namun masih menunjukkan keterbatasan dari sisi akurasi dan stabilitas model. Algoritma Random Forest dikenal memiliki keunggulan dalam menangani data berdimensi tinggi, mengurangi risiko overfitting, serta memberikan performa klasifikasi yang tinggi.

Penelitian terdahulu menunjukkan konsistensi kinerja algoritma Random Forest pada berbagai domain. [1]membuktikan Random Forest tetap stabil meskipun dikombinasikan dengan teknik optimasi, namun peningkatan akurasi tidak selalu signifikan. [2]menerapkan Random Forest pada klasifikasi kelayakan kredit koperasi dan memperoleh akurasi tinggi dalam membedakan debitur bermasalah. [3]menunjukkan Random Forest lebih unggul dan stabil dibandingkan algoritma klasifikasi lain pada data berdimensi kompleks. [4]membuktikan Random Forest menghasilkan performa terbaik pada analisis perilaku finansial. [5]menekankan pentingnya feature importance dalam meningkatkan interpretabilitas dan akurasi model Random Forest. [6]menyatakan Random Forest efektif menangani data skala besar dan heterogen dengan generalisasi yang baik. [7]menunjukkan penerapan Random Forest pada data kesehatan mampu meningkatkan akurasi dibandingkan single classifier konvensional. [8]membuktikan Random Forest memiliki sensitivitas dan spesifisitas tinggi dalam prediksi penyakit berbasis data medis. [9]menunjukkan kombinasi preprocessing data dan Random Forest meningkatkan performa klasifikasi secara signifikan. [10]menyimpulkan Random Forest lebih konsisten dibandingkan SVM dan Naïve Bayes pada kasus kesehatan. [11]menegaskan keunggulan Random Forest dalam prediksi diabetes setelah preprocessing dan seleksi fitur.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk menganalisis kinerja algoritma Random Forest dalam mengklasifikasikan penyakit diabetes menggunakan data rekam medis pasien di Puskesmas Wundulako. Analisis kinerja dilakukan melalui evaluasi metrik klasifikasi untuk menilai tingkat akurasi dan keandalan model dalam mendukung deteksi dini penyakit diabetes.

II. METODE PENELITIAN

A. Desain dan Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining pada tugas klasifikasi untuk menentukan status diabetes dan non-diabetes berdasarkan data pasien di Puskesmas Wundulako. Alur penelitian mencakup: pengumpulan data, prapemrosesan, pembentukan model, validasi, serta evaluasi kinerja model.

B. Sumber Data dan Variabel

Data yang digunakan berasal dari rekam medis pasien Puskesmas Wundulako sebanyak 227 data pasien. Variabel prediktor meliputi: usia, tinggi badan, berat badan, lingkar perut, indeks massa tubuh (IMT), gula darah sewaktu, asam urat, dan tekanan darah. Variabel target adalah status diabetes (kelas).

C. Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan kualitas data sebelum pemodelan, meliputi:

a. Pemeriksaan kelengkapan (*missing value*) dan konsistensi format data.
b. Penyesuaian tipe data (numerik/kategori) agar sesuai untuk pemodelan.
c. (Opsional, bila diterapkan di skripsi) penyeimbangan kelas atau penanganan outlier bila ditemukan ketimpangan distribusi kelas yang ekstrem.

D. Pembentukan Model Random Forest

Random Forest merupakan metode *ensemble* yang membangun banyak pohon keputusan, lalu menggabungkan prediksi tiap pohon menggunakan voting mayoritas (untuk klasifikasi). Untuk setiap

pohon, data dilatih menggunakan teknik *bootstrap sampling* dan pemilihan subset fitur secara acak pada setiap node split.

Rumus prediksi Random Forest (klasifikasi – majority voting):
Misalkan ada (*T*) pohon keputusan, dan pohon ke-*t* menghasilkan prediksi kelas $h_t(x)$ untuk data *x*. Maka prediksi akhir Random Forest adalah:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\} \quad (1)$$

atau ekuivalen dengan:

$$\hat{y} = \arg\max_{c \in C} \sum_{t=1}^{T} \mathbb{I}(h_t(x) = c) \quad (2)$$

Kriteria pemilihan split (Gini Index) pada node *t*
Jika proporsi kelas ke-*j* pada node (t) adalah $p(j|t)$, maka:

$$Gini(t) = 1 - \sum_{j=1}^{K} p(j|t)^2 \quad (3)$$

dengan (K) adalah jumlah kelas (pada penelitian ini (K=2)).

E. Validasi Model (5-Fold Cross Validation)

Untuk menguji kestabilan model, digunakan 5-Fold Cross Validation. Data dibagi menjadi 5 bagian; pada setiap iterasi, 4 fold menjadi data latih dan 1 fold menjadi data uji, lalu hasil dari semua fold dirata-ratakan sebagai performa akhir.

F. Evaluasi Kinerja

Kinerja model dievaluasi menggunakan confusion matrix dan metrik berikut[12].

- Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

- Precision

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

- Recall

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

- F1-Score

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

Selain itu, penelitian menggunakan AUC untuk melihat kemampuan pemisahan kelas, serta (jika dicantumkan di hasil) analisis feature importance untuk mengidentifikasi atribut paling berpengaruh terhadap klasifikasi.

III. HASIL DAN PEMBAHASAN

A. Hasil Prapemrosesan Data

Tahap prapemrosesan menghasilkan dataset yang siap digunakan untuk pemodelan klasifikasi. Data rekam medis pasien Puskesmas Wundulako yang berjumlah 227 data dinyatakan lengkap dan konsisten setelah dilakukan pemeriksaan kelengkapan dan penyesuaian format. Seluruh variabel prediktor, yaitu usia, tinggi badan, berat badan, lingkar perut, indeks massa tubuh (IMT), gula darah sewaktu, asam urat, dan tekanan darah, dapat digunakan secara optimal dalam proses pembentukan model Random Forest. Tahapan ini memastikan bahwa proses klasifikasi tidak dipengaruhi oleh kesalahan data, sehingga hasil evaluasi model dapat merepresentasikan kinerja algoritma secara objektif.

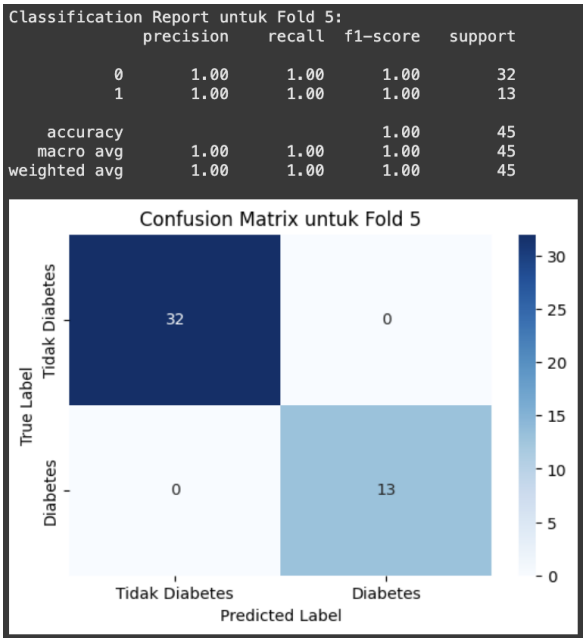
B. Hasil Pembentukan Model Random Forest

Model klasifikasi dibangun menggunakan algoritma Random Forest dengan pendekatan ensemble pohon keputusan. Setiap pohon dibentuk melalui proses bootstrap sampling dan pemilihan subset fitur secara acak pada setiap node split. Pendekatan ini memungkinkan model untuk mengurangi risiko overfitting dan meningkatkan kemampuan generalisasi terhadap data uji. Proses voting mayoritas digunakan untuk menentukan kelas akhir pada setiap data pasien, yaitu diabetes atau non-diabetes.

C. Hasil Validasi Menggunakan 5-Fold Cross Validation

Pengujian model dilakukan menggunakan 5-Fold Cross Validation untuk memastikan kestabilan performa klasifikasi. Pada setiap fold, data dibagi menjadi data latih dan data uji secara bergantian. Hasil pengujian menunjukkan bahwa model Random Forest memberikan performa yang sangat konsisten pada seluruh fold. Nilai *Area Under Curve* (AUC) yang diperoleh pada setiap fold adalah 1.00, yang menunjukkan kemampuan model yang sangat baik dalam memisahkan kelas diabetes dan non-diabetes pada berbagai skenario pembagian data.

Salah satu hasil pengujian model Random Forest ditunjukkan pada fold ke-5 dari proses 5-Fold Cross Validation. Pada fold ini, model diuji menggunakan data uji yang berbeda dari fold lainnya untuk memastikan kestabilan kinerja klasifikasi. Hasil klasifikasi pada fold ke-5 direpresentasikan dalam bentuk confusion matrix yang ditunjukkan pada Gambar 1.



Gambar 1. Hasil *Confusion Matrix* Fold 5

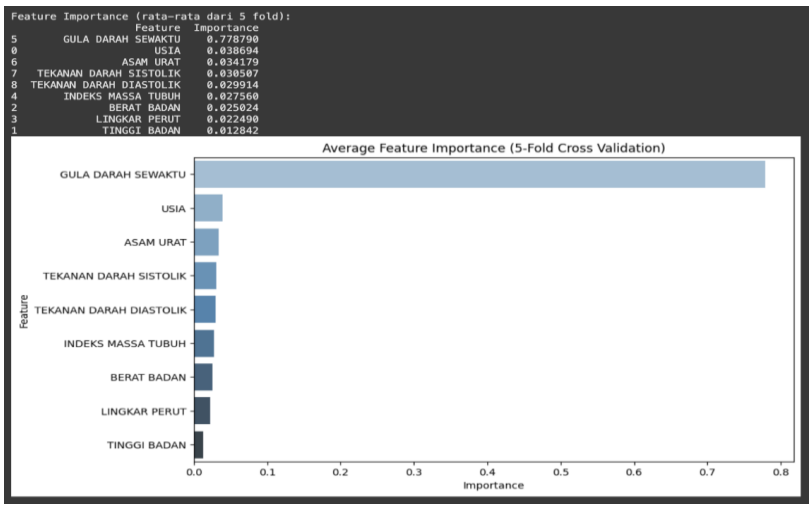
Berdasarkan Gambar 1, dapat dilihat bahwa sebagian besar data pasien berhasil diklasifikasikan dengan benar ke dalam kelas diabetes dan non-diabetes. Jumlah kesalahan klasifikasi pada fold ke-5 relatif kecil, yang menunjukkan bahwa model Random Forest memiliki kemampuan generalisasi yang baik terhadap data uji. Hasil ini konsisten dengan nilai akurasi, precision, dan recall yang tinggi pada fold lainnya.

D. Hasil Evaluasi Kinerja Model

Berdasarkan hasil pengujian keseluruhan, model Random Forest menghasilkan akurasi rata-rata sebesar 99.57%. Nilai precision mencapai 100%, yang menunjukkan bahwa seluruh data yang diprediksi sebagai diabetes benar-benar merupakan data diabetes. Nilai recall sebesar 98.75% menunjukkan bahwa hampir seluruh kasus diabetes berhasil terdeteksi oleh model. Selanjutnya, nilai F1-score sebesar 99.35% mencerminkan keseimbangan yang sangat baik antara precision dan recall.

Hasil confusion matrix menunjukkan bahwa dari total 920 data uji yang dihasilkan dari seluruh fold, hanya 4 data yang mengalami kesalahan klasifikasi. Dengan demikian, error rate yang dihasilkan hanya sebesar 0.43%, yang mengindikasikan tingkat kesalahan model sangat rendah. Temuan ini menegaskan bahwa algoritma Random Forest memiliki kinerja klasifikasi yang sangat baik pada data rekam medis Puskesmas Wundulako.

Selain evaluasi kinerja model, penelitian ini juga menganalisis kontribusi setiap variabel terhadap proses klasifikasi menggunakan feature importance pada algoritma Random Forest. Hasil analisis tersebut ditunjukkan pada Gambar 2.



Gambar 2. Hasil *Feature Importance*

Berdasarkan Gambar 2, variabel gula darah sewaktu memiliki kontribusi paling dominan dalam menentukan klasifikasi penyakit diabetes. Variabel indeks massa tubuh (IMT) dan tekanan darah juga menunjukkan pengaruh yang signifikan. Temuan ini sejalan dengan indikator medis yang umum digunakan dalam diagnosis diabetes, sehingga memperkuat validitas model dari sisi klinis.

5) Pembahasan dan Interpretasi Hasil

Tingginya nilai akurasi, precision, recall, dan F1-score menunjukkan bahwa Random Forest mampu memanfaatkan pola kompleks pada data rekam medis secara efektif. Penerapan teknik ensemble dan pemilihan fitur acak pada setiap pohon berkontribusi besar dalam meningkatkan stabilitas dan ketepatan klasifikasi. Selain itu, hasil AUC yang sempurna pada seluruh fold mengindikasikan bahwa model tidak hanya akurat pada satu ambang batas tertentu, tetapi juga memiliki kemampuan pemisahan kelas yang sangat kuat.

Analisis feature importance dalam skripsi menunjukkan bahwa gula darah sewaktu merupakan atribut paling berpengaruh dalam menentukan klasifikasi diabetes, diikuti oleh indeks massa tubuh (IMT) dan tekanan darah. Temuan ini selaras dengan indikator medis yang umum digunakan dalam praktik klinis, sehingga memperkuat relevansi model dari sisi kesehatan. Dengan demikian, hasil penelitian ini tidak hanya unggul secara teknis, tetapi juga memiliki validitas klinis yang baik.

Secara keseluruhan, hasil penelitian membuktikan bahwa algoritma Random Forest sangat layak digunakan sebagai sistem pendukung keputusan untuk membantu tenaga kesehatan dalam melakukan deteksi dini penyakit diabetes di tingkat puskesmas. Implementasi model ini diharapkan dapat meningkatkan ketepatan dan kecepatan pengambilan keputusan medis, khususnya pada layanan kesehatan dasar seperti Puskesmas Wundulako.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest memiliki kinerja yang sangat baik dalam mengklasifikasikan penyakit diabetes menggunakan data rekam medis pasien di Puskesmas Wundulako. Penerapan 5-Fold Cross Validation menunjukkan tingkat akurasi yang tinggi dan konsisten, dengan nilai precision, recall, dan F1-score yang sangat baik serta error rate yang rendah. Hasil evaluasi melalui confusion matrix menegaskan bahwa sebagian besar data pasien berhasil diklasifikasikan dengan benar, sementara analisis feature importance menunjukkan bahwa gula darah sewaktu, indeks massa tubuh, dan tekanan darah merupakan variabel yang paling berpengaruh dalam proses klasifikasi. Temuan ini tidak hanya membuktikan keunggulan Random Forest secara komputasional, tetapi juga menunjukkan kesesuaian model dengan indikator medis yang relevan, sehingga berpotensi dimanfaatkan sebagai sistem pendukung keputusan untuk membantu deteksi dini penyakit diabetes di tingkat pelayanan kesehatan dasar.

V. DAFTAR PUSTAKA

[1] L. I. A. Latif, A. A. Bakar, Z. A. Othman, M. S. A. Rais, and M. Berahim, "The Random Forest Algorithm for Modelling the Overspending Behaviour of Malaysian Households Income Class," *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 14, no. 1, pp. 40–60, 2025, doi: 10.17576/apjitm-2025-1401-03.

[2] W. A. Alansari and M. Mohd, "A Comparative Analysis of Machine Learning Algorithms for Diabetes Prediction," *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 13, no. 2, pp. 253–265, 2024, doi: 10.17576/apjitm-2024-1302-07.

[3] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021.

[4] Y. Religia, A. Nugroho, and W. Hadikristanto, "Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, 2021.

[5] Y. Iza Fajarendra, Y. Rizal Fauzan, and S. 'Uyun, "Klasifikasi Citra Eurosat Menggunakan Algoritma Knn, Decision Tree Dan Random Forest," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7754–7761, 2024, doi: 10.36040/jati.v8i4.10458.

[6] S. F. Jannah, R. Astuti, and F. M. Basysyar, "Implementasi Algoritma Random Forest Pada Aplikasi Picsart Berdasarkan Respon Pengguna," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 274–283, 2024, doi: 10.36040/jati.v8i1.8329.

[7] G. Surono and N. N. Pusparini, "Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera," *J. Technol. Inf.*, vol. 6, no. 1, pp. 7–14, 2020.

[8] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification," *Techno.COM*, vol. 20, no. 3, pp. 400–410, 2021.

[9] A. Saepudin, A. Faqih, and G. Dwilestari, "Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee," *J. Tekno Kompak*, vol. 18, no. 1, p. 178, 2024, doi: 10.33365/jtk.v18i1.3764.

[10] K. Abdi, A. Warjaya, I. Muthmainnah, and P. H. Pahutar, "Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 2, pp. 81–88, 2024, doi: 10.54082/jiki.81.

[11] S. Sudiadi and M. Meiriyama, "Penerapan Algoritma Random Forest untuk Klasifikasi Jenis Daun Herbal," *JITTER J. Ilm. Teknol. dan Komput.*, vol. 4, no. 2, p. 1700, 2023, doi: 10.24843/jtrti.2023.v04.i02.p05.

[12] Rosihan, F. Tempola, M. N. Sutoyo, and C. E. Gunawan, "Improving System Accuracy by Modifying the Transfer Learning Architecture for Detecting Clove Maturity Levels," *J. Adv. Inf. Technol.*, vol. 15, no. 3, pp. 407–413, 2024, doi: 10.12720/jait.15.3.407-413.