

Literature Review : Penerapan Algoritma Random Forest untuk Klasifikasi Penyakit Diabetes

Rian Maulana^{1*}, Moh Fachri Hasan², Farriz Raehan³, Muhammad Ramzy⁴

¹⁻⁴Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ¹rianmaulana120602@gmail.com, ²fahrihasan2277@gmail.com,

³farrizraehan@gmail.com, ⁴muhhammadramzy27@gmail.com

(* : coressponding author)

Abstrak- Diabetes adalah penyakit kronis yang berdampak signifikan pada kesehatan global, ditandai dengan peningkatan kadar gula darah akibat gangguan insulin. Deteksi dini dan klasifikasi yang akurat sangat penting untuk mendukung diagnosis dan penanganan diabetes. Penelitian ini bertujuan mengevaluasi algoritma Random Forest (RF), yang telah terbukti efektif dalam klasifikasi data medis, khususnya untuk deteksi diabetes. Algoritma RF bekerja melalui metode pembelajaran kelompok, menggabungkan berbagai pohon keputusan untuk menghasilkan prediksi yang lebih akurat. Dengan menggunakan pendekatan rapid review, studi ini meninjau literatur yang relevan guna memahami efektivitas RF dalam klasifikasi diabetes. Hasil analisis menunjukkan bahwa RF mencapai akurasi tinggi hingga 98% dengan metode evaluasi Area Under Curve (AUC) mencapai 100%, yang menunjukkan kemampuan klasifikasi yang sangat baik. Selain itu, penerapan normalisasi data, khususnya min-max normalization, meningkatkan performa model. Meskipun memiliki beberapa keterbatasan, seperti kompleksitas dan keterbatasan interpretasi, RF menunjukkan potensi signifikan dalam pengembangan sistem deteksi dini berbasis web untuk diabetes, serta potensi aplikasi lebih lanjut dalam klasifikasi penyakit lainnya.

Kata Kunci : Diabetes, Random Forest, Klasifikasi, Deteksi Dini, Rapid Review

Abstract- Diabetes is a chronic disease with a significant impact on global health, characterized by elevated blood sugar levels due to insulin dysfunction. Early detection and accurate classification are essential to support the diagnosis and management of diabetes. This study aims to evaluate the Random Forest (RF) algorithm, which has proven effective in medical data classification, specifically for diabetes detection. The RF algorithm utilizes an ensemble learning method by combining multiple decision trees to achieve more accurate predictions. Using a rapid review approach, this study reviews relevant literature to understand the effectiveness of RF in diabetes classification. Analysis results show that RF achieves a high accuracy of up to 98%, with an Area Under Curve (AUC) evaluation of 100%, indicating excellent classification capability. Additionally, data normalization methods, particularly min-max normalization, further improve model performance. Although there are some limitations, such as complexity and interpretability issues, RF demonstrates significant potential in the development of web-based early detection systems for diabetes, along with further potential applications in other disease classifications.

Keywords: Diabetes, Random Forest, Classification, Early Detection, Rapid Review

1. PENDAHULUAN

Diabetes merupakan salah satu penyakit kronis yang berdampak besar pada kesehatan global. Penyakit ini ditandai oleh tingginya kadar gula dalam darah yang disebabkan oleh gangguan produksi atau fungsi insulin. Jika tidak ditangani dengan baik, diabetes dapat menyebabkan berbagai komplikasi serius, seperti penyakit jantung, kerusakan ginjal, dan gangguan penglihatan, yang akhirnya dapat mempengaruhi kualitas hidup penderitanya. Oleh karena itu, deteksi dini dan klasifikasi diabetes yang akurat menjadi langkah penting untuk membantu proses diagnosis dan penanganan yang tepat bagi pasien.

Dalam beberapa tahun terakhir, perkembangan teknologi informasi dan ilmu data memungkinkan penggunaan algoritma pembelajaran mesin untuk mengolah dan menganalisis data medis. Algoritma ini membantu dalam klasifikasi penyakit secara otomatis dan akurat dengan memanfaatkan sejumlah data medis yang relevan. Salah satu algoritma yang terbukti efektif dalam klasifikasi data adalah algoritma Random Forest. Algoritma Random Forest merupakan teknik berbasis ensemble yang bekerja dengan menggabungkan hasil dari banyak pohon keputusan untuk mendapatkan prediksi yang lebih andal dan tahan terhadap variasi data.

Penggunaan Random Forest dalam klasifikasi penyakit diabetes bertujuan untuk membantu menentukan apakah seseorang berpotensi mengidap diabetes atau tidak, berdasarkan analisis beberapa faktor, seperti kadar glukosa, tekanan darah, usia, dan indeks massa tubuh. Dengan sifatnya yang mampu menangani data beragam dan akurasi yang tinggi, algoritma ini menjadi pilihan yang menjanjikan dalam meningkatkan efisiensi dan akurasi klasifikasi penyakit.

Penelitian ini bertujuan untuk mengevaluasi keakuratan dan efektivitas algoritma Random Forest dalam melakukan klasifikasi penyakit diabetes. Dengan menguji algoritma ini pada dataset yang relevan, penelitian diharapkan dapat memberikan kontribusi signifikan dalam bidang kesehatan, khususnya dalam pengembangan sistem deteksi dini yang akurat untuk penyakit diabetes, sehingga dapat mendukung proses pengambilan keputusan medis yang lebih cepat dan tepat.

2. METODE PENELITIAN

2.1 *Rapid Review*

Rapid review adalah metode yang dipercepat untuk melakukan literature review. Proses ini biasanya mengurangi cakupan atau mempercepat langkah-langkah tertentu dalam metode systematic review agar dapat segera memberikan informasi atau panduan kebijakan.

2.2 *Random Forest*

Random Forest (RF) adalah pendekatan yang didasarkan pada konsep pembelajaran kelompok, dimana berbagai pohon keputusan dibangun dan digabungkan untuk menghasilkan prediksi akhir (A Setiawan et al., 2024).

2.3 *Content Analysis*

Analisis konten adalah metode yang menekankan pada identifikasi tema, kata kunci, atau pola yang sering muncul dalam literatur.

2.4 *Comparative Analysis*

Comparative analysis membandingkan dan mengontraskan berbagai penelitian untuk menemukan persamaan, perbedaan, atau kontradiksi di antara studi-studi yang ada.

3. ANALISIS DAN PEMBAHASAN

Pada bagian ini berisi temuan, hasil serta pembahasan dari topik penelitian.

Tabel 1. Hasil Temuan Artikel Relevan

No	Author / Tahun	Metode Penelitian	Tujuan Penelitian	Hasil yang didapat	Nilai Akurasi
1.	(Andri Setiawan et al., 2024)	Kuantitatif	Untuk mengembangkan dan menerapkan model klasifikasi serta mengeksplorasi epidemiologi diabetes, identifikasi faktor risiko, strategi pencegahan, manajemen diabetes, dan upaya untuk mengurangi dampak penyakit	Penelitian ini memanfaatkan data yang dibagi dalam dua subset, dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian model. Hasil evaluasi menunjukkan bahwa algoritma Random Forest berhasil mencapai tingkat akurasi tinggi, yaitu sebesar 98%, dalam mengidentifikasi risiko diabetes. Temuan ini	98%

			diabetes di masa depan.	mengindikasikan bahwa model memiliki performa yang kuat dalam membedakan kategori data secara akurat. Selain itu, nilai Area Under Curve (AUC) mencapai 100%, yang termasuk dalam kategori "Klasifikasi Unggul," memperlihatkan kehandalan model dalam membedakan kelas positif dan negatif dengan sangat baik.	
2.	(Gde Agung Brahma Suryanegara et al., 2021)	Kuantitatif	Penelitian ini berfokus pada pengembangan sistem deteksi penyakit diabetes yang mengklasifikasikan hasil ke dalam dua kategori: positif diabetes dan negatif diabetes. Ketiga model tersebut diterapkan menggunakan algoritma Random Forest. Tujuan dari penelitian ini adalah untuk mengevaluasi metode normalisasi mana yang paling efektif dalam meningkatkan kinerja algoritma Random Forest dalam mendeteksi dan mengklasifikasikan kasus diabetes.	Dari ketiga pendekatan yang diuji dalam penelitian ini, dua pendekatan yang menerapkan normalisasi data menunjukkan hasil yang lebih unggul dibandingkan dengan pendekatan yang tidak menggunakan normalisasi. Pendekatan yang menggabungkan Min-max normalization dengan algoritma Random Forest mencatatkan akurasi tertinggi sebesar 95,45%, lebih tinggi daripada pendekatan yang menggunakan Z-score normalization (95%) dan yang tidak menggunakan normalisasi (92%). Temuan ini menunjukkan bahwa penggunaan Min-max normalization lebih efektif dalam mengklasifikasikan data uji yang belum diketahui, serta secara signifikan meningkatkan performa algoritma Random Forest dalam	94.45%

				mendeteksi diabetes jika dibandingkan dengan pendekatan lainnya.	
3.	(Andri Triyono et al., 2021)	Kuantitatif	untuk mengembangkan model sistem prediksi untuk penyakit diabetes mellitus menggunakan teknik pembelajaran mesin dengan algoritma Random Forest dan Information Gain. Untuk membangun sistem model prediksi yang dapat mendiagnosis diabetes mellitus secara dini, guna membantu deteksi lebih cepat terhadap kondisi diabetes pada pasien.	Penggunaan metode Random Forest dengan criterion Information Gain terbukti menjadi pilihan yang optimal untuk prediksi dini diabetes. Proses dimulai dengan tahap preprocessing untuk mengatasi data yang tidak lengkap menggunakan teknik penanganan missing value. Model Random Forest dengan Information Gain berhasil mencapai akurasi sebesar 98,27%, presisi 97,69%, dan recall 98%. Dengan performa model yang sangat baik ini, sistem tersebut cocok sebagai langkah awal dalam mengidentifikasi apakah seseorang memiliki potensi mengidap diabetes mellitus atau tidak.	98.27%
4.	(Widya Apriliah et al., 2020)	Kuantitatif	Merancang model yang mampu memprediksi kemungkinan diabetes pada pasien dengan tingkat akurasi yang tinggi.	Penelitian ini bertujuan untuk mengidentifikasi algoritma klasifikasi yang paling efektif dalam memprediksi risiko diabetes. Uji coba dilakukan menggunakan dataset Diabetes Hospital di Sylhet, Bangladesh, yang diperoleh dari UCI repository. Hasilnya menunjukkan bahwa model yang dirancang memiliki tingkat akurasi sebesar 97,88%, dengan algoritma Random Forest memberikan	97.88%

				performa terbaik. Di masa mendatang, sistem berbasis algoritma klasifikasi machine learning ini dapat diperluas untuk membantu prediksi atau diagnosis penyakit lainnya.	
5.	(Reza Fauzan Nur Iskandar et al., 2024)	Kuantitatif	Membangun model klasifikasi untuk mendeteksi penyakit Diabetes Mellitus tipe 2 dengan menggunakan metode Random Forest (RF), berdasarkan dataset yang diperoleh dari UPTD Puskesmas Jatiroto.	Penelitian ini melibatkan tujuh variabel, yang terdiri dari enam fitur dan satu label yaitu "Riwayat PTM". Model Random Forest yang dibangun dalam penelitian ini mengoptimalkan empat parameter utama, yaitu "n_estimators", "criterion", "max_features", dan "min_samples_leaf". Sebagai langkah validasi, digunakan teknik cross-validation dengan 10 lipatan (K = 10), dan untuk menilai kinerja model, digunakan *confusion matrix*. Berdasarkan hasil evaluasi, model ini mencatatkan tingkat akurasi sebesar 97,48%, presisi 95%, *recall* 97%, dan skor F1 sebesar 96%.	97%

4. KESIMPULAN

Berdasarkan hasil dari sejumlah penelitian mengenai penggunaan algoritma Random Forest dalam klasifikasi risiko diabetes, algoritma ini telah terbukti sangat efektif dan akurat dalam mendeteksi serta mengklasifikasikan tingkat risiko diabetes. Dengan pembagian data 80:20 antara pelatihan dan pengujian, model ini berhasil mencapai tingkat akurasi tinggi hingga 98%, menunjukkan kemampuannya untuk melakukan klasifikasi data dengan sangat tepat. Nilai Area Under Curve (AUC) yang mencapai 100% juga menunjukkan kemampuan luar biasa model dalam membedakan kelas positif dan negatif, masuk dalam kategori "Klasifikasi Luar Biasa."

Penelitian-penelitian ini menunjukkan bahwa algoritma Random Forest memiliki potensi besar dalam mengidentifikasi pasien dengan risiko diabetes tinggi maupun rendah. Selain itu, model ini memperkuat sistem deteksi dini, memungkinkan proses diagnosis diabetes yang lebih cepat dan akurat. Dengan kemampuan tersebut, Random Forest menjadi salah satu metode yang sangat menjanjikan untuk dikembangkan sebagai sistem pendukung keputusan medis, yang berperan penting dalam diagnosis dan pencegahan diabetes dengan cara yang lebih efisien dan efektif.

REFERENCES

- Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *SISTEMASI*, 10(1), 163. <https://doi.org/10.32520/stmsi.v10i1.1129>
- Baihaqi, M. I., Syaripudin, A., & Nugroho, F. A. (2023). *Implementation Of The Random Forest Algorithm In Stock Price Predictions Based On Historical Data. 1.*
- Gde Agung Brahmuna Suryanegara, Adiwijaya, & Mahendra Dwifabri Purbolaksono. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 114–122. <https://doi.org/10.29207/resti.v5i1.2880>
- Hardoni, A., Rini, D. P., & Sukemi, S. (2021). Integrasi SMOTE pada Naive Bayes dan Logistic Regression Berbasis Particle Swarm Optimization untuk Prediksi Cacat Perangkat Lunak. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(1), 233. <https://doi.org/10.30865/mib.v5i1.2616>
- Iskandar, R. F. N., Gutama, D. H., Wijaya, D. P., & Danianti, D. (2024). Klasifikasi Menggunakan Metode Random Forest untuk Awal Deteksi Diabetes Melitus Tipe 2. *Jurnal Teknik Industri Terintegrasi*, 7(3), 1620–1626. <https://doi.org/10.31004/jutin.v7i3.26916>
- Mahendra, I. B. S., Widiari, T., Nugroho, F. A., & Sasongko, P. S. (2024). *Implementation Of Feature Selection Chi-Square To Improve The Accuracy Of The Classification Model Using The Random Forest Algorithm On Coronary Artery Disease. 9, 1–7.*
- Rizal, A., Bryliana, F. R., Aripin, K. N. A., Wardani, S. A., & Rosyani, P. (2023). PEMANFAATAN DATA MINING UNTUK PRAKIRAAN. *Jurnal Matematika*, 1(1).
- Rosyani, P., Saprudin, S., & Amalia, R. (2021). Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO). *Jurnal Sistem dan Teknologi Informasi (Justin)*, 9(2), 132. <https://doi.org/10.26418/justin.v9i2.44120>
- Saifudin, A., Nabillah, U. U., Yulianti, & Desyani, T. (2020). Bagging Technique to Reduce Misclassification in Coronary Heart Disease Prediction Based on Random Forest. *Journal of Physics: Conference Series*, 1477(3), 032009. <https://doi.org/10.1088/1742-6596/1477/3/032009>
- Setiawan, A., Nst, Z. H., Khairi, Z., & Efrizoni, L. (2024). KLASIFIKASI TINGKAT RISIKO DIABETES MENGGUNAKAN ALGORITMA. *Jurnal Informatika*, 7(2).
- Triyono, A., Trianto, R. B., & Arum, D. M. P. (2021). *EARLY DETECTION OF DIABETES MELLITUS USING RANDOM FOREST ALGORITHM. 1.*