

Systematic Literature Review Penerapan Gradient Boosting Untuk Klasifikasi Penyakit Diabetes Tipe 2

Bintang^{1*}, Agni Tri Pratiwi², Ahmad Barizi³, Muhammad Irfan Maulana⁴, Perani Rosyani⁵

¹⁻⁵Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan. Banten 15310, Indonesia

Email: ^{1*}bintang.personal10@gmail.com, ²agnipratiwi528@gmail.com,

³ahmadbarizi464@gmail.com, ⁴maul010803@gmail.com, ⁵dosen00837@unpam.ac.id

(* : coresponding author)

Abstrak—Penerapan algoritma *Gradient Boosting* dalam klasifikasi penyakit diabetes tipe 2 telah menunjukkan hasil yang signifikan dalam meningkatkan akurasi diagnosis. Penelitian ini mengkaji berbagai studi yang menggunakan *Gradient Boosting* untuk memprediksi diabetes tipe 2 dengan memanfaatkan dataset medis. Hasil menunjukkan bahwa algoritma ini mampu mencapai akurasi hingga 82%, lebih baik dibandingkan dengan metode lain seperti *Random Forest* dan *Naïve Bayes*. Dengan mengidentifikasi fitur-fitur penting yang berkontribusi terhadap perkembangan diabetes, *Gradient Boosting* tidak hanya meningkatkan akurasi tetapi juga memberikan wawasan berharga untuk pencegahan dan pengelolaan penyakit ini. Hasil penelitian ini mendemonstrasikan potensi machine learning dalam diagnosis dini diabetes, yang penting untuk pengelolaan kesehatan masyarakat.

Kata Kunci: Algoritma *Gradient Boosting*, Klasifikasi Penyakit Diabetes Tipe 2, *Machine Learning*

Abstract—*The application of the Gradient Boosting algorithm in the classification of type 2 diabetes has shown significant results in improving diagnostic accuracy. This research examines various studies that use Gradient Boosting to predict type 2 diabetes by utilizing medical datasets. Results indicate that this algorithm can achieve an accuracy of up to 82%, surpassing other methods such as Random Forest and Naïve Bayes. By identifying key features that contribute to the development of diabetes, Gradient Boosting not only enhances accuracy but also provides valuable insights for the prevention and management of this disease. The findings of this research demonstrate the potential of machine learning in early diabetes diagnosis, which is crucial for public health management.*

Keywords: *Gradient Boosting Algorithm, Type 2 Diabetes Disease Classification, Machine Learning*

1. PENDAHULUAN

Diabetes mellitus, khususnya diabetes tipe 2, merupakan penyakit kronis yang semakin meningkat prevalensinya di seluruh dunia. Menurut International Diabetes Federation (IDF), pada tahun 2021, diperkirakan terdapat 537 juta orang dewasa di seluruh dunia hidup dengan diabetes, dan angka ini diproyeksikan akan meningkat menjadi 629 juta pada tahun 2045 (IDF, 2021). Diabetes tipe 2 ditandai dengan resistensi insulin, suatu kondisi di mana tubuh tidak dapat menggunakan glukosa dengan baik, sehingga menyebabkan peningkatan kadar gula darah. Jika tidak ditangani dengan baik, diabetes tipe 2 dapat menyebabkan berbagai komplikasi serius, seperti penyakit jantung, stroke, gagal ginjal, dan kerusakan saraf.

Penelitian ini menggunakan metode *Systematic Literature Review* yang sistematis untuk mengidentifikasi, mengevaluasi, dan mensintesis semua bukti penelitian yang relevan terhadap pertanyaan penelitian tertentu. (Kitchenham et al., 2007). SLR dimulai dengan pertanyaan penelitian yang jelas dan terfokus. Pertanyaan ini harus diformulasikan dengan cermat untuk memastikan bahwa pencarian literatur relevan dan terarah. (Booth, A et al., 2005)

Gradient Boosting merupakan algoritma pembelajaran mesin yang dikenal dengan kemampuannya dalam menangani data kompleks dan tidak seimbang (Friedman, 2001). Algoritma ini bekerja dengan membangun serangkaian model lemah (weak learners) secara bertahap, dan kemudian menggabungkannya untuk membentuk model yang kuat (strong learner). *Gradient Boosting* telah terbukti efektif dalam berbagai aplikasi, termasuk klasifikasi penyakit, prediksi, dan deteksi anomali.

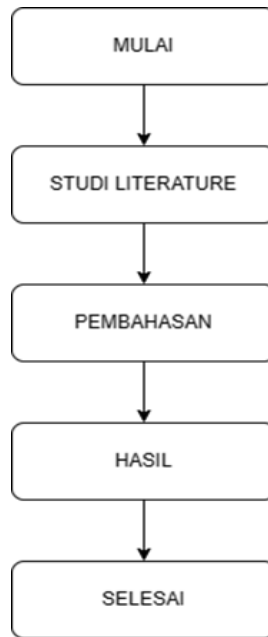
Dalam hal ini, *literature review* memiliki tujuan untuk mengeksplorasi penerapan algoritma *Gradient Boosting* untuk klasifikasi penyakit diabetes tipe 2. Penelitian ini akan menggunakan dataset klinis yang berisi informasi demografis, riwayat medis, dan hasil pemeriksaan laboratorium.

Algoritma *Gradient Boosting* akan dilatih dan dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil penelitian diharapkan dapat memberikan informasi tentang potensi *Gradient Boosting* sebagai alat bantu dalam deteksi dini dan manajemen penyakit diabetes tipe 2.

2. METODE PENELITIAN

2.1 *Systematic Literature Review*

Systematic Literature Review (SLR) merupakan teknik untuk mengidentifikasi, menilai, dan menginterpretasi seluruh temuan-temuan pada suatu topik penelitian, untuk menjawab pertanyaan penelitian (research question) yang telah ditetapkan sebelumnya. (RomiSatriaWahono.Net., 2016).



Gambar 1. Metode Penelitian

Gradient Boosting adalah algoritma machine learning yang menggabungkan beberapa model yang lemah (*weak learners*) untuk menciptakan model yang kuat (*strong learner*). Algoritma ini bekerja dengan cara berulang-ulang menambahkan model yang lemah ke dalam ensemble, dengan setiap model baru mencoba untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. (Friedman, J. H. 2001). Keuntungan utama *Gradient Boosting* adalah kemampuannya untuk menangani data kompleks dan non-linear, serta kemampuannya untuk mencapai akurasi yang tinggi. Algoritma ini juga relatif mudah untuk diimplementasikan dan dapat digunakan untuk berbagai tugas klasifikasi dan regresi. (James, G., Witten et al., 2009).

3. ANALISA DAN PEMBAHASAN

Tabel 1. Pembahasan

No	Nama Peneliti dan Tahun	Metode yang Dibahas	Tujuan Penelitian	Hasil yang didapat
1.	Boobala Muralidharan, Ahmed Yahya, Aswib Visveswar, Dhivakar,	Algoritma <i>Gradient Boosting</i>	Tujuan penelitian ini adalah untuk mengembangkan dan mengevaluasi model prediksi diabetes yang	Penelitian ini menunjukkan hasil klasifikasi menggunakan metode <i>gradient boosting</i> memiliki akurasi sebesar 80% pada data uji, Nilai sensitivitas

	NagarajR. (2023).		akurat dan handal menggunakan teknik ensemble learning Berbasis gradient boosting.untuk menunjukkan potensi model ini dalam mendeteksi dini diabetes, sehingga dapat meningkatkan hasil perawatan pasien dan mengurangi biaya kesehatan.	di angka 83%, nilai spesifisitas untuk hasil yangdidapat yaitu 85%.Selain itu, Hasil ini menunjukkan potensi model untuk membantudalam identifikasi dini pasien berisiko tinggi. menyoroti potensi peningkatan akurasi lebih lanjut dengan integrasi data tambahan seperti catatankesehatan elektronikdan data genetik di masa mendatang.
2.	J. Beschi Raja, R.Anitha, R. Sujatha, V. Roopa, S. SamPeter (2019).	Algoritma <i>Gradient Boosting</i>	Tujuan dari penelitian ini adalah Meneliti danmembandingkan kinerja Gradient Boosted Classifier, Neural Networks, dan Random Forest dalam memprediksi diabetes. Menunjukkan pentingnya deteksi dini diabetes dan peran machine learning dalam kesehatan.	Hasil penelitian ini menunjukkan bahwa Gradient Boosting lebih efektif dalam memprediksi diabetes dibandingkan dengan Neural Networks dan Random Forest. Penelitian tersebut menggunakan datasetPima Indians Diabetes dan mengevaluasi model berdasarkan beberapa metrik, termasuk recall, AUC (Area Under Curve), dan akurasi. GBC mencapai akurasi tertinggi sebesar 81,3%, sedangkan Random Forest hanya mencapai 79%
3.	Fatema Nusrat 1, Betül Uzbaş, and Ömer Kaan Baykan(2020).	Algoritma <i>Gradient Boosting</i>	Tujuan dari penelitian ini adalah Menunjukkan keandala klasifikasi Gradient Boosting Sebagai metode prediksi diabetes. Dan juga Menentukan algoritma mana yang paling efektif dalam memprediksi diabetes berdasarkan berbagaimetrik evaluasi seperti akurasi, F1 score, presisi, recall, Cohen's kappa, dan analisis kurva ROC.	Penelitian ini menyelidiki prediksi diabetes melitus menggunakan teknik machine learning, dengan fokus pada algoritma Gradient Boosting (GB) . Menggunakan dataset Pima Indian Diabetes yang terdiri dari 768 catatan, penelitian ini membandingkan kinerja tiga metode klasifikasi: Decision Tree (DT), Random Forest (RF), dan Gradient Boosting (GB). Algoritma GB mencapai akurasi tertinggi sebesar 76,30%, lebih baik dibandingkan dengan RF yang mencatat akurasi 74,50% dan DT dengan akurasi 73,69%. Ini menunjukkan bahwa GB lebih efektif dalam memprediksi diabetes dalam konteks dataset ini.

4.	Hilal KOÇAK, Gürcan ÇETİN (2023)	Algoritma <i>Gradient Boosting</i>	Tujuan penelitian ini adalah Menentukan algoritma boosting mana yang paling efektif dalam mendiagnosis diabetes berdasarkan tingkat akurasi. Menekankan pentingnya deteksi dini diabetes untuk mencegah masalah kesehatan yang serius.	Dalam penelitian Hasil penelitian ini menunjukkan bahwa algoritma CatBoost mencapai tingkat akurasi tertinggi sebesar 88% dalam mendiagnosis diabetes melitus, dibandingkan dengan algoritma boosting lainnya seperti Gradient Boosting, AdaBoost, LightGBM, dan XGBoost. Penelitian ini juga menegaskan kembali pentingnya deteksi dini diabetes untuk mencegah masalah kesehatan yang serius.
5.	Derara Duba Rufo ,TayeGirmaDe elee ,Achim Ibenthal, and Worku Gachena Negera.(2021)	Algoritma <i>Gradient Boosting</i>	Tujuan dari penelitian ini Menyelidiki efektivitas algoritma LightGBM dalam mendiagnosis diabetes mellitus (DM) di wilayah dengan sumber daya medis terbatas, seperti di Ethiopia. Menekankan pentingnya deteksi dini diabetes untuk meningkatkan hasil kesehatan dan Mengurangi beban penyakit.	Hasil penelitian ini menunjukkan bahwa algoritma LightGBM sangat efektif dalam mendiagnosis diabetes mellitus (DM) di wilayah dengan sumber daya medis terbatas, seperti di Ethiopia. Model LightGBM mencapai akurasi 98,1% dan sensitivitas 99,9%, mengungguli algoritma machine learning lainnya yang diuji.

4. KESIMPULAN

Berdasarkan *Systematic Literature Review* Penerapan Gradient Bosting Untuk Klasifikasi Penyakit Diabetes Tipe 2, dapat disimpulkan beberapa hal sebagai berikut:

1. Salah satu teknik pembelajaran mesin yang terbukti sangat efektif dalam menangani masalah klasifikasi, seperti identifikasi dan prediksi diabetes tipe 2, adalah Gradient Boosting
2. Dari tinjauan literatur ini, dapat disimpulkan bahwa algoritma Gradient Boosting memiliki potensi luar biasa untuk meningkatkan akurasi diagnosis diabetes tipe 2 secara signifikan
3. *Gradient Boosting* dapat memberikan tingkat akurasi yang sangat menjanjikan dan bahkan melebihi metode klasifikasi lain, misalnya *Random Forest* dan *Naïve Bayes*
4. Gradient Boosting mampu dengan mudah mengidentifikasi bagaimana faktor-faktor ini berinteraksi satu sama lain.
5. Studi ini menemukan bahwa algoritma Classifier Gradient Boosting memiliki akurasi 81% dalam memprediksi diabetes, lebih tinggi dari Random Forest (yang memiliki akurasi 79%)
6. Studi ini menunjukkan bahwa dalam hal klasifikasi diabetes, metode peningkatan dapat

mengungguli metode lainnya seperti Bagging dan Stacking. Karena diabetes semakin umum di masyarakat, penggunaan algoritma pembelajaran mesin seperti Gradient Boosting dapat membantu pencegahan dan pengelolaan penyakit ini.

7. Hasil penelitian menunjukkan bahwa algoritma Gradient Boosting memiliki akurasi di atas 90% dalam banyak kasus, menjadikannya pilihan yang bagus untuk klasifikasi penyakit diabetes.
8. Secara keseluruhan, penerapan algoritma ini dalam klasifikasi penyakit diabetes tipe 2 menunjukkan hasil yang menjanjikan dan memberikan dasar untuk pengembangan lebih lanjut dalam bidang kesehatan menggunakan teknologi pembelajaran mesin

Kesimpulan ini mencerminkan temuan utama dari tiga penelitian utama tentang penyakit diabetes. Dengan demikian, Gradient Boosting memiliki potensi besar untuk mengubah diagnosis diabetes tipe 2 dini. Algoritma ini dapat menjadi alat yang ampuh untuk meningkatkan pengelolaan kesehatan masyarakat, meningkatkan kualitas hidup penderita diabetes, dan mengurangi beban penyakit ini di seluruh dunia dengan mengatasi masalah yang ada.

REFERENCES

- Adiwijaya, A., & Al-Faraby, M. (2017). Comparative Analysis of Accuracy of Random Forest and Gradient Boosting Classifier for Diabetes Prediction. *Jurnal Wicida*.
- Salma Irena, & Febriastia. (2024). Klasifikasi Penyakit Diabetes Mellitus Tipe II Berbasis Machine Learning Menggunakan LightGBM. Skripsi, Universitas Lampung.
- Cendani, L. M., & Wibowo, A. (2022). Penerapan Machine Learning dalam Prediksi Diabetes Melitus dengan Algoritma Gradient Boosting Machine. *Jurnal Masyarakat Informatika*.
- International Diabetes Federation (IDF) (2021). *IDF Diabetes Atlas, 10th edition*. Brussels: International Diabetes Federation.
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report.
- RomiSatriaWahono.Net. (2016). *Systematic Literature Review: Pengantar, Tahapan dan Studi Kasus*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer Science & Business Media.
- Booth, A., & Sutton, J. (2005). *Systematic approaches to a literature review*. Sage Publications.)