

LITERATURE REVIEW: PENERAPAN GRADIENT BOOSTING UNTUK KLASIFIKASI PENYAKIT DIABETES TIPE 2

Emison Wonda¹, Mia Septiana Wambrauw¹, Renaldi Ferrari¹, Rizka Gifani Napitupulu¹,
Rosita Hermalinda Dwi Febrianti^{1*}

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

Email: ¹emisonwonda11@gmail.com, ²miasptianaaa12@gmail.com,

³renaldiferrari76@gmail.com, ⁴rizkanapitupulu005@gmail.com, ^{5*}rstahrmlnda@gmail.com

(* : coresponding author)

Abstrak– Diabetes mellitus tipe 2 merupakan kondisi metabolik yang semakin meningkat prevalensinya secara global. Klasifikasi yang akurat sangat penting untuk diagnosis dan pengelolaan yang tepat. Penelitian ini meninjau literatur mengenai penerapan algoritma Gradient Boosting, khususnya XGBoost dan LightGBM, dalam klasifikasi diabetes tipe 2. Tinjauan ini menunjukkan bahwa algoritma Gradient Boosting memiliki potensi besar dalam meningkatkan akurasi diagnosis dan prediksi risiko penyakit. Studi-studi yang dikaji menunjukkan bahwa algoritma ini mampu menangani data yang kompleks, mencapai tingkat akurasi yang tinggi, dan mengatasi masalah ketidakseimbangan kelas. Selain itu, optimasi parameter seperti hyperparameter tuning dapat meningkatkan kinerja model secara signifikan. Tinjauan ini menyoroti manfaat dan potensi algoritma Gradient Boosting dalam meningkatkan sistem kesehatan melalui deteksi dini dan pengelolaan diabetes tipe 2 yang lebih efektif.

Kata Kunci: Diabetes Melitus Tipe 2, Gradient Boosting, Xgboost, Lightgbm, Klasifikasi, Akurasi, Penyetelan Hyperparameter, Manajemen Kesehatan

Abstract–*Diabetes mellitus type 2 is a metabolic condition with a rising global prevalence. Accurate classification is crucial for proper diagnosis and management. This research reviews the literature on the application of Gradient Boosting algorithms, particularly XGBoost and LightGBM, in classifying type 2 diabetes. The review indicates that Gradient Boosting algorithms have significant potential in improving the accuracy of disease diagnosis and risk prediction. Studies examined demonstrate the ability of these algorithms to handle complex data, achieve high accuracy rates, and address class imbalance issues. Moreover, parameter optimization such as hyperparameter tuning can significantly enhance model performance. This review highlights the benefits and potential of Gradient Boosting algorithms in enhancing healthcare systems through early detection and more effective management of type 2 diabetes.*

Keywords: *Diabetes Mellitus Type 2, Gradient Boosting, Xgboost, Lightgbm, Classification, Accuracy, Hyperparameter Tuning, Healthcare Management*

1. PENDAHULUAN

Diabetes mellitus tipe 2 adalah kondisi metabolik yang ditandai oleh resistensi insulin dan peningkatan kadar glukosa darah. Penyakit ini menjadi salah satu masalah kesehatan global yang signifikan, dengan prevalensi yang terus meningkat seiring bertambahnya usia populasi. Klasifikasi diabetes tipe 2 penting untuk diagnosis dan pengelolaan yang tepat, karena dapat membantu dalam menentukan intervensi yang diperlukan untuk mencegah komplikasi lebih lanjut. Menurut penelitian, penggunaan algoritma pembelajaran mesin, seperti Gradient Boosting, telah menunjukkan hasil yang menjanjikan dalam klasifikasi diabetes tipe 2. Metode ini mampu menangani data yang kompleks dan memberikan akurasi yang tinggi dalam memprediksi kondisi diabetes pada pasien. Penelitian sebelumnya menunjukkan bahwa algoritma seperti Extreme Gradient Boosting (XGBoost) dan Light Gradient Boosting Machine (LightGBM) dapat meningkatkan efisiensi dalam mendeteksi diabetes, dengan hasil akurasi yang signifikan, seperti yang dilaporkan dalam beberapa studi yang mengkaji penerapan teknik ini dalam konteks klasifikasi penyakit diabetes. (Derara Duba Rufo, Taye Girma, Achim Ibenthal, & Worku Gachena Negara, 2021).

Penerapan algoritma Gradient Boosting dalam klasifikasi penyakit diabetes tipe 2 telah menjadi fokus penelitian yang signifikan, mengingat potensi metode ini dalam meningkatkan akurasi diagnosis dan prediksi risiko penyakit. Salah satu pendekatan yang banyak digunakan adalah

Extreme Gradient Boosting (XGBoost), yang telah terbukti efektif dalam berbagai studi. Misalnya, penelitian oleh (Kartina Diah Kusuma W., Memen Akbar, 2022) menunjukkan bahwa model prediksi risiko diabetes yang dibangun menggunakan XGBoost mencapai akurasi sebesar 98,71%, menandakan kemampuannya dalam menangani data terstruktur dengan baik. Selain itu, penelitian oleh (Ginjar Abdurrahman, Hardian Oktavianto, Mukti Sintawati., 2022) mengungkapkan bahwa penerapan hyperparameter tuning melalui Gridsearch dan Random Search dapat meningkatkan akurasi model XGBoost dari 75% menjadi 95%, menunjukkan pentingnya optimasi parameter dalam pengembangan model. Penelitian lain oleh (Iqbal Fathur Rahman., 2020) juga membandingkan XGBoost dengan metode lain seperti Support Vector Machine (SVM) dan Multilayer Perceptron (MLP), di mana meskipun XGBoost menunjukkan akurasi yang lebih rendah, analisis ini menyoroti peran penting algoritma dalam konteks klasifikasi data ekspresi gen. Dengan demikian, penerapan Gradient Boosting, khususnya XGBoost, dalam klasifikasi diabetes tipe 2 tidak hanya menawarkan hasil yang menjanjikan tetapi juga membuka peluang untuk penelitian lebih lanjut dalam optimalisasi dan aplikasi klinis.

Penerapan algoritma Gradient Boosting dalam klasifikasi penyakit diabetes tipe 2 merupakan topik yang semakin relevan dalam bidang kesehatan dan teknologi informasi. Studi literatur ini bertujuan untuk mengeksplorasi efektivitas dan efisiensi metode Gradient Boosting dalam mendiagnosis diabetes tipe 2, serta untuk memberikan pemahaman yang lebih baik tentang bagaimana algoritma ini dapat digunakan untuk meningkatkan akurasi prediksi dan deteksi dini penyakit. Berbagai penelitian menunjukkan bahwa algoritma seperti XGBoost dan LightGBM tidak hanya mampu menghasilkan model yang lebih akurat dibandingkan dengan metode tradisional, tetapi juga dapat menangani dataset besar dengan lebih efisien. Dengan demikian, tujuan dari studi literatur ini adalah untuk merangkum berbagai temuan terkait penerapan Gradient Boosting dalam klasifikasi diabetes tipe 2 dan untuk menilai kontribusinya terhadap peningkatan diagnosis dan pengelolaan penyakit.

Gradient Boosting merupakan salah satu algoritma machine learning yang efektif untuk klasifikasi, termasuk dalam diagnosis penyakit diabetes tipe 2. Metrik evaluasi seperti akurasi dan f-measure sering digunakan untuk menilai kinerja model. Akurasi mengukur proporsi prediksi yang benar, sed angka f-measure menggabungkan precision dan recall untuk memberikan gambaran yang lebih komprehensif tentang kinerja model, terutama dalam konteks ketidakseimbangan kelas. Dalam penelitian yang dilakukan oleh (Derara Duba Rufo, Taye Girma, Achim Ibenthal, & Worku Gachena Negara, 2021), model Light Gradient Boosting Machine (LightGBM) menunjukkan akurasi 98.1%, dengan nilai AUC 98.1%, sensitivitas 99.9%, dan spesifisitas 96.3%. Hasil ini menunjukkan bahwa LightGBM tidak hanya efektif dalam mendeteksi diabetes tipe 2, tetapi juga memberikan hasil yang lebih baik dibandingkan dengan algoritma lain seperti KNN, SVM, dan XGBoost. Penelitian ini menyoroti pentingnya pemilihan metrik yang tepat dalam evaluasi model untuk memastikan diagnosis yang akurat dan dapat diandalkan dalam konteks kesehatan masyarakat. (Nova Christina Sari, & Triana Linda Larasati, 2018).

Penerapan algoritma Gradient Boosting dalam klasifikasi penyakit diabetes tipe 2 menawarkan berbagai manfaat dan potensi yang signifikan dalam bidang kesehatan, terutama dalam upaya meningkatkan akurasi diagnosis dan efektivitas pengelolaan penyakit. Gradient Boosting, yang mencakup teknik seperti XGBoost dan LightGBM, dikenal karena kemampuannya dalam menangani data yang kompleks dan besar, serta dalam mengurangi kesalahan prediksi melalui pendekatan ensemble learning. Algoritma ini mampu mencapai tingkat akurasi yang lebih tinggi dibandingkan dengan metode tradisional seperti regresi logistik dan pohon keputusan, sehingga memberikan hasil yang lebih andal dalam mengidentifikasi pasien yang berisiko terkena diabetes tipe 2. (Liu, Y., Wang, Y., & Zhang, J., 2019). Dengan demikian, penerapan Gradient Boosting memiliki potensi besar untuk meningkatkan sistem kesehatan melalui deteksi dini dan pengelolaan diabetes tipe 2 yang lebih baik, memberikan manfaat tidak hanya bagi pasien tetapi juga bagi penyedia layanan kesehatan.

Literatur review ini bertujuan untuk mengeksplorasi dan menganalisis penerapan algoritma Gradient Boosting dalam klasifikasi penyakit diabetes tipe 2, dengan harapan dapat memberikan wawasan yang mendalam mengenai efektivitas dan potensi metode ini dalam meningkatkan diagnosis dan pengelolaan penyakit. Dalam konteks meningkatnya angka kejadian diabetes tipe 2 di seluruh dunia, terdapat kebutuhan mendesak untuk pendekatan yang lebih canggih dan akurat dalam

identifikasi pasien berisiko. Melalui tinjauan ini, diharapkan dapat diidentifikasi berbagai studi dan temuan yang menunjukkan bagaimana Gradient Boosting, dengan kemampuannya dalam menangani data yang kompleks dan menghasilkan model prediksi yang lebih baik, dapat berkontribusi pada deteksi dini dan intervensi yang lebih efektif. Selain itu, literatur review ini juga berharap untuk menggarisbawahi tantangan yang dihadapi dalam penerapan algoritma ini, serta memberikan rekomendasi untuk penelitian lebih lanjut. Dengan demikian, hasil dari tinjauan ini diharapkan dapat memberikan kontribusi signifikan bagi pengembangan metode berbasis data dalam bidang kesehatan, khususnya dalam upaya memerangi diabetes tipe 2 dan meningkatkan kualitas hidup pasien.

2. METODOLOGI PENELITIAN

2.1 Jenis Penelitian

Penelitian ini mengadopsi pendekatan machine learning untuk mengklasifikasi pasien Diabetes Melitus Tipe 2. Algoritma yang digunakan adalah Gradient Boosting. Metode ini dipilih karena kemampuannya dalam menangani dataset yang kompleks dan menghasilkan model prediksi yang akurat. Data yang digunakan meliputi variabel demografi, hasil pemeriksaan laboratorium, dan faktor risiko lainnya. Proses klasifikasi dilakukan dengan melatih model Gradient Boosting pada data latih, kemudian mengevaluasi performanya pada data uji. Sebagai metrik evaluasi, digunakan akurasi, presisi, recall, dan F1-score. Penelitian serupa yang telah dilakukan oleh (Salma Irena Febriastia, 2024) menggunakan algoritma LightGBM, sebuah varian dari Gradient Boosting, untuk mengklasifikasi penyakit Diabetes Melitus Tipe 2 dengan hasil yang menjanjikan. Studi ini menginspirasi kami untuk mengeksplorasi lebih lanjut potensi algoritma Gradient Boosting dalam konteks yang sama.

2.2 Gradient Boosting

Gradient Boosting adalah teknik machine learning yang efektif untuk klasifikasi penyakit diabetes tipe 2, bekerja dengan menggabungkan beberapa model prediktif sederhana ("weak learners") menjadi satu model yang lebih kuat. Proses ini dimulai dengan membangun model pertama yang memprediksi target, kemudian model-model berikutnya dibangun untuk memperbaiki kesalahan dari model sebelumnya dengan memfokuskan pada data yang sulit diprediksi. Dalam penerapannya, data yang digunakan biasanya mencakup berbagai fitur seperti kadar glukosa dan indeks massa tubuh. Setelah data dikumpulkan dan dibersihkan, Gradient Boosting digunakan untuk melatih model dengan cara iteratif, melibatkan tuning hyperparameter untuk mengoptimalkan kinerja model. Dengan pendekatan ini, Gradient Boosting dapat memberikan akurasi yang tinggi dalam mengklasifikasikan pasien diabetes tipe 2, menjadikannya pilihan yang populer dalam analisis data medis. (Silvia Elsa Suryana, Budi Warsito, Suparti, 2021).

2.3 Teknik Pengumpulan Data

Gradient Boosting adalah teknik machine learning yang efektif untuk klasifikasi penyakit diabetes tipe 2, bekerja dengan menggabungkan beberapa model prediktif sederhana ("weak learners") menjadi satu model yang lebih kuat. Proses ini dimulai dengan membangun model pertama yang memprediksi target, kemudian model-model berikutnya dibangun untuk memperbaiki kesalahan dari model sebelumnya dengan memfokuskan pada data yang sulit diprediksi. Dalam penerapannya, data yang digunakan biasanya mencakup berbagai fitur seperti kadar glukosa dan indeks massa tubuh. Setelah data dikumpulkan dan dibersihkan, Gradient Boosting digunakan untuk melatih model dengan cara iteratif, melibatkan tuning hyperparameter untuk mengoptimalkan kinerja model. Dengan pendekatan ini, Gradient Boosting dapat memberikan akurasi yang tinggi dalam mengklasifikasikan pasien diabetes tipe 2, menjadikannya pilihan yang populer dalam analisis data medis. (Silvia Elsa Suryana, Budi Warsito, Suparti, 2021).

2.4 Preprocessing Data

Penerapan Gradient Boosting dalam klasifikasi penyakit diabetes tipe 2 melibatkan proses yang sistematis, mulai dari pengumpulan data yang relevan dan terpercaya, seperti usia, BMI, dan

kadar glukosa, hingga pembersihan data untuk memastikan kualitas dataset. Setelah data bersih, eksplorasi data dilakukan untuk memahami pola dan hubungan antar variabel, mengidentifikasi fitur berpengaruh, dan mungkin melakukan ekstraksi fitur. Model Gradient Boosting kemudian diimplementasikan dengan membangun model prediktif secara bertahap, melibatkan tuning hyperparameter untuk mengoptimalkan kinerja model. Akhirnya, model yang telah dilatih dievaluasi menggunakan metrik seperti akurasi dan presisi untuk memastikan efektivitasnya dalam mengklasifikasikan pasien diabetes tipe 2. (jajang jaya purnama, 2020).

2.5 Evaluasi Model

Penelitian ini menggunakan algoritma Gradient Boosting untuk klasifikasi penyakit diabetes tipe 2, sebuah teknik ensemble learning yang menggabungkan beberapa model weak learners (pohon keputusan) untuk meningkatkan akurasi prediksi secara bertahap. Evaluasi model dilakukan dengan menggunakan metrik kinerja seperti akurasi, precision, recall, dan F1-score, serta AUC-ROC untuk mengevaluasi kemampuan model dalam membedakan kelas positif dan negatif. Penelitian sebelumnya menunjukkan bahwa Gradient Boosting dapat menghasilkan kinerja yang tinggi pada tugas klasifikasi medis, termasuk deteksi diabetes, dengan nilai AUC-ROC yang umumnya lebih tinggi dibandingkan model sederhana lainnya. Untuk meningkatkan keandalan evaluasi, dilakukan teknik cross-validation dengan membagi dataset menjadi beberapa subset untuk mengurangi risiko overfitting. Melalui pendekatan ini, Gradient Boosting diharapkan mampu memberikan prediksi yang akurat dan andal dalam klasifikasi diabetes tipe 2, sehingga dapat mendukung pengambilan keputusan klinis yang lebih baik. (Tianqi Chen, Carlos Guestrin, 2016).

2.6 Analisis Hasil

Penelitian tentang penerapan Gradient Boosting untuk klasifikasi penyakit diabetes tipe 2 umumnya bersifat eksperimen kuantitatif, menggunakan algoritma Gradient Boosting, sebuah metode ensemble learning yang menggabungkan beberapa weak learners (model decision tree) untuk meningkatkan akurasi prediksi. Penelitian ini memanfaatkan data pasien, seperti dari dataset Pima Indians, yang mencakup fitur-fitur seperti kadar glukosa, tekanan darah, dan indeks massa tubuh. Proses penelitian dimulai dengan prapemrosesan data, diikuti dengan pemisahan data menjadi set pelatihan dan pengujian. Model Gradient Boosting kemudian diterapkan untuk membangun klasifikasi, mampu mengenali pola-pola spesifik dalam data medis pasien diabetes. Studi-studi yang menggunakan algoritma ini menunjukkan bahwa Gradient Boosting cukup efektif dalam meningkatkan akurasi prediksi dibandingkan metode seperti Random Forest atau Logistic Regression untuk mendeteksi diabetes berdasarkan karakteristik kesehatan pasien. (Sahat Pandapotan Nainggolan, Ardiles Sinaga, 2023).

3. ANALISA DAN PEMBAHASAN

Tabel 1. Studi Literature Review

No.	Peneliti dan Tahun	Metode yang dibahas	Tujuan Penelitian	Hasil yang di dapat
1	- Kartina Diah Kusuma W. - Memen Akbar - Desember 2022	Extreme Gradient Boosting (XGBoost).	membangun model prediksi risiko diabetes pada pasien menggunakan algoritma Extreme Gradient Boosting (XGBoost).	- Model prediksi diabetes yang dihasilkan menggunakan XGBoost ditampilkan dalam bentuk tree. - Model ini mencapai akurasi sebesar 98,71%. - Penelitian ini menunjukkan bahwa algoritma XGBoost dapat

				<p>diimplementasikan pada pemodelan dataset diabetes untuk menghasilkan prediksi risiko diabetes dengan akurasi yang tinggi.</p> <ul style="list-style-type: none"> - XGBoost memiliki performa yang baik untuk data yang terstruktur.
2	<ul style="list-style-type: none"> - Ginanjar Abdurrahman - Hardian Oktavianto - Mukti Sintawati - 2022 	Extreme Gradient Boosting (XGBoost)	mengoptimalkan algoritma XGBoost Classifier dengan menggunakan hyperparameter tuning melalui Gridsearch dan Random Search	<p>Penelitian ini menemukan bahwa menggunakan hyperparameter tuning dengan Gridsearch dan Random Search pada algoritma XGBoost meningkatkan akurasi model secara signifikan.</p> <ul style="list-style-type: none"> - Tanpa hyperparameter tuning, akurasi model XGBoost mencapai 75%. - Dengan hyperparameter tuning menggunakan Gridsearch dan Random Search, akurasi model meningkat menjadi 95%. <p>Hasil ini menunjukkan bahwa hyperparameter tuning merupakan langkah penting dalam mengoptimalkan performa algoritma XGBoost.</p>
3	<ul style="list-style-type: none"> - M. Jishnu Sai - Pratiksha Chettri - Ranjit Panigrahi - Amik Garg - Akash Kumar Bhoi - Poalo Barsocchi - Januari 2023 	Metode ensemble learning	membangun model prediksi penyakit diabetes tipe-2 yang lebih akurat dengan menggunakan metode ensemble learning.	<ul style="list-style-type: none"> - Model ensemble yang diusulkan berhasil mencapai akurasi deteksi 90,76% dengan menggunakan k-NN, Adaboost, dan LightGBM. - Model ensemble ini menunjukkan kinerja yang lebih baik dibandingkan dengan model-model lain yang ada. - Analisis kurva operasi penerima (ROC) menunjukkan bahwa

				k-NN, RF, dan LightGBM berhasil mengatasi masalah ketidakseimbangan kelas pada dataset yang digunakan.
4	<ul style="list-style-type: none"> - Iqbal Fathur Rahman - 2020 	<ul style="list-style-type: none"> - Support Vector Machine (SVM) - Multilayer Perceptron (MLP) - Xtreme Gradient Boosting (XGBoost) 	<ul style="list-style-type: none"> - Menganalisis data ekspresi gen skeletal muscle NGT, IGT, dan diabetes melitus tipe 2 pada data microarray. - Mengklasifikasikan data ekspresi gen skeletal muscle NGT, IGT, dan diabetes melitus tipe 2 menggunakan metode SVM, MLP, dan XGBoost. - Mengevaluasi akurasi dari ketiga metode klasifikasi tersebut. 	<ul style="list-style-type: none"> - Analisis deskriptif menunjukkan bahwa data terdiri dari 118 sampel dengan 3 kelas: NGT, IGT, dan diabetes melitus tipe 2. - Model SVM dengan kernel linear memiliki akurasi tertinggi (91,30%) dalam mengklasifikasikan data ekspresi gen. - Model MLP dengan 1 hidden layer dan 100 hidden nodes, serta epoch 400, mencapai akurasi 78,26%. - Model XGBoost dengan parameter yang dioptimalkan mencapai akurasi 73,91%.
5	<ul style="list-style-type: none"> - Sahat Pandapotan Nainggolan - Ardiles Sinaga - Juni 2023 	<ul style="list-style-type: none"> - Random Forest Algorithm - Gradient Boosting Classifier Algorithm 	Untuk menentukan akurasi dari kedua algoritma tersebut dalam mengklasifikasikan data diabetes	penelitian ini menunjukkan bahwa Gradient Boosting Classifier Algorithm lebih efektif dalam mengklasifikasikan data diabetes dibandingkan dengan Random Forest Algorithm.

4. KESIMPULAN

Dari rangkuman hasil penelitian yang disajikan, dapat disimpulkan bahwa:

- a. Algoritma Extreme Gradient Boosting (XGBoost) secara konsisten menunjukkan akurasi yang tinggi dalam memprediksi risiko diabetes, mencapai 98,71%. Keunggulan XGBoost terletak pada akurasinya yang sangat baik, efisiensinya dalam pemodelan, dan fleksibilitasnya dalam mengatasi berbagai jenis data terstruktur.
- b. Penggunaan hyperparameter tuning dengan Gridsearch dan Random Search pada algoritma XGBoost secara signifikan meningkatkan akurasi model dari 75% menjadi 95%. Hal ini menunjukkan pentingnya optimasi parameter untuk meningkatkan performa algoritma XGBoost.
- c. Model ensemble yang menggabungkan berbagai algoritma (k-NN, Naive Bayes, Random Forest, Adaboost, LightGBM) berhasil meningkatkan akurasi prediksi

penyakit diabetes hingga 90,76%. Model ensemble ini efektif dalam mengatasi masalah ketidakseimbangan kelas pada dataset.

- d. Metode klasifikasi seperti Support Vector Machine (SVM), Multilayer Perceptron (MLP), dan Xtreme Gradient Boosting (XGBoost) dapat digunakan untuk mengklasifikasikan data ekspresi gen, dengan SVM kernel linear menunjukkan akurasi tertinggi.
- e. Algoritma Gradient Boosting Classifier lebih unggul daripada Random Forest dalam memprediksi diabetes, dengan akurasi mencapai 81%. Penggunaan parameter yang sesuai dapat meningkatkan performa model dalam memprediksi penyakit diabetes.

Secara keseluruhan, penelitian ini menyoroti keberhasilan XGBoost dalam memprediksi risiko diabetes, pentingnya optimasi parameter, potensi model ensemble dalam deteksi penyakit, serta perbandingan keunggulan berbagai metode klasifikasi dalam konteks bioinformatika. Dengan demikian, penerapan algoritma yang tepat dan strategi tuning yang cermat dapat memberikan kontribusi signifikan dalam mendiagnosis dan mengidentifikasi penyakit diabetes secara efektif.

REFERENCES

- Derara Duba Rufo, Taye Girma, Achim Ibenthal, & Worku Gachena Negara. (2021). Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM). *Diagnostics*.
- Fiska R. (2021). Teknik Pengumpulan Data dalam Rancangan Penelitian. *statistik*.
- GINANJAR ABDURRAHMAN, HARDIAN OKTAVIANTO, MUKTI SINTAWATI. (2022). Optimasi algoritma XGBoost Classifier dengan hyperparameter tuning. *informatics journal*.
- Iqbal Fathur Rahman. (2020). Analisis data ekspresi gen skeletal muscle NGT, IGT, dan diabetes melitus tipe 2. *jurnal internasional sistem intelijen komputasi*.
- Jajang Jaya Purnama. (2020). Analisis Algoritma Klasifikasi Neural Network Untuk Diagnosis Penyakit Diabetes. *IJCIT*.
- Aura Amalia Warzuqni, Divia Putri Sabilla, Zara Agustin, Perani Rosyani. (2022). Analisa Sistem Presensi Kelas Menggunakan Pengenal Wajah Dengan Metode Haar Cascade Classifier. *MANEKIN*.
- Kartina Diah Kusuma W., Memen Akbar. (2022). Extreme Gradient Boosting (XGBoost): Membangun model prediksi risiko diabetes pada pasien. *join*.
- Liu, Y., Wang, Y., & Zhang, J. (2019). A novel predictive model for diabetes based on gradient boosting decision tree. *BMC Medical Informatics and Decision*.
- Nova Christina Sari, & Triana Linda Larasati. (2018). Komparasi Algoritma Naïve Bayes dan Gradient Boosting untuk Prediksi Pasien Diabetes. *Jurnal Nasional Teknologi & Sistem Informasi*.
- Sahat Pandapotan Nainggolan, Ardiles Sinaga. (2023). COMPARATIVE ANALYSIS OF ACCURACY OF RANDOM FOREST AND GRADIENT BOOSTING CLASSIFIER ALGORITHM FOR DIABETES CLASSIFICATION. *sebatik*.
- Erni, Affandi Agung Laksosno, Muchlas Syahlanisyiam, Perani Rosyani. (2023). Sistem Pakar Diagnosa Penyakit Kulit Dengan Menggunakan Metode Forward Chaining. *MANEKIN*.
- Salma Irena Febriastia. (2024). KLASIFIKASI PENYAKIT DIABETES MELLITUS TIPE II BERBASIS MACHINE LEARNING MENGGUNAKAN LIGHTGBM. *digital repository unila*.
- Silvia Elsa Suryana, Budi Warsito, Suparti. (2021). PENERAPAN GRADIENT BOOSTING DENGAN HYPEROPT UNTUK . *JURNAL GAUSSIAN*.
- Tianqi Chen, Carlos Guestrin. (2016). XGBoost: A Scalable Tree Boosting System. *cornell university*.
- Anisa Maulida, Arisky Rahmatulloh, Irwan Ahussalim, Robby Alvian, Perani Rosyani. (2023). Analisis Metode Forward Chaining pada Sistem Pakar Systematic Literatur Review. *MANEKIN*.