



Evaluasi Komparatif Algoritma Machine Learning untuk Prediksi Dini Diabetes

Aniq Astofa*, Perani Rosyani, Rahmawati, Sopiyan Apandi

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Tangerang Selatan, Indonesia

Email: ^{1,*}dosen02360@unpam.ac.id, ²dosen00837@unpam.ac.id, ³dosen02394@unpam.ac.id, ⁴dosen02601@unpam.ac.id

Email Penulis Korespondensi: dosen02360@unpam.ac.id

Abstrak—Diabetes merupakan salah satu penyakit tidak menular yang sering terdeteksi pada tahap lanjut, sehingga meningkatkan risiko komplikasi serius. Pemanfaatan *machine learning* berpotensi mendukung deteksi dini diabetes, namun sebagian besar penelitian sebelumnya berfokus pada dataset berskala besar dan pencapaian akurasi tinggi, sementara evaluasi metodologis pada data klinis berukuran kecil masih terbatas. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja beberapa algoritma *machine learning* dalam prediksi dini diabetes berbasis data klinis dengan jumlah data yang terbatas, serta mengkaji dampak karakteristik data terhadap kinerja model. Dataset yang digunakan terdiri atas 22 sampel dengan delapan fitur klinis dan satu variabel target, yang dibagi menjadi 17 data latih dan 5 data uji. Tahapan penelitian meliputi prapemrosesan data, pembagian data latih dan data uji, pelatihan model, serta evaluasi kinerja menggunakan metrik akurasi, presisi, *recall*, *F1-score*, dan ROC-AUC. Algoritma yang digunakan meliputi Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan XGBoost. Hasil pengujian menunjukkan bahwa seluruh model belum mampu mendeteksi kelas diabetes secara efektif, yang ditunjukkan oleh nilai presisi, *recall*, dan *F1-score* yang bernilai nol. Random Forest dan XGBoost memperoleh nilai akurasi sebesar 0,6, namun nilai tersebut dipengaruhi oleh dominasi kelas non-diabetes pada data uji yang sangat terbatas. Analisis korelasi menunjukkan bahwa Glucose, BMI, dan Diabetes Pedigree Function merupakan fitur yang paling berpengaruh terhadap status diabetes. Kontribusi utama penelitian ini terletak pada penyajian evaluasi metodologis yang realistis terhadap penerapan *machine learning* pada data klinis berukuran kecil, dengan menegaskan bahwa keterbatasan jumlah data dan pembagian data latih–uji memiliki dampak signifikan terhadap kinerja model dan interpretasi metrik evaluasi. Temuan ini memberikan dasar penting bagi pengembangan model prediksi dini diabetes selanjutnya yang lebih andal pada kondisi data klinis terbatas.

Kata Kunci: Machine Learning; Prediksi Dini, Diabetes, Data Klinis, Klasifikasi

Abstract—Diabetes is one of the non-communicable diseases that is often detected at an advanced stage, thereby increasing the risk of serious complications. The application of machine learning has the potential to support early diabetes detection; however, most previous studies have focused on large-scale datasets and high predictive accuracy, while methodological evaluations on small-sized clinical data remain limited. This study aims to evaluate and compare the performance of several machine learning algorithms for early diabetes prediction using a limited clinical dataset, with particular emphasis on analyzing the impact of data characteristics on model performance. The dataset used in this study consists of 22 samples with eight clinical features and one target variable, which were divided into 17 training samples and 5 testing samples. The research stages include data preprocessing, training–testing data splitting, model training, and performance evaluation using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The algorithms evaluated include Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and XGBoost. The experimental results indicate that none of the evaluated models were able to effectively detect the diabetes class, as reflected by precision, recall, and F1-score values of zero across all models. Although Random Forest and XGBoost achieved an accuracy of 0.6, this value was largely influenced by the dominance of the non-diabetes class in the very limited test set. Correlation analysis further reveals that Glucose, BMI, and Diabetes Pedigree Function are the most influential features associated with diabetes status. The main contribution of this study lies in providing a realistic methodological evaluation of machine learning models applied to small-sized clinical datasets, highlighting that limited sample size and training–testing data partitioning have a substantial impact on model performance and the interpretation of evaluation metrics. These findings provide an important methodological reference for future studies aiming to develop more reliable early diabetes prediction models under constrained clinical data conditions.

Keywords: Machine Learning; Early Prediction; Diabetes; Clinical Data; Classification

1. PENDAHULUAN

Diabetes melitus merupakan salah satu penyakit tidak menular dengan tingkat pertumbuhan tercepat di dunia dan menjadi tantangan serius bagi sistem kesehatan global. Peningkatan prevalensi diabetes terjadi secara konsisten seiring perubahan gaya hidup, pola konsumsi tinggi kalori, serta rendahnya aktivitas fisik.[1][2] Permasalahan utama dalam pengendalian diabetes bukan hanya tingginya jumlah penderita, tetapi juga besarnya proporsi kasus yang tidak terdiagnosis secara dini, sehingga banyak pasien baru mendapatkan penanganan ketika komplikasi telah berkembang. Kondisi ini berdampak pada meningkatnya risiko komplikasi kronis seperti penyakit kardiovaskular, gagal ginjal[3], gangguan penglihatan, serta neuropati, yang pada akhirnya menurunkan kualitas hidup pasien dan meningkatkan beban biaya layanan kesehatan[4]

Deteksi dini diabetes menjadi langkah strategis dalam upaya pencegahan dan pengendalian penyakit ini.[5] Namun, proses diagnosis konvensional masih bergantung pada pemeriksaan laboratorium yang bersifat reaktif dan dilakukan setelah gejala klinis muncul. Seiring berkembangnya teknologi komputasi dan ketersediaan data klinis dalam jumlah besar, *machine learning* berkembang sebagai pendekatan yang menjanjikan untuk mendukung deteksi dini diabetes[6]. *Machine learning* mampu mempelajari pola kompleks dari data kesehatan dan menghasilkan model prediksi yang lebih adaptif dibandingkan metode statistik tradisional, khususnya dalam menangani hubungan non-linear antarvariabel klinis[7][8]



Berdasarkan karakteristik klinis diabetes dan kemampuan machine learning dalam mengenali pola non-linear, kerangka konseptual penelitian ini dibangun dengan menempatkan data klinis berukuran kecil sebagai tantangan utama dalam prediksi dini. Pada kondisi keterbatasan data, proses pembelajaran model tidak hanya bergantung pada kompleksitas algoritma, tetapi juga pada distribusi kelas, relevansi fitur klinis, serta pemilihan metrik evaluasi yang tepat. Oleh karena itu, pendekatan prediksi dini diabetes dalam penelitian ini diposisikan sebagai evaluasi perilaku algoritma machine learning pada skenario data terbatas, bukan semata-mata optimalisasi performa prediktif.

Permasalahan penelitian ini diformulasikan sebagai permasalahan metodologis dalam penelitian komputasional, yaitu bagaimana kinerja algoritma machine learning dalam melakukan klasifikasi dini diabetes ketika dihadapkan pada data klinis berukuran kecil dan distribusi kelas yang tidak seimbang. Kondisi ini berpotensi menyebabkan bias prediksi, kegagalan deteksi kelas minoritas, serta distorsi interpretasi kinerja model apabila hanya mengandalkan metrik akurasi.

Meskipun berbagai penelitian sebelumnya telah melaporkan keberhasilan penerapan machine learning dalam prediksi penyakit, sebagian besar studi tersebut menggunakan dataset berskala besar atau berfokus pada peningkatan akurasi model. Perbedaan mendasar penelitian ini terletak pada fokus evaluatif terhadap kegagalan model dalam mendeteksi kelas diabetes pada data klinis berukuran kecil. Berbeda dari penelitian terdahulu yang menempatkan akurasi sebagai indikator utama, penelitian ini memosisikan recall dan ROC-AUC sebagai indikator kritis dalam konteks medis. Dengan demikian, penelitian ini mengambil posisi sebagai studi evaluatif yang menelaah keterbatasan algoritma, bukan sekadar mengejar performa optimal. [9]

Di sisi lain, penelitian berbasis *continuous glucose monitoring* (CGM)[10] menitikberatkan pada efektivitas intervensi klinis terhadap pengendalian glukosa darah, bukan pada pengembangan model prediksi berbasis data klinis. Pendekatan ini memberikan kontribusi penting dalam praktik klinis, tetapi tidak membahas permasalahan klasifikasi dan generalisasi model *machine learning* pada data dengan ukuran kecil dan distribusi kelas yang tidak seimbang.

Penelitian lain[11] yang menerapkan *machine learning* untuk klasifikasi penyakit, seperti deteksi kanker paru, umumnya melaporkan tingkat akurasi yang tinggi, namun masih terbatas dalam membahas dampak ukuran data, ketidakseimbangan kelas, serta kegagalan model dalam mendeteksi kelas minoritas. Sebagian besar penelitian tersebut berfokus pada pencapaian performa optimal, tanpa mengkaji secara kritis keterbatasan model ketika diterapkan pada data klinis yang bersifat terbatas dan heterogen.

Berdasarkan beberapa penelitian terdahulu, masih terdapat celah penelitian terkait evaluasi kinerja algoritma *machine learning*[12][13] dalam prediksi dini diabetes pada kondisi data klinis berukuran kecil dan tidak seimbang. Penelitian ini mengisi celah tersebut dengan mengevaluasi perilaku dan keterbatasan beberapa algoritma *machine learning* dalam mendeteksi kelas diabetes, serta menekankan pentingnya penggunaan metrik evaluasi yang relevan secara klinis, seperti *recall* dan ROC-AUC, dibandingkan hanya mengandalkan akurasi.

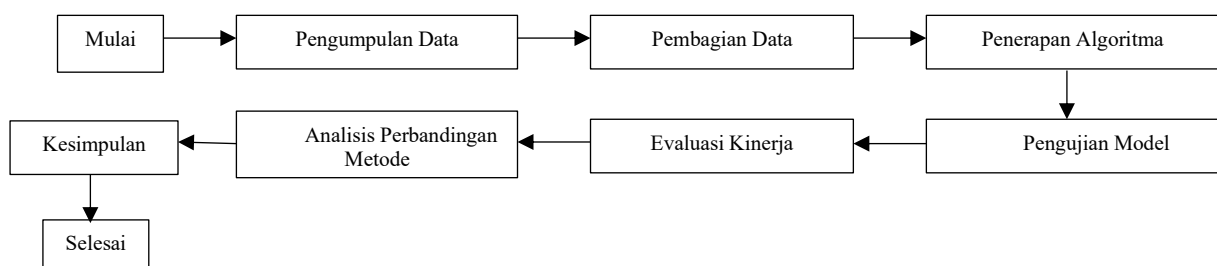
Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja beberapa algoritma machine learning dalam prediksi dini diabetes pada data klinis berukuran kecil dengan distribusi kelas yang tidak seimbang.

Kontribusi utama penelitian ini adalah memberikan evaluasi empiris terhadap kinerja algoritma machine learning pada skenario data klinis berukuran kecil, khususnya dalam konteks prediksi dini diabetes. Penelitian ini tidak hanya membandingkan performa antar algoritma, tetapi juga menegaskan keterbatasan penggunaan metrik akurasi sebagai indikator tunggal. Dengan demikian, penelitian ini berkontribusi sebagai referensi metodologis bagi pengembangan sistem prediksi dini berbasis data klinis terbatas pada tahap awal implementasi.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental untuk mengevaluasi kinerja beberapa algoritma *machine learning* dalam melakukan prediksi dini diabetes berbasis data klinis. Alur penelitian dimulai dari pengumpulan data hingga analisis hasil prediksi, yang secara ringkas disajikan dalam bentuk diagram alir pada Gambar 1. Diagram tersebut menggambarkan hubungan antar tahapan penelitian dan peran masing-masing tahapan dalam membangun serta mengevaluasi model prediksi dini diabetes.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, tahapan penelitian dapat dijelaskan sebagai berikut.



a. Pengumpulan Data

Tahap awal penelitian adalah pengumpulan data klinis yang digunakan sebagai dasar pemodelan. Dataset yang digunakan dalam penelitian ini adalah *Pima Indians Diabetes Dataset*, yang berisi atribut klinis pasien dan satu variabel target berupa status diabetes. Dataset ini dipilih karena sering digunakan sebagai *benchmark* dalam penelitian prediksi diabetes, sehingga memungkinkan perbandingan hasil secara metodologis dengan studi terdahulu.

b. Prapemrosesan Data

Data yang telah dikumpulkan kemudian melalui tahap prapemrosesan untuk meningkatkan kualitas data sebelum digunakan dalam pemodelan. Tahapan ini meliputi penanganan *missing values*, normalisasi fitur numerik agar berada pada skala yang seragam, serta proses encoding apabila diperlukan. Prapemrosesan data bertujuan untuk mengurangi bias model dan memastikan bahwa setiap algoritma dapat mempelajari pola data secara optimal.

c. Pembagian Data

Data yang telah diproses selanjutnya dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Pada penelitian ini, total dataset setelah pembersihan berjumlah 22 sampel, yang dibagi menjadi 17 data latih dan 5 data uji. Pembagian ini mencerminkan kondisi keterbatasan data klinis yang umum ditemui pada tahap awal pengembangan sistem prediksi kesehatan.

d. Penerapan Algoritma Machine Learning

Pada tahap ini, beberapa algoritma *machine learning* bertipe *supervised learning* diterapkan pada data latih. Algoritma yang digunakan meliputi Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan XGBoost. Penggunaan beberapa algoritma bertujuan untuk membandingkan perilaku dan kinerja masing-masing metode pada kondisi data berukuran kecil dan distribusi kelas yang tidak seimbang.

e. Pengujian Model

Model yang telah dilatih kemudian diuji menggunakan data uji yang tidak terlibat dalam proses pelatihan. Tahap pengujian ini bertujuan untuk menilai kemampuan generalisasi model dalam memprediksi status diabetes pada data baru. Dengan jumlah data uji yang sangat terbatas, hasil pengujian lebih merepresentasikan perilaku model dibandingkan performa prediktif jangka panjang.

f. Evaluasi Kinerja Model

Evaluasi kinerja dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC. Penggunaan lebih dari satu metrik bertujuan untuk memperoleh gambaran kinerja model secara komprehensif, khususnya dalam konteks medis di mana kemampuan mendeteksi kelas diabetes sebagai kelas minoritas lebih penting dibandingkan nilai akurasi semata.

g. Analisis Perbandingan Metode

Hasil evaluasi dari seluruh algoritma selanjutnya dianalisis secara komparatif untuk mengidentifikasi perbedaan kinerja, keunggulan, dan keterbatasan masing-masing metode. Analisis ini difokuskan pada perilaku algoritma dalam menghadapi keterbatasan jumlah data dan ketidakseimbangan kelas.

h. Kesimpulan

Tahap akhir penelitian adalah penarikan kesimpulan berdasarkan hasil evaluasi dan analisis perbandingan metode. Kesimpulan difokuskan pada implikasi metodologis penggunaan *machine learning* untuk prediksi dini diabetes pada data klinis berukuran kecil serta rekomendasi pengembangan penelitian selanjutnya.

2.2 Algoritma yang Digunakan

Penelitian ini menggunakan beberapa algoritma *machine learning* bertipe *supervised learning* yang umum digunakan dalam permasalahan klasifikasi data medis. Pemilihan algoritma didasarkan pada pertimbangan karakteristik data klinis, kemampuan algoritma dalam menangani hubungan non-linear, serta tingkat interpretabilitas model.

Algoritma pertama adalah Logistic Regression[14], yang digunakan sebagai model dasar (*baseline*) dalam penelitian ini. Logistic Regression memodelkan probabilitas kejadian biner menggunakan fungsi logistik dan memiliki keunggulan dalam hal interpretabilitas. Algoritma ini memungkinkan identifikasi pengaruh masing-masing fitur klinis terhadap risiko diabetes, sehingga sering digunakan dalam aplikasi medis meskipun memiliki keterbatasan dalam menangani hubungan non-linear.[15]

Algoritma kedua adalah Random Forest[16], yaitu metode *ensemble learning* yang menggabungkan banyak *decision tree* untuk meningkatkan akurasi dan stabilitas prediksi. Random Forest mampu menangani data berdimensi tinggi dan hubungan non-linear antar fitur, serta relatif tahan terhadap *overfitting*. Selain itu, algoritma ini menyediakan informasi *feature importance* yang berguna untuk analisis faktor risiko diabetes.[17]

Algoritma ketiga adalah Support Vector Machine (SVM)[18], yang bekerja dengan mencari *hyperplane* optimal untuk memisahkan kelas data. SVM efektif digunakan pada data berdimensi tinggi dan mampu menangani klasifikasi non-linear melalui penggunaan fungsi kernel. Namun, performa SVM sangat dipengaruhi oleh pemilihan parameter dan jenis kernel yang digunakan.

Algoritma keempat adalah K-Nearest Neighbors (KNN)[11], yang merupakan metode klasifikasi berbasis jarak. KNN mengklasifikasikan data baru berdasarkan mayoritas kelas dari sejumlah tetangga terdekat. Algoritma ini sederhana dan mudah diimplementasikan, tetapi sensitif terhadap skala data dan kurang efisien pada dataset berukuran besar.

Algoritma terakhir adalah XGBoost[1], yang merupakan pengembangan dari metode *gradient boosting* dengan efisiensi komputasi yang lebih baik. XGBoost memiliki kemampuan prediksi yang tinggi, dilengkapi dengan mekanisme



regularisasi untuk mengurangi *overfitting*, sehingga sangat cocok untuk data kompleks seperti data klinis. Namun, algoritma ini memiliki tingkat kompleksitas yang lebih tinggi dibandingkan metode lainnya.

Melalui penerapan dan perbandingan beberapa algoritma tersebut, penelitian ini bertujuan untuk memperoleh gambaran kinerja model prediksi dini diabetes serta mempertimbangkan keseimbangan antara akurasi dan interpretabilitas dalam konteks medis.[19]

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian serta pembahasan yang diperoleh dari penerapan metode *machine learning* dalam prediksi dini diabetes berbasis data klinis. Penyajian hasil difokuskan pada analisis kinerja model klasifikasi yang telah dibangun melalui tahapan prapemrosesan data, pelatihan model, dan pengujian menggunakan data uji. Hasil penelitian disajikan secara sistematis untuk memberikan gambaran menyeluruh mengenai kemampuan masing-masing algoritma dalam mengidentifikasi risiko diabetes, serta untuk menilai efektivitas metode yang digunakan dalam menjawab permasalahan penelitian.

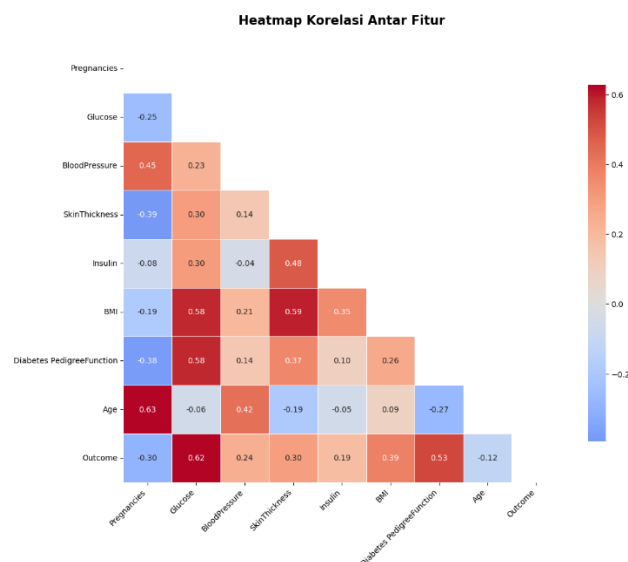
Pembahasan pada bagian ini tidak hanya menampilkan hasil numerik, tetapi juga menginterpretasikan makna dari setiap hasil evaluasi berdasarkan konteks medis dan karakteristik data yang digunakan. Oleh karena itu, hasil penelitian didukung oleh penyajian tabel dan gambar, seperti tabel perbandingan metrik evaluasi, visualisasi korelasi antar variabel klinis, serta grafik kinerja model. Penyajian ini bertujuan untuk mempermudah pemahaman pembaca terhadap pola data, perbedaan performa antar algoritma, serta implikasi hasil prediksi terhadap upaya deteksi dini diabetes.[4]

Secara umum, pembahasan diarahkan untuk mengaitkan hasil penelitian dengan tujuan yang telah ditetapkan pada bagian pendahuluan, yaitu mengevaluasi dan membandingkan kinerja beberapa algoritma *machine learning* dalam prediksi dini diabetes. Selain itu, bagian ini juga membandingkan temuan penelitian dengan hasil penelitian terdahulu guna menilai keterbaruan dan kontribusi penelitian. Dengan pendekatan tersebut, bagian hasil dan pembahasan diharapkan dapat memberikan penjelasan yang komprehensif mengenai keunggulan, keterbatasan, serta potensi penerapan metode yang digunakan dalam konteks layanan kesehatan.

3.1 Hasil Penerapan Metode Prediksi Dini Diabetes

Bagian ini menyajikan hasil penerapan metode *machine learning* dalam melakukan prediksi dini diabetes menggunakan data klinis. Proses analisis dimulai dari pemahaman karakteristik data, dilanjutkan dengan penerapan algoritma klasifikasi, hingga diperoleh hasil prediksi yang selanjutnya dievaluasi menggunakan metrik kinerja yang relevan secara klinis. Penyajian hasil dilakukan secara deskriptif dan komparatif untuk memberikan gambaran menyeluruh mengenai performa setiap metode yang digunakan.

Dataset yang digunakan dalam penelitian ini terdiri dari sejumlah atribut klinis numerik, antara lain jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan lipatan kulit, kadar insulin, indeks massa tubuh (BMI), *diabetes pedigree function*, dan usia, dengan satu atribut target berupa status diabetes. Sebelum dilakukan pemodelan, data terlebih dahulu melalui tahap prapemrosesan untuk memastikan kualitas data yang digunakan sudah layak dianalisis. Setelah prapemrosesan, data dibagi menjadi data latih dan data uji. Pada penelitian ini, data dibagi menjadi data latih sebanyak 17 sampel dan data uji sebanyak 5 sampel. Proporsi kelas pada data latih terdiri atas 70,6% non-diabetes dan 29,4% diabetes, sedangkan pada data uji terdiri atas 60% non-diabetes dan 40% diabetes. Pembagian data dengan ukuran yang sangat terbatas ini berpotensi memengaruhi kemampuan model dalam mempelajari pola kelas minoritas serta berdampak pada hasil evaluasi kinerja model.



Gambar 2. Heatmap setiap variabel

Gambar 2 menampilkan *heatmap* korelasi antar variabel klinis yang digunakan dalam penelitian ini. Visualisasi ini bertujuan untuk menggambarkan kekuatan dan arah hubungan linear antara setiap pasangan variabel, termasuk hubungan antara fitur klinis dengan variabel target (*Outcome*). Skala warna pada *heatmap* merepresentasikan nilai koefisien korelasi Pearson, yang berkisar dari -1 hingga $+1$.

Warna merah menunjukkan korelasi positif, yang mengindikasikan bahwa peningkatan nilai suatu variabel cenderung diikuti oleh peningkatan nilai variabel lainnya. Sebaliknya, warna biru menunjukkan korelasi negatif, yang menandakan hubungan berlawanan arah antar variabel. Sementara itu, warna yang mendekati putih merepresentasikan nilai korelasi mendekati nol, yang berarti tidak terdapat hubungan linear yang signifikan antara dua variabel.

Berdasarkan baris *Outcome* pada *heatmap*, dapat diamati bahwa variabel Glucose memiliki korelasi positif tertinggi dengan status diabetes ($r = 0,62$). Temuan ini menunjukkan bahwa semakin tinggi kadar glukosa darah, semakin besar kemungkinan seseorang terklasifikasi sebagai penderita diabetes. Hasil ini sejalan dengan pengetahuan klinis yang menempatkan kadar glukosa sebagai indikator utama dalam diagnosis diabetes. Variabel Diabetes Pedigree Function juga menunjukkan korelasi positif yang cukup kuat ($r = 0,53$), yang mengindikasikan bahwa faktor keturunan berperan penting dalam meningkatkan risiko diabetes. Selain itu, variabel BMI memiliki korelasi positif moderat ($r = 0,39$), yang menunjukkan keterkaitan antara tingkat obesitas dan kejadian diabetes.

Sebaliknya, variabel Pregnancies menunjukkan korelasi negatif terhadap *Outcome* ($r = -0,30$). Hubungan ini mengindikasikan adanya kecenderungan hubungan terbalik pada dataset yang digunakan, yang dapat dipengaruhi oleh karakteristik sampel dan ukuran data yang terbatas. Oleh karena itu, interpretasi terhadap korelasi negatif ini perlu dilakukan secara hati-hati dan tidak dapat digeneralisasi secara klinis tanpa dukungan data yang lebih besar.

Selain hubungan antara fitur dan variabel target, *heatmap* juga menunjukkan adanya korelasi yang cukup kuat antar fitur pendukung. Korelasi positif antara Age dan Pregnancies ($r = 0,63$) mencerminkan hubungan logis secara biologis, di mana bertambahnya usia cenderung diikuti oleh meningkatnya jumlah kehamilan. Korelasi kuat juga terlihat antara SkinThickness dan BMI ($r = 0,59$), yang mengindikasikan adanya potensi multikolinearitas karena kedua variabel tersebut sama-sama merepresentasikan aspek komposisi tubuh.

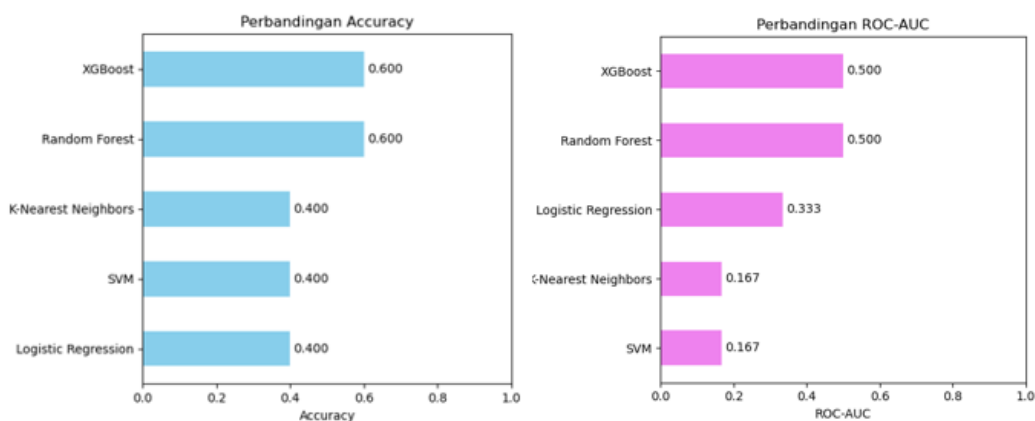
Meskipun beberapa variabel, khususnya Glucose, menunjukkan korelasi yang cukup kuat dengan status diabetes, hasil *heatmap* ini juga memberikan penjelasan terhadap kinerja model yang kurang optimal pada tahap pemodelan. Pada kondisi jumlah data yang sangat terbatas dan distribusi kelas yang tidak seimbang, informasi korelasi yang kuat tersebut belum tentu dapat dipelajari secara efektif oleh algoritma *machine learning*. Model cenderung memprioritaskan prediksi kelas mayoritas untuk mempertahankan nilai akurasi, sehingga kemampuan mendeteksi kelas minoritas (diabetes) menjadi sangat rendah. Dengan demikian, *heatmap* ini menegaskan bahwa keterbatasan kinerja model lebih disebabkan oleh faktor data dan proses pembagian data, bukan semata-mata oleh relevansi fitur yang digunakan.

3.2 Implementasi/Pengujian

Implementasi dan pengujian model dilakukan dengan menggunakan data latih sebanyak 17 sampel dan data uji sebanyak 5 sampel. Seluruh algoritma dilatih menggunakan data latih yang telah melalui proses standarisasi fitur, kemudian diuji pada data uji yang sepenuhnya terpisah untuk mengukur kemampuan generalisasi model. Dengan ukuran data uji yang sangat terbatas, hasil pengujian lebih merefleksikan perilaku model dalam kondisi data klinis awal daripada performa prediktif yang stabil.

Tahap implementasi dan pengujian model dilakukan untuk menilai kemampuan algoritma *machine learning* dalam mengklasifikasikan data pasien ke dalam kelas diabetes dan non-diabetes. Implementasi difokuskan pada aspek yang relevan dengan tujuan penelitian, tanpa menampilkan detail teknis pemrograman secara berlebihan. Pendekatan ini bertujuan agar pembahasan tetap berada dalam koridor ilmiah dan sesuai dengan karakteristik artikel jurnal.

Proses implementasi dimulai dengan pelatihan model menggunakan data latih yang telah melalui tahap prapemrosesan. Setiap algoritma dilatih menggunakan parameter awal yang seragam untuk menjaga objektivitas perbandingan. Selanjutnya, dilakukan pengujian menggunakan data uji yang tidak terlibat dalam proses pelatihan. Pengujian ini bertujuan untuk mengukur kemampuan generalisasi model terhadap data baru.

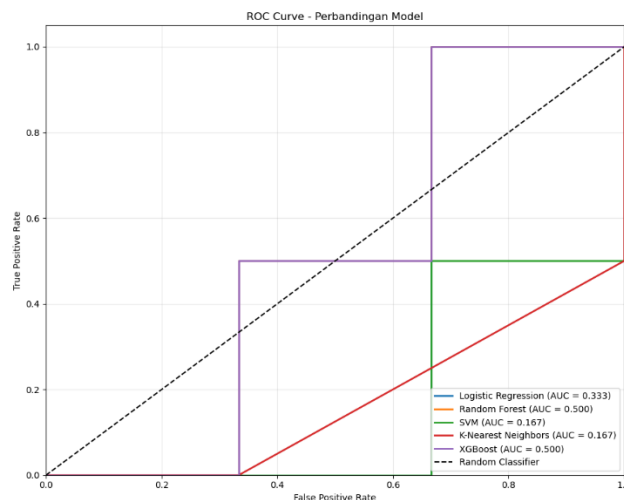


Gambar 3. Nilai akurasi

Perbedaan nilai akurasi antar algoritma pada Gambar 3 tidak dapat dijadikan satu-satunya indikator keunggulan model, karena nilai tersebut sangat dipengaruhi oleh ukuran data uji yang kecil serta dominasi kelas non-diabetes pada data pengujian. Berdasarkan hasil pengujian, algoritma Random Forest dan XGBoost memperoleh nilai *accuracy* sebesar 0,600, yang menunjukkan bahwa kedua model tersebut mampu mengklasifikasikan 60% data uji dengan benar. Sementara itu, Logistic Regression, KNN, dan SVM hanya mencapai nilai *accuracy* sebesar 0,400. Meskipun nilai *accuracy* Random Forest dan XGBoost terlihat lebih tinggi, metrik ini perlu ditafsirkan secara hati-hati. Pada kondisi distribusi kelas yang tidak seimbang dan jumlah data uji yang sangat terbatas, nilai *accuracy* yang relatif tinggi dapat terjadi akibat kecenderungan model untuk memprediksi kelas mayoritas secara konsisten, tanpa benar-benar mempelajari pola yang membedakan kedua kelas.

Selanjutnya, evaluasi menggunakan metrik ROC-AUC menunjukkan bahwa Random Forest dan XGBoost menghasilkan nilai sebesar 0,500, yang merepresentasikan kemampuan klasifikasi yang setara dengan tebakan acak. Hal ini mengindikasikan bahwa kedua model belum memiliki kemampuan diskriminatif yang memadai dalam membedakan antara pasien diabetes dan non-diabetes. Lebih lanjut, Logistic Regression, KNN, dan SVM menghasilkan nilai ROC-AUC di bawah 0,500, yang menunjukkan bahwa prediksi yang dihasilkan cenderung berlawanan dengan kondisi sebenarnya. Nilai ROC-AUC yang rendah ini mencerminkan bahwa model belum mampu menangkap pola karakteristik kelas diabetes secara efektif pada dataset yang digunakan.

Secara keseluruhan, hasil evaluasi ini menegaskan bahwa meskipun beberapa model menunjukkan nilai *accuracy* yang lebih tinggi, metrik tersebut tidak cukup untuk menggambarkan kinerja model secara komprehensif dalam konteks prediksi dini diabetes. Oleh karena itu, penggunaan metrik ROC-AUC dan *recall* menjadi sangat penting untuk menilai kemampuan model dalam mendeteksi kelas minoritas, khususnya pada dataset klinis dengan ukuran data terbatas.



Gambar 4. Kurva ROC Perbandingan Algoritma

Berdasarkan hasil pengujian yang ditunjukkan pada Gambar 4, dataset yang digunakan dalam penelitian ini terdiri atas 22 sampel setelah proses pembersihan data, dengan sembilan fitur klinis utama dan satu variabel target berupa status diabetes. Proporsi kelas menunjukkan bahwa 31,8% data merupakan pasien diabetes dan 68,2% merupakan non-diabetes, sehingga terdapat ketidakseimbangan kelas yang berpotensi memengaruhi kinerja model.

3.3 Pembahasan

Untuk menegaskan posisi hasil penelitian ini, pembahasan berikut difokuskan pada perbandingan hasil yang diperoleh dengan penelitian-penelitian sejenis yang telah diuraikan pada bagian pendahuluan. Perbandingan ini dilakukan untuk menilai kesesuaian, perbedaan, serta kontribusi penelitian dalam konteks penerapan *machine learning* untuk prediksi dini diabetes, khususnya pada kondisi data klinis berukuran kecil dan distribusi kelas yang tidak seimbang. Hasil penelitian ini menunjukkan bahwa keterbatasan jumlah data dan distribusi kelas yang tidak seimbang merupakan faktor dominan yang memengaruhi kinerja algoritma *machine learning* dalam prediksi dini diabetes, bahkan lebih besar dibandingkan pemilihan jenis algoritma yang digunakan. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa ukuran dataset dan proporsi kelas sangat berpengaruh terhadap kemampuan model dalam mendeteksi kelas minoritas, khususnya pada konteks data klinis [20][21]. Pada penelitian ini, meskipun beberapa algoritma menghasilkan nilai *accuracy* yang relatif lebih tinggi, metrik evaluasi klinis yang lebih penting, yaitu *recall* dan F1-score, menunjukkan bahwa model belum mampu mendeteksi kasus diabetes secara efektif.

Jika dibandingkan dengan penelitian terdahulu yang telah dibahas pada bagian pendahuluan, khususnya studi prediksi diabetes dan hipertensi dengan dataset berskala besar (lebih dari 17.000 sampel), algoritma *machine learning* pada penelitian tersebut mampu mencapai performa yang lebih stabil dan akurat. Hal ini menunjukkan bahwa ketersediaan data yang memadai memungkinkan model mempelajari pola kelas minoritas dengan lebih baik. Sebaliknya, pada penelitian ini yang menggunakan data klinis berukuran kecil, model cenderung bias terhadap kelas mayoritas, sehingga kemampuan deteksi dini diabetes menjadi terbatas. Perbedaan kondisi ini menegaskan bahwa hasil penelitian



tidak dapat dibandingkan secara langsung hanya berdasarkan nilai akurasi, tetapi harus dilihat dalam konteks karakteristik data yang digunakan.

Penelitian ini juga menunjukkan kesesuaian dengan temuan Zhang et al. yang melaporkan bahwa metode berbasis *ensemble*, seperti Random Forest dan XGBoost, cenderung menunjukkan performa yang lebih baik dibandingkan metode linier. Namun demikian, pada penelitian ini keunggulan tersebut belum tercermin secara signifikan pada metrik *recall* dan ROC-AUC, yang mengindikasikan bahwa kompleksitas algoritma belum mampu mengatasi keterbatasan data. Dengan demikian, penelitian ini memberikan perspektif tambahan bahwa keunggulan algoritma *ensemble* sangat bergantung pada kondisi dan ukuran dataset.

Dari sisi interpretabilitas, hasil penelitian ini menunjukkan bahwa Logistic Regression tetap memiliki keunggulan sebagai model yang mudah dipahami oleh tenaga medis, meskipun performa prediktifnya masih terbatas. Temuan ini memperkuat hasil penelitian sebelumnya yang menekankan adanya *trade-off* antara akurasi dan interpretabilitas dalam penerapan *machine learning* di bidang kesehatan. Oleh karena itu, pemilihan model tidak hanya bergantung pada performa numerik, tetapi juga pada tujuan penggunaan sistem, apakah untuk skrining awal atau sebagai alat bantu diagnosis lanjutan.

Secara keseluruhan, pembahasan ini menegaskan bahwa penelitian ini tidak berfokus pada pencapaian performa prediksi yang tinggi, melainkan pada evaluasi realistis terhadap keterbatasan penerapan *machine learning* pada data klinis berukuran kecil. Kontribusi utama penelitian ini terletak pada penekanan pentingnya pemilihan metrik evaluasi yang relevan secara klinis serta pemahaman terhadap dampak karakteristik data terhadap kinerja model. Dengan pendekatan tersebut, penelitian ini diharapkan dapat menjadi rujukan awal dalam pengembangan sistem prediksi dini diabetes yang lebih andal pada penelitian selanjutnya.

Dengan demikian, hasil dan pembahasan penelitian ini secara langsung menjawab tujuan penelitian yang telah dirumuskan pada bagian pendahuluan, yaitu mengevaluasi kinerja algoritma *machine learning* dalam kondisi keterbatasan data klinis serta menilai implikasi metodologisnya terhadap penerapan prediksi dini diabetes.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja beberapa algoritma *machine learning* dalam melakukan prediksi dini diabetes berbasis data klinis dengan jumlah sampel yang terbatas. Dataset yang digunakan terdiri atas 22 sampel setelah proses pembersihan data, yang selanjutnya dibagi menjadi 17 data latih dan 5 data uji. Kondisi ini merepresentasikan keterbatasan yang umum dijumpai pada data klinis tahap awal, khususnya terkait jumlah data yang kecil dan distribusi kelas yang tidak seimbang, yang secara signifikan memengaruhi kemampuan model dalam mempelajari pola kelas minoritas. Hasil penelitian menunjukkan bahwa seluruh algoritma yang digunakan, yaitu Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan XGBoost, belum mampu mendeteksi kelas diabetes secara efektif pada data uji. Hal ini tercermin dari nilai *precision*, *recall*, dan *F1-score* yang bernilai nol pada seluruh model. Meskipun Random Forest dan XGBoost menghasilkan nilai akurasi sebesar 0,6, nilai tersebut tidak merepresentasikan kemampuan klasifikasi yang sesungguhnya, melainkan dipengaruhi oleh dominasi kelas non-diabetes pada data uji yang sangat terbatas. Temuan ini menegaskan bahwa penggunaan metrik akurasi saja tidak memadai dalam konteks prediksi dini penyakit, dan perlu dilengkapi dengan metrik evaluasi lain yang lebih relevan secara klinis. Keterbatasan utama penelitian ini terletak pada ukuran dataset yang kecil, proporsi kelas yang tidak seimbang, serta belum dilakukannya validasi eksternal menggunakan dataset independen. Keterbatasan tersebut menyebabkan model cenderung bias terhadap kelas mayoritas dan belum mampu menunjukkan kemampuan generalisasi yang baik. Namun demikian, analisis korelasi dan *feature importance* menunjukkan bahwa variabel Glucose, BMI, dan Diabetes Pedigree Function memiliki kontribusi yang relatif lebih besar terhadap risiko diabetes, yang sejalan dengan pengetahuan klinis dan penelitian terdahulu. Kontribusi utama penelitian ini terletak pada penyajian evaluasi yang realistis terhadap penerapan *machine learning* pada data klinis berukuran kecil, serta penekanan pentingnya pemilihan metrik evaluasi yang tepat dalam prediksi dini diabetes. Penelitian ini memberikan pemahaman bahwa keterbatasan performa model tidak selalu disebabkan oleh pemilihan algoritma, melainkan sangat dipengaruhi oleh karakteristik data yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan representatif, menerapkan teknik penyeimbangan kelas, serta melakukan validasi eksternal agar model prediksi dini diabetes yang dikembangkan dapat memiliki kinerja yang lebih andal dan aplikatif dalam konteks layanan kesehatan.

REFERENCES

- [1] H. Zhou, S. Rahman, M. Angelova, C. R. Bruce, and C. Karmakar, "A robust and generalized framework in diabetes classification across heterogeneous environments," *Comput. Biol. Med.*, vol. 186, no. January, p. 109720, 2025, doi: 10.1016/j.combiomed.2025.109720.
- [2] I. Kurniastuti, A. Andini, and M. R. Dwisapta, "Implementation of Neural Network for Classification of Diabetes Mellitus through Finger Nail Image," *Procedia Comput. Sci.*, vol. 234, no. 2023, pp. 1625–1632, 2024, doi: 10.1016/j.procs.2024.03.166.
- [3] M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," *Heliyon*, vol. 10, no. 2, p. e24536, 2024, doi: 10.1016/j.heliyon.2024.e24536.
- [4] J. Wei et al., "Metadatum information and fundus image fusion neural network for hyperuricemia classification in diabetes," *Comput. Methods Programs Biomed.*, vol. 256, no. July, pp. 1–12, 2024, doi: 10.1016/j.cmpb.2024.108382.



- [5] M. Z. Atwany, A. H. Sahyoun, and M. Yaqub, "Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey," *IEEE Access*, vol. 10, pp. 28642–28655, 2022, doi: 10.1109/ACCESS.2022.3157632.
- [6] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab, and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," *IEEE Access*, vol. 9, pp. 43711–43735, 2021, doi: 10.1109/ACCESS.2021.3059343.
- [7] F. Ariska, V. Sihombing, and I. Irmayani, "Student Graduation Predictions Using Comparison of C5.0 Algorithm With Linear Regression," *Sinkron*, vol. 7, no. 1, pp. 256–266, Feb. 2022, doi: 10.33395/sinkron.v7i1.11261.
- [8] S. Tilki, H. B. Dogru, A. A. Hameed, A. Jamil, and J. Rasheed, "Gender Classification using Deep Learning Techniques," *Manchester J. Artif. Intell. Appl. Sci.*, vol. 2, no. May, 2021.
- [9] M. Sinsirimongkhon, S. Arwathananukul, and P. Temdee, "Multi-Class Classification Method with Feature Engineering for Predicting Hypertension with Diabetes," *J. Mob. Multimed.*, vol. 19, no. 3, pp. 799–822, 2023, doi: 10.13052/jmm1550-4646.1937.
- [10] S. Uhl, A. Choure, B. Rouse, A. Loblack, and P. Reaven, "Effectiveness of Continuous Glucose Monitoring on Metrics of Glycemic Control in Type 2 Diabetes Mellitus: A Systematic Review and Meta-analysis of Randomized Controlled Trials," *J. Clin. Endocrinol. Metab.*, vol. 109, no. 4, pp. 1119–1131, 2024, doi: 10.1210/clinem/dgad652.
- [11] S. Apandi *et al.*, "Classification of Lung Diseases Using the Desicison Tree Method," *Formosa J. Sci. Technol.*, vol. 4, no. 1, pp. 393–412, Jan. 2025, doi: 10.55927/fjst.v4i1.13442.
- [12] P. I. Ritharson, K. Raimond, X. A. Mary, J. E. Robert, and A. J., "DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes," *Artif. Intell. Agric.*, vol. 11, pp. 34–49, Mar. 2024, doi: 10.1016/j.aiaa.2023.11.001.
- [13] L. Otero Sanchez *et al.*, "A machine learning-based classification of adult-onset diabetes identifies patients at risk of liver-related complications," *JHEP Reports*, vol. 5, no. 8, p. 100791, 2023, doi: 10.1016/j.jhepr.2023.100791.
- [14] Y. Zhang *et al.*, "A new classification method for gestational diabetes mellitus: a study on the relationship between abnormal blood glucose values at different time points in oral glucose tolerance test and adverse maternal and neonatal outcomes in pregnant women with gest," *AJOG Glob. Reports*, vol. 4, no. 4, p. 100390, 2024, doi: 10.1016/j.xagr.2024.100390.
- [15] F. J. Lara-Abelenda, D. Chushig-Muzo, C. B. Acosta, A. M. Wagner, C. Granja, and C. Soguero-Ruiz, "Evaluating Time Series Classification Models for Nocturnal Hypoglycemia: From Predictive Performance to Environmental Impact," *IEEE Access*, vol. 13, no. September, pp. 150756–150771, 2025, doi: 10.1109/ACCESS.2025.3600917.
- [16] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.
- [17] O. U. Lenz, H. Bollaert, and C. Cornelis, "A unified weighting framework for evaluating nearest neighbour classification," *Fuzzy Sets Syst.*, vol. 519, Nov. 2025, doi: 10.1016/j.fss.2025.109516.
- [18] R. Amalia, A. F. Zaidan, S. Ramadhan, F. Septian, A. M. Aqsha, and P. Rosyani, "Classification of Autoimmune Diseases Using the K-Nearest Neighbors Algorithm," *Formosa J. Sci. Technol.*, vol. 4, no. 1, pp. 337–348, Jan. 2025, doi: 10.55927/fjst.v4i1.13443.
- [19] R. Pambudi, A. R. Harahap, F. D. Saputra, and M. Jusub, "Klasifikasi Penyakit Paru-paru Menggunakan Metode Decision Tree," vol. 3, no. 9, pp. 2397–2402, 2024.
- [20] S. Dwi, Y. Kusuma, H. Al Islami, and D. P. Rosyani, "Penerapan Naive Bayes Untuk Klasifikasi Penyakit Endokrin Pada Pasien Lansia," vol. 5, no. 2, 2024, doi: 10.31284/j.kernel.2024.
- [21] S. Raj, S. Raj, V. Namdeo, and A. Srivastava, "Decoding the gene-disease associations in type 2 diabetes: A curated dataset for text mining-based classification," *Data Br.*, vol. 54, 2024, doi: 10.1016/j.dib.2024.110418.