TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Perbandingan Kinerja XGBoost dan Random Forest Menggunakan SMOTE pada Klasifikasi Diabetes Multi-Kelas

Java Sika Maulana, Safitri Juanita*

Fakultas Teknologi Informasi, Program Studi Sistem Informasi, Universitas Budi Luhur, Jakarta, Indonesia
Email: [1]2212500967@student.budiluhur.ac.id, [2*]safitri.juanita@budiluhur.ac.id
Email Penulis Korespondensi: safitri.juanita@budiluhur.ac.id

Abstrak–Diabetes melitus tipe dua merupakan gangguan metabolik yang membutuhkan sistem deteksi dini untuk mendukung identifikasi status diagnosis secara akurat. Salah satu tantangan utama dalam membangun model klasifikasi berbasis rekam medis klinis adalah ketidakseimbangan kelas data (*class imbalance*), yang berpotensi menyebabkan model lebih cenderung mengenali kelas mayoritas, khususnya pada klasifikasi multi-kelas yang melibatkan kelas Prediabetes dengan proporsi hanya 5,3% dari total data. Kegagalan mengidentifikasi pasien dengan Prediabetes memiliki konsekuensi klinis yang serius karena intervensi dini pada tahap ini masih dapat mencegah perkembangan penyakit menjadi Diabetes. Penelitian ini membandingkan kinerja XGBoost dan Random Forest dalam klasifikasi diabetes multi-kelas (Normal, Prediabetes, Diabetes) menggunakan data klinis dari Medical City Hospital dan Al-Kindy Teaching Hospital, Irak, dengan menerapkan prosedur *Split-First, Resample-Later* untuk mencegah kebocoran data, SMOTE untuk penyeimbangan data latih, serta Grid Search untuk optimasi hiperparameter pada berbagai rasio pembagian data (60:40, 70:30, 80:20, 90:10)**.** Kontribusi penelitian ini adalah memberikan evaluasi komprehensif terhadap kinerja XGBoost dan Random Forest pada data klinis multi-kelas yang tidak seimbang dengan membandingkan performa sebelum dan sesudah penerapan SMOTE pada berbagai rasio pembagian data menggunakan prosedur *Split-First, Resample-Later* untuk mencegah kebocoran informasi dari data uji ke data latih. Hasil eksperimen menunjukkan bahwa Random Forest memiliki performa yang lebih stabil dibandingkan XGBoost pada seluruh skenario pengujian, baik sebelum maupun sesudah penerapan SMOTE. Kedua model mencapai performa terbaik pada rasio pembagian data 80:20, sedangkan SMOTE memberikan peningkatan performa yang signifikan pada XGBoost dengan rasio 60:40. Analisis *feature importance* menunjukkan bahwa HbA1c, BMI, dan AGE merupakan atribut klinis yang paling berpengaruh dalam proses klasifikasi diabetes.

Kata Kunci: Diabetes Melitus; Klasifikasi Multi-Kelas; SMOTE; XGBoost; Random Forest

Abstract–Type 2 diabetes mellitus is a metabolic disorder that requires an early detection system to support accurate diagnosis identification. One of the major challenges in developing classification models based on clinical medical records is class imbalance, which may cause models to be biased toward the majority class, particularly in multiclass classification involving the Prediabetes class, which represents only 5.3% of the total data. Failure to accurately identify the Prediabetes class may have serious clinical consequences, as this stage still provides an opportunity for early intervention to prevent progression to Diabetes. This study compares the performance of Extreme Gradient Boosting (XGBoost) and Random Forest for multiclass diabetes classification (Normal, Prediabetes, and Diabetes) using clinical data obtained from Medical City Hospital and Al-Kindy Teaching Hospital, Iraq. A Split-First, Resample-Later procedure was employed to prevent data leakage, while Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the training data, and Grid Search was used for hyperparameter optimization across different train-test split ratios (60:40, 70:30, 80:20, and 90:10). This study provides a comprehensive evaluation of XGBoost and Random Forest on imbalanced multiclass clinical data by comparing their performance before and after SMOTE application across different train–test split ratios using the Split-First, Resample-Later procedure to prevent information leakage between the training and test set. The experimental results demonstrate that Random Forest achieved more stable performance than XGBoost across all evaluation scenarios, both before and after SMOTE application. Both models achieved their best performance with an 80:20 train–test split ratio, whereas SMOTE significantly improved the performance of XGBoost only under the 60:40 split ratio. Furthermore, feature importance analysis identified HbA1c, BMI, and AGE as the most influential clinical attributes for diabetes classification.

Keywords: Diabetes Mellitus; Multiclass Classification; SMOTE; XGBoost; Random Forest

1. PENDAHULUAN

Diabetes melitus adalah gangguan metabolik yang ditandai dengan hiperglikemia kronis akibat kegagalan pankreas memproduksi insulin yang cukup atau ketidakmampuan tubuh menggunakannya secara efektif (Wongso Prawiro et al., 2025). Prevalensi diabetes terus meningkat secara global, sebanyak 589 juta atau 11,1% orang dewasa usia 20–79 tahun di dunia hidup dengan diabetes, dan lebih dari 4 dari 10 di antaranya tidak menyadari kondisi tersebut. Indonesia berada di peringkat ke-5 dunia dengan jumlah penderita terbesar, yaitu sekitar 20,4 juta jiwa, dengan prevalensi 11,3% dari total populasi dewasa yang melampaui rata-rata kawasan Asia Tenggara sebesar 10,8% (International Diabetes Federation, 2025). Kondisi ini semakin kritis mengingat layanan skrining diabetes di Indonesia masih belum merata, terutama di luar kota besar, ditambah diabetes tidak menunjukkan gejala yang jelas pada tahap awal sehingga pasien baru terdiagnosis setelah mengalami komplikasi serius seperti neuropati, nefropati, dan penyakit kardiovaskular sudah terjadi. Di sisi lain, sebagian besar dari kasus tersebut sebenarnya bisa dicegah jika terdeteksi lebih awal (Goldney et al., 2023). Deteksi dini memungkinkan tenaga medis melakukan tindakan pencegahan dan pengelolaan penyakit sebelum kondisi pasien berkembang menjadi lebih parah. Oleh karena itu, membangun sistem deteksi dini yang akurat menjadi kebutuhan mendesak untuk menekan angka mortalitas dan beban biaya perawatan jangka panjang.

Salah satu pendekatan yang berkembang untuk mengatasi masalah ini adalah pemanfaatan *machine learning*, yang dapat mengenali pola dari data laboratorium yang sudah tersedia tanpa menunggu munculnya gejala klinis. Berbagai algoritma *machine learning* telah digunakan untuk mendukung deteksi dini diabetes, namun performanya

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

sangat dipengaruhi oleh karakteristik data yang digunakan. Data yang digunakan dalam penelitian ini bersumber dari rekam medis pasien di *Medical City Hospital dan Specializes Center for Endocrinology and Diabetes–Al-Kindy Teaching Hospital*, yang memuat informasi klinis dan hasil analisis laboratorium berupa 16 fitur, meliputi kadar gula darah, usia, jenis kelamin, BMI, HbA1c, profil lipid (kolesterol total, LDL, VLDL, HDL, trigliserida), urea, dan rasio kreatinin. Dari total 999 rekam medis, sebanyak 843 sampel (84,4%) termasuk kelas diabetes, 103 sampel (10,3%) normal, dan 53 sampel (5,3%) prediabetes (Rashid, 2026).

Karakteristik dataset tersebut menghadirkan dua tantangan utama, yaitu klasifikasi multi-kelas dan ketidakseimbangan distribusi data (*class imbalance*). Berbeda dengan klasifikasi biner yang hanya membedakan pasien diabetes dan non-diabetes, penelitian ini melibatkan tiga kelas diagnosis, yaitu Normal, Prediabetes, dan Diabetes. Keberadaan kelas prediabetes menjadi sangat penting karena merupakan tahap awal yang masih memungkinkan dilakukan upaya pencegahan sebelum kondisi pasien berkembang menjadi diabetes melalui perubahan gaya hidup maupun penanganan medis yang tepat. Namun, distribusi jumlah sampel pada dataset ini tidak seimbang (*class imbalance*), dengan kelas diabetes mendominasi sebanyak 84,4% total data, sedangkan kelas prediabetes hanya mewakili 5,3%. Ketidakseimbangan distribusi kelas seperti ini merupakan karakteristik yang umum dijumpai pada data klinis dunia nyata dan menjadi tantangan dalam pengembangan model *machine learning*, karena algoritma cenderung lebih mudah mempelajari pola dari kelas mayoritas dibandingkan kelas minoritas (Astofa et al., 2025). Akibatnya, model yang tampak akurat secara keseluruhan justru gagal mendeteksi pasien Prediabetes yang seharusnya mendapat penanganan lebih awal, dan kegagalan ini memiliki konsekuensi klinis yang lebih serius dibanding kesalahan prediksi pada arah sebaliknya (Agyemang et al., 2025).

Oleh karena itu, diperlukan evaluasi terhadap algoritma klasifikasi yang mampu mempertahankan performa secara seimbang pada seluruh kelas diagnosis. Berdasarkan permasalahan tersebut, penelitian ini membandingkan kinerja algoritma XGBoost dan Random Forest pada klasifikasi diabetes multi-kelas menggunakan data klinis yang tidak seimbang. Evaluasi dilakukan untuk mengetahui kemampuan kedua model dalam mengklasifikasikan seluruh kelas diagnosis, khususnya kelas Prediabetes yang memiliki jumlah sampel paling sedikit. Selain itu, SMOTE diterapkan sebagai teknik penyeimbangan data untuk menganalisis pengaruhnya terhadap performa kedua model pada kondisi distribusi kelas yang tidak seimbang (Jang, 2025).

Sejumlah penelitian sebelumnya telah menerapkan berbagai algoritma klasifikasi untuk prediksi diabetes, namun masing-masing memiliki keterbatasan yang belum sepenuhnya diatasi. Penelitian yang mengombinasikan Random Forest dan XGBoost dengan SMOTE menggunakan dataset *Diabetes Prediction* dari *Kaggle* melaporkan akurasi Random Forest hingga 0,97, namun tidak menguji stabilitas model pada berbagai skenario rasio pembagian data (Sembiring Depari et al., 2025). Penelitian lain yang menerapkan metode *tree-based ensemble* pada data PIMA *Indian Diabetes* menemukan bahwa variabel prediktor yang hanya berjumlah 9 fitur membuat model kurang mampu menangkap keragaman pola klinis yang lebih luas (Aghware et al., 2025). Algoritma C4.5 berbasis seleksi fitur PSO yang diuji pada data yang sama hanya mencapai akurasi maksimal 81%, yang menunjukkan keterbatasan pohon keputusan tunggal (Bagus et al., 2022). Random Forest dipilih dalam penelitian ini karena arsitektur *bagging*-nya terbukti tangguh terhadap data tabular medis dengan fitur klinis yang beragam (Vlachas et al., 2022). Sedangkan XGBoost ditetapkan sebagai model pembanding karena kemampuannya menangani data kompleks melalui mekanisme boosting bertahap dengan regularisasi untuk mencegah *overfitting* (Adhi Pratama & Wahyu Utomo, 2026).

Penelitian lain yang mengklasifikasikan diabetes menggunakan KNN yang dikombinasikan dengan metode SMOTE-ENN pada rasio 70:30 dengan metode split dan mencatat akurasi 96,95%, namun penelitian tersebut hanya menangani klasifikasi biner (diabetes dan tidak diabetes) sehingga belum mampu membedakan kelas Prediabetes yang justru penting untuk deteksi dini (Amri et al., 2025). Meskipun SMOTE-ENN membantu menangani ketidakseimbangan data, KNN tetap kurang stabil dan sensitif terhadap pemilihan parameter pada data medis sehingga performanya tidak konsisten untuk klasifikasi multi-kelas dengan distribusi kelas yang tidak seimbang seperti pada penelitian ini (Haris Yunianto & Rosi Subhiyakto, 2025). Pengujian Regresi Logistik pada data diabetes hanya mampu mencapai akurasi 82% pada rasio 80:20 dengan nilai *precision* 0,81 dan *recall* 0,82, yang menunjukkan bahwa model ini belum mampu menangani klasifikasi multi-kelas dengan baik, khususnya ketika distribusi kelas tidak seimbang seperti pada penelitian ini yang memiliki kelas Prediabetes hanya sebesar 5,3% dari total data (Pratama et al., 2023). Kondisi tersebut menyebabkan Regresi Logistik cenderung bias terhadap kelas mayoritas sehingga kemampuannya untuk mendeteksi kelas minoritas seperti Prediabetes menjadi sangat terbatas (Astofa et al., 2025). Penggunaan *Decision Tree* yang interpretatif pada data diabetes menunjukkan hasil yang cukup baik, namun terbukti rentan *overfitting* karena mengandalkan satu pohon keputusan tunggal (Prayoga & Ridla, 2026). Kelemahan ini sejalan dengan temuan bahwa algoritma berbasis pohon tunggal memiliki keterbatasan dalam menangani kompleksitas data klinis (Syahla et al., 2025). Ketiga keterbatasan tersebut, yaitu efisiensi komputasi, sensitivitas terhadap ketidakseimbangan kelas, dan *overfitting*, menjadi pertimbangan utama pemilihan XGBoost dan Random Forest dalam penelitian ini. XGBoost mengatasi masalah efisiensi dan bias kelas melalui mekanisme *gradient boosting* yang adaptif (Utomo et al., 2025), sementara Random Forest meminimalkan risiko *overfitting* dengan mengagregasikan banyak pohon keputusan secara paralel (Syahla et al., 2025). Penerapan SMOTE pada data latih melengkapi pendekatan ini untuk memastikan kedua model belajar dari distribusi kelas yang seimbang tanpa membuang informasi dari data asli (Iftikhar et al., 2026).

Berdasarkan kajian penelitian sebelumnya, terdapat dua celah penelitian utama yang menjadi dasar dilakukannya penelitian ini, yaitu kurangnya pengujian stabilitas performa model pada berbagai variasi rasio pembagian data pada penelitian-penelitian sebelumnya, sehingga belum diketahui secara pasti pada rasio split berapa XGBoost dan Random

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Forest mampu mempertahankan performanya secara konsisten pada data klinis multi-kelas yang tidak seimbang dan risiko *data leakage* akibat penerapan SMOTE sebelum proses pemisahan data latih dan data uji, yang berpotensi menghasilkan estimasi performa yang lebih tinggi daripada kondisi sebenarnya karena informasi dari data uji dapat ikut memengaruhi proses pelatihan model.

Sehingga, tujuan penelitian ini adalah membandingkan kinerja XGBoost dan Random Forest pada klasifikasi diabetes multi-kelas dengan data klinis tidak seimbang, melalui penerapan prosedur *Split-First, Resample-Later*, yaitu data dipisahkan terlebih dahulu menjadi data latih dan data uji sebelum SMOTE diterapkan pada data latih untuk mencegah kebocoran informasi. Penelitian ini juga mengevaluasi performa model pada pengujian beberapa skenario pembagian data untuk menganalisis konsistensi performa XGBoost dan Random Forest pada klasifikasi diabetes multi-kelas. Selain itu, optimasi hiperparameter menggunakan Grid Search diterapkan untuk memperoleh konfigurasi model terbaik, sekaligus mengidentifikasi fitur/atribut klinis yang paling berpengaruh terhadap proses klasifikasi melalui analisis *feature importance*. Kontribusi penelitian ini adalah memberikan evaluasi komprehensif terhadap kinerja XGBoost dan Random Forest pada data klinis multi-kelas yang tidak seimbang dengan membandingkan performa sebelum dan sesudah penerapan SMOTE pada berbagai rasio pembagian data menggunakan prosedur *Split-First, Resample-Later* untuk mencegah kebocoran informasi dari data uji ke data latih.

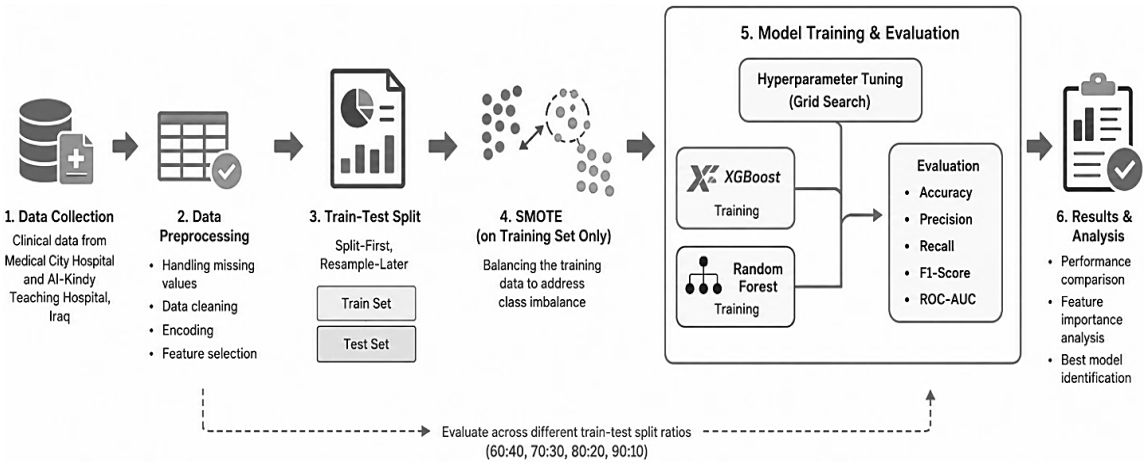
2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini membandingkan performa dari algoritma XGBoost dan Random Forest untuk klasifikasi diabetes dengan multi-kelas melalui eksperimen pembagian data (*split*) menggunakan multi-rasio. Dataset berasal dari rekam medis klinis *Medical City Hospital* dan *Al-Kindy Teaching Hospital*, Irak (Rashid, 2026), dengan tiga kelas target yaitu Normal, Prediabetes , dan Diabetes . Pengujian dilakukan pada multi-rasio (60:40, 70:30, 80:20, 90:10) dengan SMOTE diterapkan hanya pada data latih menggunakan prosedur *Split-First, Resample-Later* untuk mencegah *data leakage*. Hiperparameter dioptimasi menggunakan *GridSearchCV* dengan *3-Fold Cross Validation*, dan evaluasi dilakukan melalui *confusion matrix*, *classification report*, serta analisis *feature importance*. Alur tahapan penelitian disajikan pada Gambar 1.

2.2 Tahapan Penelitian

Penelitian ini dilaksanakan melalui enam tahapan utama secara berurutan, mulai dari pengumpulan data hingga analisis hasil akhir. Alur tahapan penelitian secara lengkap disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, penelitian ini terdiri atas enam tahapan, yaitu: (1) *Data Collection*, (2) *Data Preprocessing*, (3) *Train-Test Split*, (4) SMOTE, (5) Model *Training & Evaluation*, dan (6) *Results & Analysis.* Keenam tahapan tersebut diuraikan secara rinci sebagai berikut.

a. Pengumpulan Data (*Data Collection*)
Dataset yang digunakan bersumber dari *Laboratorium Medical City Hospita*l dan *Center for Endocrinology and Diabetes di Al-Kindy Teaching Hospital* (Rashid, 2026), berupa rekam medis klinis yang mencakup profil biokimia dan informasi demografis pasien. Data diklasifikasikan ke dalam tiga kelas Diabetes, Normal, dan Prediabetes. Fitur dalam dataset terbagi menjadi dua kelompok. Pertama, parameter laboratorium yang meliputi kadar gula darah, HbA1c, kreatinin (Cr), urea, kolesterol total, *Low-Density Lipoprotein* (LDL), *Very Low-Density Lipoprotein* (VLDL), trigliserida (TG), dan *High-Density Lipoprotein* (HDL). Kedua, fitur demografis dan fisik yang terdiri dari usia, jenis kelamin, dan indeks massa tubuh (BMI). Rincian seluruh fitur beserta karakteristik statistiknya disajikan pada Tabel 1, yang menunjukkan bahwa dataset terdiri atas 14 kolom, mencakup dua fitur identitas (ID dan

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

No_pation), satu fitur kategorik (Gender), sepuluh fitur numerik hasil pemeriksaan laboratorium dan fisik, serta satu variabel target (Class) dengan tiga label diagnosis (N = Normal, P = Prediabetes, Y = Diabetes).

Tabel 1. Deskripsi Fitur Dataset

Fitur	Deskripsi	Tipe Data	Satuan/Format
ID	Nomor Urut	Numerik	-
No_pation	Nomor Pasien	Numerik	-
Gender	Jenis Kelamin	Kategorik	Kategorikal
Age	Usia	Numerik	Tahun
Urea	Kadar urea	Numerik	mg/dL
Cr	Rasio kreatinin (Cr) dalam darah	Numerik	mg/dL
HbA1c	Kadar hemoglobin terglikasi	Numerik	Persentase (%)
Chol	kolesterol total	Numerik	mg/dL
TG	Cadangan energi	Numerik	mg/dL
HDL	Kadar kolesterol baik	Numerik	mg/dL
LDL	Kadar kolesterol jahat	Numerik	mg/dL
VLDL	Pembawa lemak	Numerik	mg/Dl
BMI	Indeks massa tubuh pasien	Numerik	Kg/m^2
Class	Status diabetes	Target Label	N, P, Y

Berdasarkan Tabel 1, fitur-fitur tersebut mencakup parameter laboratorium seperti HbA1c, kadar gula darah, dan profil lipid, serta fitur demografis seperti usia, jenis kelamin, dan BMI, yang keseluruhannya akan digunakan sebagai input dalam proses klasifikasi tiga kelas diagnosis diabetes.

b. Pemrosesan Data (*Data Preprocessing*)

Tahap ini bertujuan memastikan kualitas data sebelum proses pemodelan. Fitur identitas pasien seperti nomor pasien (*No_Pation*) dan ID dihapus karena tidak relevan terhadap klasifikasi. Selanjutnya, baris yang mengandung nilai kosong (*missing value*) dihapus menggunakan fungsi *.dropna*(). Dataset hasil pembersihan kemudian digunakan pada seluruh tahap berikutnya. Struktur data disajikan pada Tabel 2, yang menampilkan contoh sebelas fitur numerik (Gender hingga BMI) beserta label kelas (CLASS). Selanjutnya, dilakukan transformasi data dengan mengonversi fitur kategorikal menjadi numerik agar dapat diproses oleh algoritma klasifikasi. Fitur *Gender* dikonversi menjadi representasi numerik dengan ketentuan nilai 1 untuk pria (M) dan 0 untuk wanita (F). Variabel target *CLASS* juga ditransformasikan menjadi tiga kelas diagnosis, yaitu Normal dipetakan menjadi nilai 0, Prediabetes menjadi nilai 1, dan Diabetes menjadi nilai 2. Dataset hasil transformasi ditampilkan pada Tabel 2, dan digunakan pada seluruh tahap pemodelan berikutnya. Seluruh proses pembersihan dan transformasi data diimplementasikan menggunakan Python.

Tabel 2. Sampel dataset publik berisi data rekam medis klinis

Gender	Age	Urea	Cr	HbA1c	Chol	TG	HDL	LDL	VLDL	BMI	CLASS
0.0	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	0
1.0	26	4.5	62	4.9	3.7	1.4	1.1	2.1	0.6	23.0	0
0.0	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	0
0.0	50	4.7	46	4.9	4.2	0.9	2.4	1.4	0.5	24.0	0
1.0	33	7.1	46	4.9	4.9	1.0	0.8	2.0	0.4	21.0	0
…	…	…	…	…	…	…	…	…	…	…	…
1.0	38	5.8	59	6.7	5.3	2.0	1.6	2.9	14.0	40.5	2
1.0	54	5.0	67	6.9	3.8	1.7	1.1	2.0	0.7	33.0	2

Berdasarkan Tabel 2, data telah sepenuhnya dikonversi ke dalam format numerik dan siap digunakan pada tahap pemodelan berikutnya.

c. Train-Test Split

Sebelum tahap penyeimbangan data dilakukan, dataset terlebih dahulu dibagi menjadi data latih (*train set*) dan data uji (*test set*) menggunakan prosedur *Split-First, Resample-Later*. Pembagian data menggunakan fungsi train_test_split() dengan parameter stratify=y agar distribusi tiga kelas, yaitu Normal, Prediabetes, dan Diabetes, tetap proporsional pada kedua subset. Skenario pembagian data dilakukan pada empat rasio, yaitu 60:40, 70:30, 80:20, dan 90:10, untuk mengevaluasi konsistensi performa model pada berbagai proporsi volume data latih.

d. SMOTE

Distribusi kelas pada dataset ini tidak seimbang, khususnya kelas Prediabetes yang hanya mewakili 5,3% dari total data, diperlukan teknik penyeimbangan data sebelum model dilatih. Penelitian ini menerapkan SMOTE (*Synthetic Minority Over-sampling Technique*), yaitu teknik *oversampling* yang membangkitkan sampel sintetis baru pada kelas minoritas melalui interpolasi antar sampel minoritas yang berdekatan menggunakan K-Nearest Neighbor, sehingga model dapat belajar dari distribusi kelas yang lebih seimbang tanpa membuang informasi dari data asli (Iftikhar et al., 2026). Penerapan SMOTE dibatasi hanya pada data latih menggunakan prosedur *Split-First*,

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Resample-Later untuk mencegah *data leakage*, sehingga data uji tetap dalam kondisi aslinya di seluruh skenario rasio.

e. Model *Training & Evaluation*
Tahap ini mencakup proses pemodelan, optimasi hiperparameter, dan evaluasi hasil prediksi. Penelitian ini menggunakan dua algoritma *ensemble learning* berbasis pohon, yaitu XGBoost dan Random Forest, yang dipilih karena karakteristiknya saling melengkapi. XGBoost membangun pohon keputusan secara sekuensial di mana setiap pohon memperbaiki kesalahan pohon sebelumnya melalui mekanisme *gradient descent*, dilengkapi regularisasi L1 dan L2 untuk mencegah *overfitting* (Adhi Pratama & Wahyu Utomo, 2026). Sementara itu, Random Forest membangun pohon secara paralel menggunakan subset data dan fitur yang dipilih secara acak, lalu menggabungkan seluruh prediksi melalui *voting* mayoritas sehingga lebih stabil dan tahan terhadap *overfitting* pada data tabular medis (Vlachas et al., 2022). Kedua model dijalankan dengan parameter *default* untuk menghasilkan *baseline* sebagai acuan perbandingan sebelum optimasi diterapkan. Data klinis digunakan dalam bentuk aslinya tanpa normalisasi atau standardisasi, karena algoritma berbasis pohon keputusan bersifat non-parametrik dan tidak sensitif terhadap perbedaan skala fitur. Selanjutnya, optimasi hiperparameter dilakukan menggunakan *GridSearchCV* dengan skema *3-Fold Cross Validation*. Ruang pencarian parameter untuk Random Forest mencakup jumlah pohon (*n_estimators*), kedalaman maksimum (*max_depth*), dan kriteria pemisahan (*criterion*), sedangkan untuk XGBoost meliputi tingkat pembelajaran (*learning_rate*), kedalaman pohon (*max_depth*), dan jumlah estimator (*n_estimators*). Parameter terbaik yang diperoleh digunakan untuk meminimalkan risiko *overfitting* dan meningkatkan generalisasi model (Zhou et al., 2025). Setelah parameter terbaik diperoleh, XGBoost dan Random Forest dilatih ulang pada data latih yang telah diseimbangkan dengan SMOTE, kemudian diuji pada data uji untuk setiap skenario rasio (60:40, 70:30, 80:20, 90:10), sebagaimana ditunjukkan pada Gambar 1. Prediksi model pada data uji selanjutnya dievaluasi menggunakan *confusion matrix* untuk memperoleh nilai akurasi, *precision*, *recall*, dan *F1-score* (Susanto & Cahyana, 2025).

f. Hasil dan Analisis (*Results & Analysis*)
Tahap ini mencakup evaluasi performa model dan interpretasi hasil klasifikasi. Performa dari seluruh skenario rasio dibandingkan menggunakan grafik garis untuk melihat konsistensi dan stabilitas kedua model. Evaluasi per kelas diagnosis dilakukan melalui *classification report* yang memuat nilai *precision*, *recall*, dan *F1-score*, serta *confusion matrix* untuk melihat distribusi kesalahan prediksi. Analisis hubungan antar fitur klinis dilakukan menggunakan korelasi Pearson dengan rentang nilai -1,00 hingga +1,00, yang mengukur kekuatan hubungan linear antara setiap fitur terhadap variabel target *CLASS*. Hasil korelasi divisualisasikan dalam bentuk *heatmap*.

Kontribusi masing-masing fitur dianalisis menggunakan *feature importance*, yaitu *Gini Importance* pada Random Forest dan *Gain* pada XGBoost. *Gini Importance* mengukur seberapa besar suatu fitur mampu menurunkan ketidakmurnian (*impurity*) data di setiap simpul pohon karena semakin besar penurunannya, semakin penting fitur tersebut (Alfarisi et al., 2025). Sedangkan metrik *Gain* pada XGBoost mengukur rata-rata perolehan informasi yang dihasilkan setiap kali fitur digunakan sebagai titik pemisahan dalam struktur *gradient boosting* (Munshi et al., 2025). Perbandingan hasil *feature importance* dari kedua model dilakukan untuk mengidentifikasi atribut yang paling berpengaruh dalam proses klasifikasi serta mengevaluasi konsistensi fitur yang digunakan oleh masing-masing model. Selanjutnya, lima fitur dengan nilai kepentingan tertinggi pada setiap skenario rasio dianalisis dan diinterpretasikan berdasarkan relevansinya terhadap diagnosis diabetes.

3. HASIL DAN PEMBAHASAN

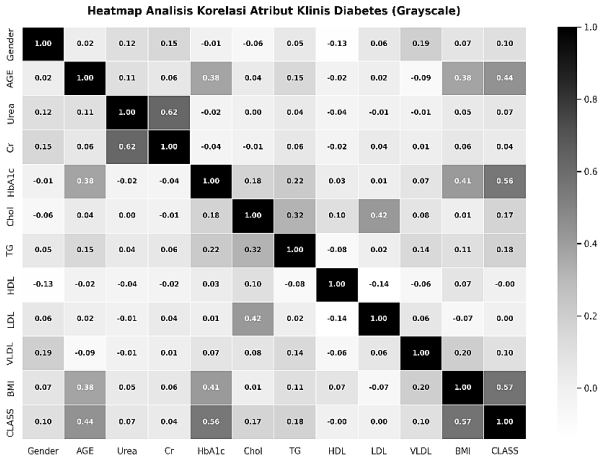
Bagian ini menyajikan hasil klasifikasi multi-kelas risiko pada penyakit Diabetes menggunakan XGBoost dan Random Forest. Performa kedua model dievaluasi melalui *confusion matrix* pada berbagai skenario rasio, dilanjutkan dengan analisis *feature importance* untuk mengidentifikasi fitur klinis yang paling berpengaruh terhadap keputusan klasifikasi.

3.1 Analisis Korelasi Antar Fitur Klinis

Hubungan antara setiap fitur klinis dan variabel target *CLASS* dianalisis menggunakan *koefisien korelasi Pearson*. Analisis ini bertujuan untuk memahami sejauh mana masing-masing fitur berkontribusi secara linear terhadap status diagnosis, sehingga dapat memberikan gambaran awal mengenai fitur mana yang berpotensi paling berpengaruh dalam proses klasifikasi. Hasil analisis divisualisasikan dalam bentuk *heatmap* pada Gambar 2. Berdasarkan *heatmap* pada Gambar 2, BMI dan HbA1c memiliki korelasi positif tertinggi terhadap variabel target *Class*, dengan nilai koefisien masing-masing 0,57 dan 0,56. Artinya, peningkatan indeks massa tubuh (BMI), dan kadar gula darah jangka panjang (HbA1c) cenderung sejalan dengan peningkatan status diagnosis dari Normal ke Prediabetes hingga Diabetes.

Sebaliknya, sebagian besar fitur lainnya menunjukkan korelasi rendah terhadap *Class*. Fitur fungsi ginjal, yaitu Urea (0,07) dan kreatinin/Cr (0,04) hampir tidak berkorelasi dengan target, meskipun keduanya berkorelasi cukup kuat satu sama lain (0,62). Pola serupa terlihat pada fitur lemak darah (lipid) seperti kolesterol total (Chol). Rendahnya korelasi antarfitur terhadap target justru menguntungkan pada setiap fitur membawa informasi yang berbeda dan saling melengkapi, sehingga memberi kedua model lebih banyak pola yang dapat dipelajari selama klasifikasi.

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339



Gambar 2. Heatmap Korelasi Fitur Klinis Diabetes

3.2 Hasil Optimasi Hiperparameter

Pencarian parameter terbaik dilakukan menggunakan *GridSearchCV* pada data latih yang telah diseimbangkan dengan SMOTE. GridSearchCV bekerja dengan mencoba seluruh kombinasi parameter yang telah ditentukan dalam ruang pencarian, kemudian memilih kombinasi yang menghasilkan performa terbaik berdasarkan rata-rata skor validasi dari 3-*Fold Cross Validation*. Pendekatan ini memastikan parameter yang dipilih tidak hanya optimal pada satu subset data, tetapi konsisten pada berbagai partisi data latih sehingga risiko *overfitting* dapat diminimalkan. Parameter optimal yang diperoleh untuk Random Forest disajikan pada Tabel 3, sedangkan untuk XGBoost pada Tabel 4. Hasil ini selanjutnya digunakan sebagai konfigurasi model pada eksperimen multi-rasio.

3.2.1 Hiperparameter Optimal Random Forest

Grid Search menghasilkan parameter optimal untuk Random Forest jumlah pohon (*n_estimators*) sebesar 100, kriteria pemisahan sebesar *entropy*, dan kedalaman maksimum (*max_depth*) sebesar 10. Nilai max_depth sebesar 10 dipilih karena memberikan keseimbangan antara kemampuan model dalam menangkap pola data dan mencegah pohon tumbuh terlalu dalam yang berpotensi menyebabkan *overfitting*. Parameter ini digunakan sebagai konfigurasi model pada seluruh skenario pengujian, sebagaimana disajikan pada Tabel 3.

Tabel 3. Optimalisasi Hiperparameter Hasil Ekstraksi GridSearch Random Forest

Hiperparameter	Rentang nilai uji	Nilai terbaik
n_estimators	[50,100,200]	100
max_depth	[*none*,10,20]	10
Criterion	[‘*gini*’,’*entropy*’]	*Entropy*

Berdasarkan Tabel 3, kombinasi parameter terbaik Random Forest menggunakan 100 pohon dengan kedalaman maksimum 10 dan kriteria pemisahan *entropy*, yang selanjutnya digunakan pada seluruh skenario pengujian.

3.2.2 Hiperparameter Optimal *Extreme Gradient Boosting* (XGBoost)

Grid Search menghasilkan parameter optimal untuk XGBoost: *learning rate* sebesar 0,1, kedalaman maksimum (*max_depth*) sebesar 3, dan jumlah pohon (*n_estimators*) sebesar 100. Nilai max_depth yang relatif kecil, yaitu 3, menunjukkan bahwa XGBoost cukup menggunakan pohon yang dangkal namun banyak secara iteratif untuk membangun model yang kuat, sejalan dengan prinsip *gradient boosting* yang mengandalkan akumulasi pohon lemah. Parameter ini digunakan sebagai konfigurasi model pada seluruh skenario pengujian, sebagaimana disajikan pada Tabel 4.

Tabel 4. Optimalisasi Hiperparameter Hasil Ekstraksi GridSearch XGBoost

Hiperparameter	Rentang nilai uji	Nilai terbaik
n_estimators	[50,100,150]	100
max_depth	[3,6,9]	3
learning_rate	[0.1,0.2]	0.1

Berdasarkan Tabel 4, kombinasi parameter terbaik XGBoost menggunakan 100 pohon dengan kedalaman maksimum 3 dan *learning rate* 0,1, yang selanjutnya digunakan pada seluruh skenario pengujian.

3.3 Perbandingan Kinerja XGBoost dan Random Forest pada Skenario Multi-Rasio

Sebelum proses pemodelan, dataset dibagi empat skenario (multi-rasio) data latih dan data uji. Distribusi jumlah sampel pada masing-masing kelas untuk setiap skenario ditampilkan pada Tabel 5.

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Tabel 5. Jumlah Sampel Per Kelas pada Skenario Multi-Rasio

Rasio	Jumlah Data Latih						Jumlah Data Uji		
	Sebelum SMOTE			Sesudah SMOTE					
	Normal	Prediabetes	Diabetes	Normal	Prediabetes	Diabetes	Normal	Prediabetes	Diabetes
60:40	62	32	505	505	505	505	41	21	338
70:30	72	37	590	590	590	590	31	16	253
80:20	82	43	674	674	674	674	21	10	169
90:10	93	48	758	758	758	758	10	5	85

Berdasarkan Tabel 5, penerapan SMOTE berhasil menyeimbangkan jumlah sampel pada ketiga kelas di data latih menjadi sama rata di seluruh skenario rasio, sedangkan data uji dibiarkan dalam kondisi aslinya untuk tetap mencerminkan distribusi data nyata. Hal ini penting agar evaluasi performa model tidak bias dan benar-benar mencerminkan kemampuan model dalam menghadapi data yang tidak seimbang

3.3.1 Perbandingan Kinerja XGBoost dan Random Forest Sebelum dan Sesudah SMOTE

Pengujian awal dilakukan pada data asli tanpa penyeimbangan sebagai *baseline*. Dataset memiliki distribusi kelas yang tidak seimbang sebagaimana ditunjukkan pada Tabel 5, dengan contoh data pada Tabel 2. Kedua model dijalankan menggunakan parameter *default*, dan hasil perbandingan akurasi pada seluruh skenario rasio disajikan pada Tabel 6.

Tabel 6. Perbandingan Nilai Akurasi Model XGBoost dan Random Forest

Algoritma	Skenario rasio	Akurasi	
		Sebelum SMOTE	Setelah SMOTE
XGBoost	60:40	0.9850	0.9875
	70:30	0.9933	0.9867
	80:20	0.9900	**0.9950**
	90:10	0.9900	0.9900
Random Forest	60:40	0.9850	0.9800
	70:30	0.9900	0.9800
	80:20	**0.9950**	**0.9950**
	90:10	0.9900	0.9900

*Angka yang ditebalkan performa terbaik

Berdasarkan Tabel 6, kedua model mencatat akurasi tinggi di seluruh skenario rasio, berkisar antara 0,9800 hingga 0,9950. Sebelum SMOTE, akurasi dari algoritma Random Forest lebih baik dari XGBoost dengan skor 0,9950 pada rasio 80:20, sementara XGBoost hanya mencapai 0,9933 pada rasio 70:30. Perbedaan ini wajar karena mekanisme *bagging* pada Random Forest memang dirancang untuk stabil meski data tidak seimbang, karena setiap pohon dilatih pada subset data yang berbeda sehingga pengaruh kelas mayoritas tidak menumpuk pada satu arah prediksi.

Setelah penerapan SMOTE, nilai akurasi Random Forest pada rasio 80:20 tetap sebesar 0.9950, menunjukkan bahwa model ini sudah mampu mengenali pola ketiga kelas dengan baik bahkan sebelum penyeimbangan data dilakukan. Sebaliknya, XGBoost mengalami peningkatan akurasi setelah penerapan SMOTE dan mencapai nilai yang sama dengan Random Forest pada rasio 80:20, yaitu sebesar 0.9950. Dengan demikian, Random Forest dinilai lebih konsisten dan tidak bergantung pada penyeimbangan data untuk mencapai performa terbaiknya. Rincian evaluasi per kelas disajikan pada Tabel 7.

Tabel 7. Matriks Evaluasi Klasifikasi Per Kelas pada Data Klinis dengan Model XGBoost dan Random Forest

Algoritma	Rasio	Kelas	Sebelum SMOTE			Setelah SMOTE		
			Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*
XGBoost	60:40	Normal	1.0000	0.9268	0.9620	0.9512	0.9512	0.9512
		Prediabetes	1.0000	0.8571	0.9231	1.0000	0.9524	0.9756
		Diabetes	0.9826	1.0000	0.9912	0.9912	0.9941	0.9926
	70:30	Normal	1.0000	0.9677	0.9836	0.9375	0.9677	0.9524
		Prediabetes	1.0000	0.9375	0.9677	1.0000	0.9375	0.9677
		Diabetes	0.9922	1.0000	0.9961	0.9921	0.9921	0.9921
	80:20	Normal	1.0000	0.9524	0.9756	1.0000	0.9524	0.9756
		Prediabetes	1.0000	0.9000	0.9474	1.0000	0.9000	0.9474
		Diabetes	0.9883	1.0000	0.9898	0.9883	0.9883	0.9941
	90:10	Normal	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		Prediabetes	1.0000	0.8000	0.8889	1.0000	0.8000	0.8889
		Diabetes	0.9884	1.0000	0.9942	0.9884	1.0000	0.9942
Random Forest	60:40	Normal	0.9524	0.9756	0.9639	0.9111	1.0000	0.9535
		Prediabetes	1.0000	0.8571	0.9231	1.0000	0.8571	0.9231

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Algoritma	Rasio	Kelas	Sebelum SMOTE			Setelah SMOTE		
			Precision	*Recall*	*F1-Score*	*Precision*	*Recall*	*F1-Score*
	70:30	Diabetes	0.9882	0.9941	0.9912	0.9911	0.9882	0.9896
		Normal	1.0000	0.9677	0.9836	0.8857	1.0000	0.9394
		Prediabetes	1.0000	0.8750	0.9333	1.0000	0.8750	0.9333
	80:20	Diabetes	0.9883	1.0000	0.9941	0.9920	0.9842	0.9881
		Normal	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		Prediabetes	1.0000	0.9000	0.9474	1.0000	0.9000	0.9474
	90:10	Diabetes	0.9941	1.0000	0.9971	0.9941	1.0000	0.9971
		Normal	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
		Prediabetes	1.0000	0.8000	0.8889	1.0000	0.8000	0.8889
		Diabetes	0.9884	1.0000	0.9942	0.9884	1.0000	0.9942

*Angka yang ditebalkan performa terkecil

Berdasarkan Tabel 7, berikut ini adalah analisis performa model pada setiap kelas.

a. Pada kelas Normal, Sebelum SMOTE, kedua model mencatat precision sempurna 1,0000 di hampir seluruh skenario, kecuali Random Forest pada rasio 60:40 yang mencatat *precision* terendah 0,9524. XGBoost mencatat *F1-Score* terendah 0,9620 pada rasio 60:40, sedangkan Random Forest mempertahankan *recall* dan *F1-Score* sempurna 1,0000 pada rasio 80:20. Setelah SMOTE, *precision* XGBoost pada rasio 60:40 justru turun menjadi 0,9512, sementara Random Forest tetap stabil. Ini menunjukkan SMOTE tidak memberi manfaat pada kelas Normal, bahkan sedikit menurunkan performa XGBoost. Sehingga dapat disimpulkan bahwa, skenario terbaik untuk mendeteksi kelas Normal adalah Random Forest rasio 80:20 dengan *precision*, *recall*, dan *F1-Score* sempurna 1,0000 baik sebelum maupun sesudah SMOTE.
b. Pada kelas Prediabetes, kedua model mencatat *precision* sempurna 1,0000 di seluruh skenario, namun *recall*-nya rendah. *Recall* terendah 0,8000 terjadi pada rasio 90:10 di kedua model, baik sebelum maupun sesudah SMOTE, yang berarti masih ada pasien Prediabetes yang tidak berhasil terdeteksi dan SMOTE tidak memberikan perbaikan sama sekali pada kelas ini. Skenario terbaik untuk mendeteksi Prediabetes adalah Random Forest rasio 80:20 *dengan F1-Score* 0,9474, *precision* 1,0000, dan *recall* 0,9000.
c. Pada kelas Diabetes, kedua model menunjukkan performa yang konsisten karena jumlah sampelnya lebih besar dibandingkan kelas lain sehingga membuat polanya mudah dipelajari. Nilai *Recall* 1,0000 tercatat di hampir seluruh skenario pada kedua model, baik sebelum maupun sesudah SMOTE. Skenario terbaik adalah Random Forest rasio 80:20 sebelum SMOTE dengan *F1-Score* tertinggi 0,9971, *precision* 0,9941, dan *recall* 1,0000. Pada skenario yang sama, XGBoost menghasilkan *F1-Score* 0,9898. Hasil ini menunjukkan bahwa Random Forest skenario 80:20, baik sebelum maupun sesudah SMOTE lebih unggul dalam mengklasifikasikan kelas Diabetes.

Secara keseluruhan, ketidakseimbangan data paling berdampak pada kelas Prediabetes yang ditunjukkan dengan nilai *recall* terendah 0,8000 pada rasio 90:10, yang berarti masih ada pasien Prediabetes yang tidak berhasil terdeteksi oleh kedua model. Penerapan SMOTE tidak menunjukkan terjadinya meningkatkan performa secara konsisten pada kedua model. Random Forest lebih stabil dibanding XGBoost di seluruh skenario, dengan skenario terbaik pada rasio 80:20 yang menghasilkan *F1-Score* tertinggi dan paling konsisten di seluruh kelas diagnosis.

3.3.2 Analisis Pengaruh SMOTE terhadap Performa Model

Untuk mengukur dampak SMOTE secara konkret, dihitung selisih (Δ) nilai *precision*, *recall*, dan *F1-score* sebelum dan sesudah penyeimbangan data. Hasil perbandingan untuk XGBoost disajikan pada Tabel 8, sedangkan Random Forest pada Tabel 9. Nilai yang digunakan merupakan rata-rata dari seluruh kelas diagnosis. Pada tahap ini, evaluasi pengaruh SMOTE pada kedua model difokuskan pada nilai *F1-Score* karena metrik ini mampu memberikan gambaran yang lebih seimbang antara *precision* dan *recall* pada permasalahan klasifikasi multi-kelas dengan distribusi data yang tidak merata.

Tabel 8. Analisis Perubahan Performa Model XGBoost Terhadap SMOTE

Rasio	Akurasi		Δ	*Precision*		Δ	*Recall*		Δ	*F1-Score*		Δ
	A*	B**		A*	B**		A*	B**		A*	B**	
60:40	0.9850	0.9875	+0.0025	0.9942	0.9808	-0.0134	0.9280	0.9659	+0.0379	0.9588	0.9731	+0.0143
70:30	0.9933	0.9867	-0.0066	0.9974	0.9765	-0.0209	0.9684	0.9658	-0.0026	0.9825	0.9707	-0.0118
80:20	0.9900	0.9950	+0.0050	0.9961	0.9961	0.0000	0.9508	0.9508	0.0000	0.9724	0.9724	0.0000
90:10	0.9900	0.9900	0.0000	0.9961	0.9961	0.0000	0.9333	0.9333	0.0000	0.9610	0.9610	0.0000

*A=Sebelum; **B=Sesudah; Δ = Selisih

Berdasarkan Tabel 8, hasil menunjukkan bahwa peningkatan *F1-Score* hanya terjadi pada rasio 60:40, yaitu dari 0,9588 menjadi 0,9731 atau meningkat sebesar 0,0143. Hal ini dikuatkan juga dengan menggunakan SMOTE dapat meningkatkan *recall* XGBoost sebesar +0,0379 pada rasio 60:40, dari 0,9280 menjadi 0,9659 yang merupakan peningkatan terbesar dibandingkan metrik lainnya. Sehingga dapat disimpulkan bahwa meningkatnya nilai *F1-Score* pada skenario SMOTE terutama dipengaruhi oleh meningkatnya nilai recall, meskipun terjadi sedikit penurunan pada

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

skor *precision*. Hal ini mengindikasikan bahwa pada kondisi data latih yang terbatas seperti rasio 60:40, SMOTE membantu XGBoost mengenali pola kelas minoritas yang sebelumnya kurang terwakili, sehingga kemampuan model dalam mendeteksi seluruh kelas meningkat. Peningkatan nilai *F1-Score* dan *recall* membuktikan bahwa model XGBoost yang menggunakan skenario SMOTE mampu meningkatkan kemampuan model dalam mengklasifikasi data multi-kelas yang tidak seimbang.

Tabel 9. Analisis Perubahan Performa Model Random Forest Terhadap SMOTE

Rasio	Akurasi		Δ	*Precision*		Δ*	*Recall*		Δ*	*F1-Score*		Δ*
	A*	B**		A*	B**		A*	B**		A*	B**	
60:40	0.9850	0.9800	-0.0050	0.9802	0.9674	-0.0128	0.9423	0.9484	+0.0061	0.9594	0.9554	-0.0040
70:30	0.9900	0.9800	-0.0100	0.9961	0.9592	-0.0369	0.9476	0.9531	+0.0055	0.9703	0.9536	-0.0167
80:20	0.9950	0.9950	0.0000	0.9980	0.9980	0.0000	0.9667	0.9667	0.0000	0.9815	0.9815	0.0000
90:10	0.9900	0.9900	0.0000	0.9961	0.9961	0.0000	0.9333	0.9333	0.0000	0.9610	0.9610	0.0000

*A=Sebelum; **B=Sesudah; Δ = Selisih

Berdasarkan Tabel 9, hasil menunjukkan bahwa SMOTE tidak menghasilkan peningkatan *F1-Score* pada Random Forest di seluruh skenario rasio. Pada rasio 60:40, *F1-Score* justru turun dari 0,9594 menjadi 0,9554 atau menurun sebesar -0,0040, dan pada rasio 70:30 penurunan lebih dalam terjadi dari 0,9703 menjadi 0,9536 atau sebesar -0,0167. Penurunan ini dipicu oleh turunnya nilai *precision*, yaitu -0,0128 pada rasio 60:40 dan -0,0369 pada rasio 70:30, meskipun *recall* sedikit meningkat sebesar +0,0061 pada rasio 60:40 dan +0,0055 pada rasio 70:30. Kondisi ini terjadi karena sampel sintetis yang dihasilkan SMOTE justru memperkenalkan noise pada batas keputusan yang sudah terbentuk dengan baik oleh Random Forest, sehingga beberapa prediksi yang sebelumnya benar menjadi salah. Hal ini menunjukkan bahwa Random Forest sudah mampu menangani distribusi kelas yang tidak seimbang pada klasifikasi multi-kelas secara stabil ketika volume data latih mencukupi, tanpa memerlukan penyeimbangan data tambahan.

3.3.3 Pengaruh Rasio Pembagian Data terhadap Kinerja Model

Berdasarkan Tabel 6, Tabel 8, dan Tabel 9, volume data latih berpengaruh langsung terhadap stabilitas performa kedua model. Pada Tabel 6 menunjukkan bahwa Random Forest sudah mencapai akurasi tertinggi 0,9950 pada rasio 80:20 sebelum SMOTE diterapkan, sedangkan XGBoost menggunakan rasio yang sama mencapai angka yang sama setelah menerapkan SMOTE.

Pada rasio 60:40 dan 70:30, penerapan SMOTE menyebabkan penurunan *precision* pada kedua model. Namun menggunakan rasio 60:40 meningkatkan nilai *recall* di kedua model meskipun kecil, yaitu +0,0061 pada Random Forest di rasio 60:40 (Tabel 9), dan +0,0379 pada XGBoost di rasio 60:40 (Tabel 8). Ini menunjukkan bahwa, penggunaan jumlah data latih yang terbatas, menyebabkan kedua model masih kesulitan mengenali kelas minoritas secara konsisten meskipun distribusi kelas sudah diseimbangkan.

Berdasarkan Tabel 9, penurunan *precision* pada Random Forest (RF) pada rasio 70:30 mencapai -0,0369, lebih besar dibanding XGBoost yang hanya mengalami penurunan sebesar -0,0209 pada rasio yang sama. Namun pada rasio 80:20 dan 90:10, seluruh selisih metrik Random Forest bernilai 0,0000, menunjukkan bahwa performa model ini tetap stabil meskipun tanpa penerapan SMOTE. Hasil ini mengindikasikan bahwa Random Forest mampu menangani ketidakseimbangan data pada klasifikasi multi-kelas. Dengan *F1-Score* tertinggi sebesar 0,9815 serta nilai *precision* dan *recall* yang konsisten, rasio 80:20 menjadi skenario terbaik bagi Random Forest. Sementara itu, skenario terbaik untuk XGBoost diperoleh pada rasio 80:20 setelah penerapan SMOTE dengan *F1-Score* sebesar 0,9724 dan akurasi tertinggi sebesar 0,9950.

3.3.4 Analisis Komparatif XGBoost dan Random Forest

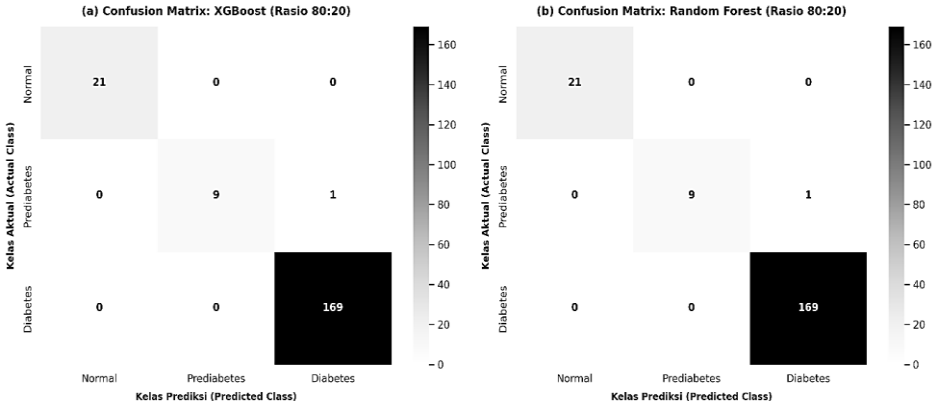
Berdasarkan Tabel 6, Tabel 8, dan Tabel 9, Random Forest menunjukkan performa yang lebih stabil dalam menangani klasifikasi multi-kelas dengan distribusi data yang tidak seimbang. Berdasarkan Tabel 9, pada rasio 80:20 dan 90:10, *F1-Score* Random Forest tidak berubah setelah penyeimbangan data (Δ = 0,0000), didukung oleh nilai *recall* yang juga tetap stabil di 0,9667 dan 0,9333. Hasil tersebut menunjukkan bahwa Random Forest sudah mampu mempertahankan performanya untuk mendeteksi ketiga kelas (multi-kelas) diagnosis tanpa terpengaruh oleh data yang tidak seimbang.

Sebaliknya, berdasarkan Tabel 8, XGBoost pada rasio 70:30 mengalami penurunan *F1-Score* dari 0,9825 menjadi 0,9707, yang disertai penurunan *precision* sebesar -0,0209 dan *recall* sebesar -0,0026. Ini menunjukkan bahwa XGBoost kurang konsisten dalam mempertahankan kemampuan untuk mendeteksi kelas minoritas ketika distribusi data latih berubah. Perbedaan perilaku ini mencerminkan karakteristik mendasar kedua algoritma yaitu Random Forest yang berbasis *bagging* secara alami lebih toleran terhadap ketidakseimbangan kelas karena setiap pohon dilatih pada subset data yang berbeda, sementara XGBoost yang berbasis *boosting* lebih sensitif terhadap perubahan distribusi data karena setiap iterasi dipengaruhi langsung oleh kesalahan iterasi sebelumnya. Dengan demikian, Random Forest lebih baik untuk klasifikasi diabetes multi-kelas pada data klinis yang tidak seimbang, karena *F1-Score* dan *recall*-nya tetap stabil di seluruh skenario rasio tanpa bergantung pada penyeimbangan distribusi data.

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

3.4 Analisis Confusion Matrix XGBoost dan Random Forest

Visualisasi confusion matrix untuk kedua model disajikan pada skenario rasio 80:20 karena skenario ini menghasilkan selisih (Δ) bernilai 0,0000 di seluruh metrik setelah SMOTE, yang menunjukkan bahwa model sudah stabil dengan volume data latih terbesar. Rincian prediksi benar dan salah pada setiap kelas diagnosis dianalisis untuk melihat konsistensi model dalam mengklasifikasikan ketiga kelas diagnosis diabetes.



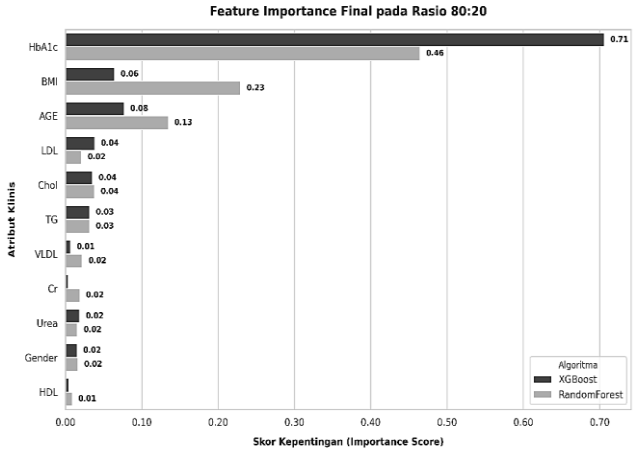
(a) (b)

Gambar 3. Confusion Matrix XGBoost dan Random Forest Rasio 80:20

Berdasarkan Gambar 3, kedua model menunjukkan pola kesalahan yang sama pada skenario rasio 80:20 sebelum SMOTE. *Confusion matrix* XGBoost ditampilkan pada Gambar 3a, sedangkan Random Forest pada Gambar 3b. Pada kedua matriks, kesalahan hanya terjadi pada kelas Prediabetes. Dari 10 sampel Prediabetes yang diuji, masing-masing model salah mengklasifikasikan 1 sampel ke dalam kelas Diabetes, sementara kelas Normal (21 sampel) dan kelas Diabetes (169 sampel) seluruhnya diprediksi dengan benar. Akibatnya, *recall* kelas Prediabetes pada kedua model hanya mencapai 0,9000 pada rasio 80:20 sebagaimana tercatat pada Tabel 7. Kesamaan pola kesalahan ini memperkuat temuan bahwa keterbatasan deteksi kelas Prediabetes bukan disebabkan oleh kelemahan salah satu algoritma, melainkan karena profil klinis pasien Prediabetes dan Diabetes memang sangat mirip sehingga sulit dibedakan oleh model manapun. Hal ini sejalan dengan performa Random Forest yang sudah stabil tanpa SMOTE.

3.5 Analisis *Feature Importance* dan Interpretasi Medis

Analisis *feature importance* untuk mengevaluasi kontribusi setiap fitur klinis terhadap keputusan klasifikasi. Hasil dari kedua model disajikan pada skenario rasio 80:20 karena skenario ini menghasilkan *F1-Score* tertinggi dan paling stabil pada kedua model sebagaimana ditunjukkan pada Tabel 8 dan Tabel 9, sehingga nilai kepentingan fitur yang dihasilkan lebih representatif terhadap kondisi model terbaiknya. Kedua model ditampilkan berdampingan untuk melihat apakah keduanya konsisten menunjukan dalam menentukan fitur mana yang paling berpengaruh. Jika XGBoost dan Random Forest sama-sama memberi bobot tinggi pada fitur yang sama, maka fitur tersebut memiliki kontribusi yang konsisten dan relevan dalam membedakan ketiga kelas diagnosis diabetes, bukan hanya hasil dari karakteristik satu algoritma saja.



Gambar 4. *Feature Importance* Model XGBoost dan Random Forest Rasio 80:20

Berdasarkan Gambar 4, kedua model menempatkan HbA1c, BMI, dan AGE sebagai tiga fitur dengan kontribusi tertinggi. Pada Random Forest, HbA1c mendominasi dengan skor 0,46, diikuti BMI (0,23) dan AGE (0,13). Pola yang

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

sama terlihat pada XGBoost, namun dominasi HbA1c jauh lebih kuat dengan skor 0,71, sementara BMI (0,06) dan AGE (0,08) berada jauh di bawahnya. Perbedaan distribusi bobot ini mencerminkan perbedaan mekanisme kedua algoritma dalam mengukur kepentingan fitur Random Forest menggunakan *Gini Importance* yang mengakumulasi kontribusi fitur dari seluruh pohon secara merata, sedangkan XGBoost menggunakan metrik *Gain* yang lebih sensitif terhadap fitur yang memberikan peningkatan informasi terbesar pada setiap iterasi *boosting*. Meskipun mekanisme kedua algoritma berbeda, keduanya mengandalkan fitur klinis yang sama dalam mengambil keputusan klasifikasi.

Ketiga fitur ini memang dikenal sebagai indikator utama diabetes secara klinis. HbA1c mencerminkan rata-rata kadar gula darah dalam tiga bulan terakhir dan menjadi acuan utama diagnosis diabetes. BMI berkaitan langsung dengan resistensi insulin, sedangkan AGE menggambarkan peningkatan risiko diabetes seiring bertambahnya usia. Konsistensi kedua model dalam menempatkan ketiga fitur ini sebagai yang paling dominan memperkuat validitas hasil penelitian, sekaligus mengindikasikan bahwa HbA1c, BMI, dan AGE layak dijadikan prioritas utama dalam sistem pendukung keputusan klinis untuk deteksi dini diabetes tanpa harus memproses seluruh fitur yang tersedia.

3.6 Diskusi

Pemilihan Random Forest pada penelitian ini didasarkan pada argumen bahwa arsitektur *bagging*-nya tangguh terhadap data tabular medis dengan fitur klinis yang beragam (Vlachas et al., 2022), sedangkan XGBoost dipilih karena mekanisme *gradient boosting* yang adaptif dalam mengatasi bias kelas (Utomo et al., 2025). Hasil penelitian mendukung argumen tersebut. Berdasarkan Tabel 9, Random Forest menunjukkan performa yang stabil pada rasio 80:20 dan 90:10, sedangkan XGBoost mengalami penurunan F1-Score dari 0,9825 menjadi 0,9707 pada rasio 70:30 setelah penerapan SMOTE (Tabel 8). Temuan ini mengindikasikan bahwa mekanisme *bagging* lebih tahan terhadap perubahan distribusi data dibandingkan *boosting*, yang sensitif terhadap kesalahan pada iterasi sebelumnya (Syahla et al., 2025).

Terkait penerapan SMOTE, tujuan awalnya adalah memastikan kedua model belajar dari distribusi kelas yang lebih seimbang tanpa membuang informasi dari data asli (Iftikhar et al., 2026). namun manfaatnya tidak konsisten pada kedua model. Pada XGBoost, SMOTE meningkatkan *recall* sebesar +0,0379 pada rasio 60:40, sedangkan pada Random Forest tidak memberikan peningkatan F1-Score yang berarti dan bahkan menurunkan *precision* pada beberapa rasio (Tabel 8 dan Tabel 9). Selain itu, *recall* kelas Prediabetes tetap rendah (0,8000) pada rasio 90:10, menunjukkan bahwa penyeimbangan data saja belum cukup untuk meningkatkan deteksi kelas minoritas ketika jumlah data latih terbatas.

Dibandingkan penelitian sebelumnya, Random Forest pada rasio 80:20 mencapai akurasi 0,9950, sedikit lebih tinggi dibanding akurasi 0,97 menurut (Sembiring Depari et al., 2025), meskipun perbandingan ini perlu diinterpretasikan secara hati-hati karena menggunakan dataset yang berbeda. Perbedaan utama penelitian ini terletak pada evaluasi stabilitas model pada berbagai rasio pembagian data serta penerapan prosedur *Split-First, Resample-Later* untuk mengurangi risiko kebocoran informasi antara data latih dan data uji. Sementara itu, dibandingkan model tunggal seperti Regresi Logistik dan C4.5 yang hanya mencapai akurasi sekitar 81–82% (Pratama et al., 2023)(Bagus et al., 2022), pendekatan *ensemble* menunjukkan performa yang lebih baik, meskipun keunggulan tersebut tetap perlu divalidasi menggunakan dataset yang sama.

Penelitian terdahulu yang menggunakan KNN dengan SMOTE-ENN melaporkan akurasi 96,95%, tetapi hanya pada klasifikasi biner (Amri et al., 2025), sehingga tidak mencerminkan kompleksitas deteksi kelas Prediabetes pada klasifikasi multi-kelas. Rendahnya *recall* Prediabetes yang konsisten sebelum maupun sesudah SMOTE menunjukkan bahwa tantangan utama tidak hanya berasal dari ketidakseimbangan kelas, tetapi juga dari kemiripan karakteristik klinis antara kelas Prediabetes dan Diabetes, sebagaimana terlihat pada hasil *confusion matrix*.

Secara keseluruhan, hasil penelitian menunjukkan bahwa Random Forest memberikan performa yang lebih stabil dibandingkan XGBoost pada berbagai rasio pembagian data, sedangkan manfaat SMOTE bergantung pada karakteristik model dan distribusi data. Temuan ini melengkapi penelitian sebelumnya dengan menunjukkan pentingnya evaluasi pada berbagai rasio pembagian data menggunakan prosedur *Split-First, Resample-Later* untuk memperoleh hasil klasifikasi diabetes multi-kelas yang lebih andal.

4. KESIMPULAN

Penelitian ini membandingkan kinerja XGBoost dan Random Forest dalam klasifikasi diabetes multi-kelas menggunakan data klinis dari Medical City Hospital dan Al-Kindy Teaching Hospital yang memiliki distribusi kelas tidak seimbang, dengan kelas Prediabetes hanya mewakili 5,3% dari total data. Proses penelitian mencakup penerapan prosedur *Split-First, Resample-Later* untuk mencegah *data leakage*, SMOTE untuk penyeimbangan data latih, optimasi hiperparameter menggunakan GridSearchCV dengan 3-*Fold Cross Validation*, serta pengujian pada empat skenario rasio pembagian data (60:40, 70:30, 80:20, dan 90:10). Hasil eksperimen menunjukkan bahwa Random Forest memiliki performa yang lebih stabil dibanding XGBoost karena mampu mempertahankan F1-*Score* dan *recall* tanpa perubahan pada rasio 80:20 dan 90:10 meskipun tanpa penerapan SMOTE, sementara XGBoost mengalami penurunan F1-*Score* pada rasio 70:30 setelah SMOTE diterapkan, yang mengindikasikan bahwa performanya lebih sensitif terhadap perubahan distribusi data latih. Rasio 80:20 terbukti menjadi skenario terbaik bagi kedua model karena menghasilkan F1-*Score* tertinggi dan paling konsisten di seluruh kelas diagnosis. Penerapan SMOTE hanya memberikan manfaat berarti pada XGBoost di rasio 60:40 melalui peningkatan *recall* sebesar +0,0379, sedangkan pada Random Forest justru


TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

menyebabkan penurunan *precision* di rasio 60:40 dan 70:30 karena model ini sudah mampu menangani ketidakseimbangan kelas secara alami melalui mekanisme *bagging*-nya. Analisis *feature importance* dari kedua model secara konsisten menempatkan HbA1c, BMI, dan AGE sebagai fitur klinis paling dominan dalam keputusan klasifikasi, yang sejalan dengan relevansi klinisnya sebagai indikator utama diagnosis diabetes. Kontribusi penelitian ini adalah memberikan evaluasi komprehensif terhadap kinerja XGBoost dan Random Forest pada data klinis multi-kelas yang tidak seimbang dengan membandingkan performa sebelum dan sesudah penerapan SMOTE pada berbagai rasio pembagian data menggunakan prosedur *Split-First, Resample-Later* untuk mencegah kebocoran informasi dari data uji ke data latih. Penelitian ini terbatas pada algoritma berbasis pohon keputusan dan dataset dari satu sumber, sehingga penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain seperti *deep learning* serta memperluas variasi dataset dari berbagai fasilitas kesehatan untuk meningkatkan generalisasi model.

REFERENCES

Adhi Pratama, N., & Wahyu Utomo, D. (2026). Deteksi diabetes mellitus dengan menggunakan teknik ensemble XGBoost dan LightGBM. *Jurnal Informatika Sunan Kalijaga*, *11*(1), 1–12. https://doi.org/10.14421/jiska.4908

Aghware, F. O., Akazue, M. I., Okpor, M. D., Malasowe, B. O., Aghaunor, T. C., Ugbotu, E. V., Ojugo, A. A., Ako, R. E., Geteloma, V. O., Odiakaose, C. C., Eboka, A. O., & Onyemenem, S. I. (2025). Effects of data balancing in diabetes mellitus detection: A comparative XGBoost and Random Forest learning approach. *NIPES - Journal of Science and Technology Research*, *7*(1), 1–11. https://doi.org/10.37933/nipes/7.1.2025.1

Agyemang, E. F., Mensah, J. A., Nyarko, E., Arku, D., Mbeah-Baiden, B., Opoku, E., & Noye Nortey, E. N. (2025). Addressing class imbalance problem in health data classification: Practical application from an oversampling viewpoint. *Applied Computational Intelligence and Soft Computing*, *2025*(1), 1–20. https://doi.org/10.1155/acis/1013769

Alfarisi, M. A., Sumanto, B., Putu, I., & Rachmawan, F. (2025). Pengembangan model machine learning untuk deteksi penyakit diabetes menggunakan analisis gini importance. *Journal of Internet and Software Engineering*, *6*(2), 127–134. https://doi.org/10.22146/jise.v6i2.16551

Amri, Z., Rodi, M., Wathani, N., Bagja, A., & Zulkipli. (2025). Prediksi diabetes menggunakan algoritma K-Nearest (KNN) teknik SMOTE-ENN. *Infotek: Jurnal Informatika Dan Teknologi*, *8*(1), 193–204. https://doi.org/10.29408/jit.v8i1.27975

Astofa, A., Rosyani, P., & Apandi, S. (2025). Evaluasi komparatif algoritma machine learning untuk prediksi dini diabetes. *Bulletin of Computer Science and Research*, *6*(1), 558–565. https://doi.org/10.47065/bulletincsr.v6i1.859

Bagus, G., Sidi, A., Arsana, M., & Gunawan, R. (2022). Peningkatan akurasi algoritma C4.5 menggunakan particle swarm optimization untuk mendeteksi penyakit diabetes. *Bit*, *19*(2), 90–97. https://doi.org/10.36080/bit.v19i2.2044

Goldney, J., Sargeant, J. A., & Davies, M. J. (2023). Incretins and microvascular complications of diabetes: Neuropathy, nephropathy, retinopathy and microangiopathy. *Diabetologia*, *66*(10), 1832–1845. https://doi.org/10.1007/s00125-023-05988-3

Haris Yunianto, A., & Rosi Subhiyakto, E. (2025). Perbandingan kinerja algoritma k-nearest neighbors dan decision tree untuk klasifikasi diabetes. *Technology and Science (BITS)*, *6*(4), 2601–2611. https://doi.org/10.47065/bits.v6i4.6550

Iftikhar, W., Yaseen, M., Rahman, G., Nauman, M. A., & Khattak, U. F. (2026). Improving diabetes prediction accuracy and interpretability with SMOTE and SHAP. *Engineering, Technology & Applied Science Research*, *16*(2), 34276–34282. https://doi.org/10.48084/etasr.16247

International Diabetes Federation. (2025). *IDF Diabetes Atlas* (D. J. Magliano, E. J. Boyko, I. Genitsaridi, L. Piemonte, P. Riley, & P. Salpea, Eds.; 11th ed.). International Diabetes Federation. https://diabetesatlas.org/resources/idf-diabetes-atlas-2025/

Jang, Y. (2025). Feature-based ensemble modeling for addressing diabetes data imbalance using the SMOTE, RUS, and random forest methods: A prediction study. *Ewha Medical Journal*, *48*(2), 1–8. https://doi.org/10.12771/emj.2025.00353

Munshi, R. M., Munshi, L. R., Himdi, H., Qashlan, A., Munshi, R., Alyahyawy, O. Y., & Khayyat, M. M. (2025). Optimising hyperparameters with a tree structured parzen estimator to improve diabetes prediction. *Scientific Reports*, *15*(1), 1–10. https://doi.org/10.1038/s41598-025-19295-x

Pratama, A., Nurcahyo, A. C., & Firgia, L. (2023). Penerapan machine learning dengan algoritma logistik regresi untuk memprediksi diabetes. *Prosiding CORISINDO*, 116–121. https://ojs.stmikpontianak.ac.id/corisindo/article/view/30

Prayoga, N., & Ridla, M. A. (2026). Analisis prediksi diabetes menggunakan decision tree pada dataset diabetes pima indians. *Journal of Information System and Application Development*, *4*(1), 146–153. https://doi.org/10.26905/jisad.v4i1.16494

Rashid, A. (2026, April 30). *Diabetes Dataset* [Data set]. Mendeley Data. https://doi.org/10.17632/wj9rwkp9c2.1

Sembiring Depari, A. D., Tania, K. D., & Sevtiyuni, P. E. (2025). Penerapan metode machine learning dan teknik SMOTE untuk prediksi diabetes. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *7*(2), 436–447. https://doi.org/10.30865/json.v7i2.9032

TIN: Terapan Informatika Nusantara
Vol 7, No 1, June 2026, page 688-700
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i1.10339

Susanto, E. R., & Cahyana, A. (2025). Penerapan algoritma XGBoost untuk prediksi diabetes: Analisis confusion matrix dan ROC curve. *Fountain of Informatics Journal*, *10*(1), 40–50. https://doi.org/10.21111/fij.v10i1.14311

Syahla, H., Izzudin, H., Pratama, F. A., Rahmatullah, B., Wahidin, A. J., & Kurniawati, I. (2025). Klasifikasi indikator kesehatan diabetes menggunakan algoritma random forest. *Jurnal Teknik Informatika Dan Teknologi Informasi*, *5*(3), 401–416. https://doi.org/10.55606/jutiti.v5i3.6338

Utomo, S., Sigit, S., Sulistyowati, N., Arman Prasetya, D., & Maulana Fahrudin, T. (2025). Perbandingan kinerja algoritma XGBoost dan CatBoost dalam klasifikasi risiko penyakit diabetes. *Alinier*, *6*(2), 84–96. https://doi.org/10.36040/alinier.v6i2.14772

Vlachas, C., Damianos, L., Gousetis, N., Mouratidis, I., Kelepouris, D., Kollias, K.-F., Asimopoulos, N., & Fragulis, G. F. (2022). Random forest classification algorithm for medical industry data. *SHS Web of Conferences*, 1–6. https://doi.org/10.1051/shsconf/202213903008

Wongso Prawiro, A., Rofilah, A. K., Putri, I. P. H., Praditna, L. M. A., Hasanah, M., & Husodo, D. P. (2025). Diabetic ketoacidosis and hyperosmolar hyperglycemic state: Diagnosis and management in emergency condition : A literature review. *Jurnal Biologi Tropis*, *25*(4a), 620–626. https://doi.org/10.29303/jbt.v25i4a.11083

Zhou, F., Hu, S., Du, X., & Lu, Z. (2025). Anston attentional network for structured data based stroke risk prediction in smart aging. *Scientific Reports*, *15*(1), 1–20. https://doi.org/10.1038/s41598-025-18758-5