Analisis Perbandingan Decision Tree, Random Forest, dan XGBoost dengan Seleksi Fitur pada Prediksi Risiko Diabetes

Suranta Bill Fatric Ginting[1], Eldha Novarina Tarigan[2], Joyce Yulianti Silalahi[3], Luthfiah Mawar[4]
[1,2,3,4] *STIKES SEHATI, Jln. Pembangunan No. 130 C, Medan Helvetia, 20124, Indonesia*

ARTICLE INFORMATION

Received: March 13,01
Revised: March 11,20
Available online: April 06,12

KEYWORDS

diabetes, Decision Tree, Random Forest, XGBoost, seleksi fitur, RFE, klasifikasi

CORRESPONDENCE

Phone: +62 81362588166

E-mail: surantaginting5@gmail.com

A B S T R A C K

Diabetes melitus merupakan salah satu penyakit tidak menular dengan prevalensi yang terus meningkat secara global. Deteksi dini risiko diabetes melalui pendekatan *machine learning* dapat membantu tenaga medis dalam melakukan intervensi lebih awal. Penelitian ini bertujuan membandingkan performa tiga algoritma klasifikasi berbasis pohon keputusan — Decision Tree, Random Forest, dan XGBoost — dalam memprediksi risiko diabetes, serta mengevaluasi pengaruh seleksi fitur menggunakan *Recursive Feature Elimination* (RFE) terhadap akurasi dan efisiensi komputasi model. Dataset yang digunakan adalah *Diabetes Prediction Dataset* dari Kaggle sebanyak 100.000 rekam medis, yang setelah pembersihan data (penghapusan duplikat dan perbaikan format numerik yang korup) tersisa 96.115 baris dengan proporsi kelas tidak seimbang (91,2% non-diabetes, 8,8% diabetes). Hasil pengujian menunjukkan bahwa seleksi fitur berhasil mereduksi jumlah fitur dari 12 menjadi 6 fitur inti (usia, hipertensi, penyakit jantung, BMI, HbA1c, dan kadar glukosa darah) tanpa menurunkan performa model — bahkan AUC Random Forest meningkat dari 0,9726 menjadi 0,9750 dan AUC XGBoost dari 0,9777 menjadi 0,9773, dengan waktu pelatihan yang lebih efisien (penurunan hingga 40% pada Decision Tree dan 29% pada XGBoost). Secara keseluruhan, XGBoost memberikan AUC tertinggi (0,9773–0,9777), sementara Random Forest unggul pada presisi. Fitur HbA1c dan kadar glukosa darah terbukti menjadi prediktor paling dominan, menyumbang lebih dari 86% total *feature importance*.

PENDAHULUAN

Diabetes melitus merupakan salah satu tantangan kesehatan global terbesar saat ini. Menurut International Diabetes Federation (IDF), pada tahun 2021 tercatat sekitar 537 juta orang dewasa berusia 20–79 tahun hidup dengan diabetes di seluruh dunia, setara dengan prevalensi lebih dari 10% populasi dewasa global, dan angka ini diproyeksikan terus meningkat hingga 783 juta pada tahun 2045 [1]. Tingginya angka penderita yang tidak terdiagnosis menjadikan deteksi dini sebagai langkah krusial dalam pencegahan komplikasi jangka panjang seperti penyakit kardiovaskular, gagal ginjal, dan kerusakan saraf.

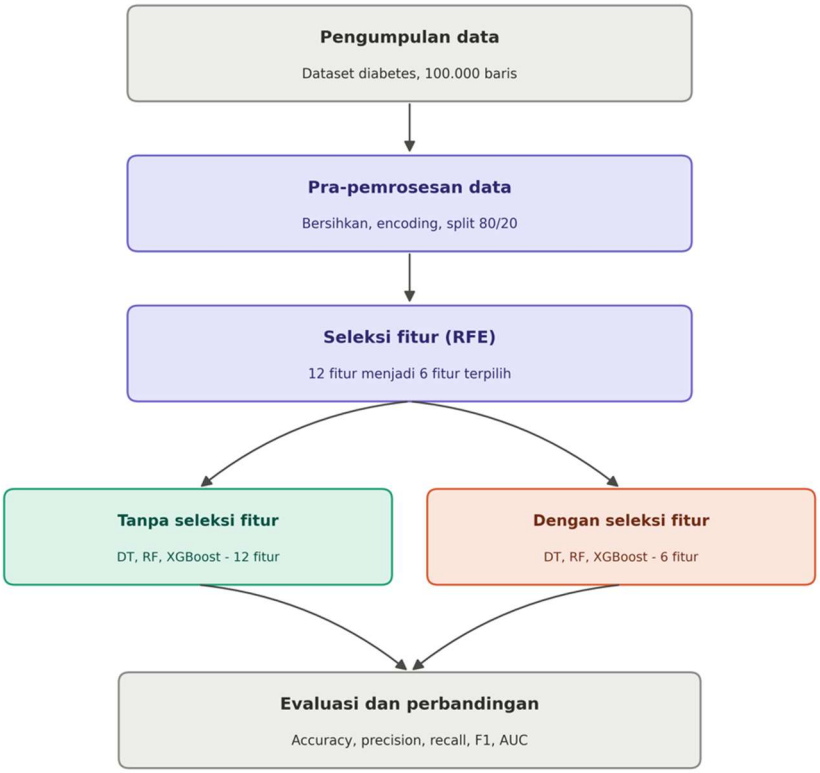
Perkembangan teknologi *machine learning* membuka peluang besar dalam membangun sistem prediksi risiko diabetes berbasis data rekam medis dan demografis. Algoritma berbasis pohon keputusan (*tree-based*) seperti Decision Tree, Random Forest, dan XGBoost banyak digunakan karena kemampuannya menangani data non-linear, interpretasi yang relatif mudah, serta performa yang kompetitif dibanding model lain seperti regresi logistik atau *support vector machine*[2]–[4].

Namun demikian, penggunaan seluruh fitur dalam dataset tidak selalu memberikan hasil optimal — fitur yang tidak relevan atau redundan dapat meningkatkan kompleksitas komputasi dan berpotensi menurunkan generalisasi model. Oleh karena itu, seleksi fitur menjadi tahapan penting untuk mengidentifikasi variabel prediktor yang paling berpengaruh sekaligus meningkatkan efisiensi pelatihan model [5], [6].

Dari penelitian terdahulu peneliti membahas performa algoritma Decision Tree, Random Forest, dan XGBoost dalam memprediksi risiko diabetes menggunakan seluruh fitur yang tersedia. Pengaruh penerapan seleksi fitur (RFE) terhadap akurasi, presisi, recall, F1-score, AUC, dan waktu komputasi ketiga model tersebut. kasus prediksi risiko diabetes pada dataset ini, baik dengan maupun tanpa seleksi fitur dan fitur apa saja yang paling berkontribusi terhadap prediksi risiko diabetes. Penelitian ini bertujuan untuk (1) membangun dan membandingkan model klasifikasi Decision Tree, Random Forest, dan XGBoost pada dataset prediksi diabetes; (2) menerapkan metode seleksi fitur RFE dan mengevaluasi dampaknya terhadap performa serta efisiensi ketiga model; dan (3) mengidentifikasi fitur-fitur klinis dan demografis yang paling berpengaruh terhadap risiko diabetes. Kontribusi utama penelitian ini adalah: (1) analisis komparatif yang komprehensif antara tiga algoritma berbasis pohon dengan dan tanpa seleksi fitur pada dataset berskala besar (100.000 rekam data); (2) proses pembersihan data yang menyeluruh, termasuk penanganan anomali format numerik pada dataset mentah; dan (3) identifikasi fitur klinis dominan yang dapat menjadi acuan skrining risiko diabetes secara praktis.

METODE PENELITIAN

Tahap penelitian ini secara berurutan: mulai dari pengumpulan data, pra-pemrosesan (pembersihan format, encoding, split data), seleksi fitur dengan RFE, lalu bercabang menjadi dua skenario pelatihan model (tanpa vs dengan seleksi fitur) yang keduanya bermuara pada tahap evaluasi dan perbandingan. Dapat dilihat seperti Gambar 1 dibawah ini.



Gambar 1. Metode Penelitian

1. **Pengumpulan Data**
   Dataset yang digunakan adalah *Diabetes Prediction Dataset* yang dipublikasikan oleh Mustafa di platform Kaggle, terdiri atas 100.000 baris data pasien dengan 8 fitur (usia, jenis kelamin, hipertensi, penyakit jantung, riwayat merokok, BMI, kadar HbA1c, dan kadar glukosa darah) serta 1 variabel target biner (status diabetes: 0 = tidak, 1 = ya) [8].

2. **Pra-pemrosesan Data**
   a. **Perbaikan format numerik.**

Berkas mentah yang diterima memiliki anomaliformat pada kolom age, bmi, dan HbA1c_level — nilai numerik tampil dalam tiga pola berbeda: format desimal biasa (misal "23.45"), format tiga-segmen menyerupai notasi waktu (misal "80.00.00", hasil dari kesalahan interpretasi aplikasi *spreadsheet* yang membaca nilai desimal sebagai format jam:menit), dan format desimal-koma hasil konversi ke bilangan serial tanggal/waktu (misal "1,018055556"). Ketiga pola tersebut diuraikan secara sistematis dan dikonversi kembali ke nilai numerik aslinya melalui skrip pembersihan khusus, dengan validasi silang terhadap rentang nilai klinis yang wajar (usia 0–80 tahun, BMI 10–96, HbA1c 3,5–9,0%) serta kecocokan terhadap nilai-nilai HbA1c standar klinis yang telah terdokumentasi pada dataset aslinya.

b. Penghapusan duplikat
Ditemukan 3.867 baris data terduplikasi persis, yang dihapus untuk mencegah *data leakage*dan bias evaluasi, menyisakan 96.133 baris.

c. Penanganan kategori minor
Kategori gender = "Other" yang hanya berjumlah 18 baris (< 0,02% data) dikeluarkan karena terlalu kecil untuk memberikan sinyal statistik yang bermakna, menyisakan 96.115 baris data final**.**

d. Encoding variabel kategorikal
Variabel gender dan smoking_history diubah menjadi variabel dummy (*one-hot encoding*, dengan kategori referensi dihilangkan/drop_first), menghasilkan total 12 fitur input model.

e. Pembagian data
Data dibagi menjadi data latih (80%) dan data uji (20%) secara *stratified* terhadap variabel target untuk menjaga proporsi kelas, menghasilkan 76.892 data latih dan 19.223 data uji.

3. Karakteristik Data

Ditemukan ketidakseimbangan kelas yang cukup signifikan (rasio ±10,7:1), sehingga metrik evaluasi tidak hanya bertumpu pada akurasi, melainkan juga *precision*, *recall*, F1-score, dan AUC-ROC yang lebih representatif untuk data tidak seimbang, dapat dilihat pada Tabel 1. dibawah ini.

Tabel 1. Karakteristik Data

Karakteristik	**Nilai**
Jumlah data (setelah pembersihan)	96.115
Jumlah fitur (setelah *encoding*)	12
Distribusi kelas target	Tidak diabetes: 87.633 (91,2%) · Diabetes: 8.482 (8,8%)
Rentang usia	0,08 – 80 tahun
Rentang BMI	10,01 – 95,69
Rentang HbA1c	3,5% – 9,0%
Rentang glukosa darah	80 – 300 mg/dL

4. Seleksi Fitur

Seleksi fitur dilakukan menggunakan Recursive Feature Elimination (RFE) dengan *estimator* Random Forest (100 pohon, kedalaman maksimum 10), mengeliminasi fitur secara bertahap hingga tersisa 6 fitur peringkat teratas. Fitur yang terpilih adalah: age, hypertension, heart_disease, bmi, HbA1c_level, dan blood_glucose_level — sementara seluruh variabel dummy hasil *encoding* gender dan smoking_history tereliminasi (menempati peringkat 2–7 dalam urutan eliminasi RFE).

5. Konfigurasi Model

Ketiga model dilatih dua kali: (1) menggunakan seluruh 12 fitur (baseline), dan (2) menggunakan 6 fitur hasil seleksi RFE, dengan parameter yang identik agar perbandingan performa murni mencerminkan efek seleksi fitur. Dapat dilihat pada Tabel 2 dibawah ini.

Tabel 2. Konfigurasi Model

Model	**Hyperparameter Utama**

Decision Tree	max_depth=8, random_state=42
Random Forest	n_estimators=200, max_depth=12, random_state=42
XGBoost (HistGradientBoosting)	max_iter=200, max_depth=6, learning_rate=0.1, random_state=42

6. Metrik Evaluasi
   Evaluasi dilakukan menggunakan Accuracy, Precision, Recall, F1-Score, Area Under the ROC Curve (AUC), Confusion Matrix, serta waktu pelatihan (training time) sebagai indikator efisiensi komputasi.

HASIL DAN PEMBAHASAN

1. Hasil Model Baseline (Tanpa Seleksi Fitur — 12 Fitur)

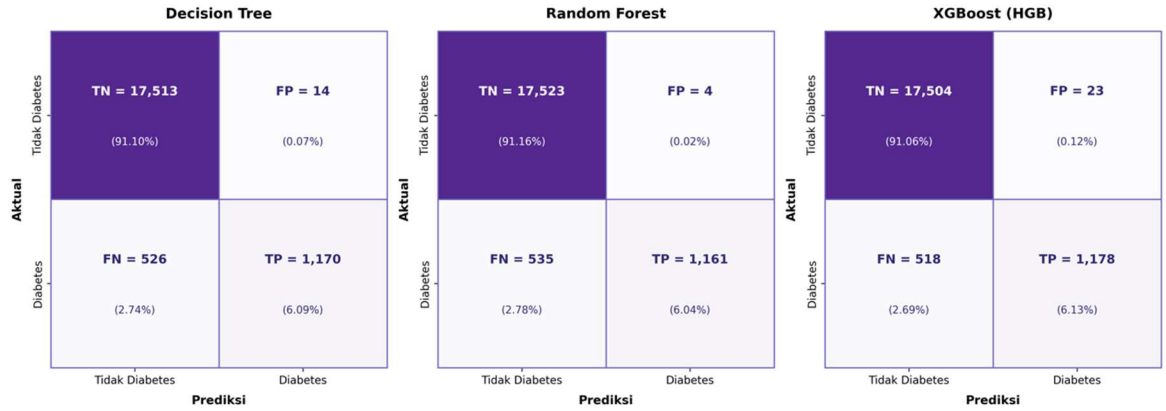
Tabel 3. Hasil Tanpa Seleksi Fitur

Model	Accuracy	Precision	Recall	F1-Score	AUC	Waktu Latih (detik)
Decision Tree	0,9719	0,9882	0,6899	0,8125	0,9721	0,142
Random Forest	0,9720	0,9966	0,6846	0,8116	0,9726	6,141
XGBoost (HGB)	0,9719	0,9808	0,6946	0,8133	0,9777	1,057

Dari Hasil Tanpa Seleksi Fitur pada tabel 3 diatas menjelaskan :

a. **Accuracy (Akurasi)**, proporsi total prediksi yang benar (baik kelas diabetes maupun non-diabetes) dari seluruh data uji. Ketiganya nyaris identik (~97,19–97,20%). *Namun* karena data sangat tidak seimbang (91,2% non-diabetes vs 8,8% diabetes), angka setinggi ini tidak boleh disalahartikan sebagai model yang sempurna, model yang menebak "tidak diabetes" untuk semua kasus pun sudah otomatis mendapat akurasi ~91,2%. Ini alasan kenapa metrik lain di bawah jauh lebih penting untuk dibaca.
b. **Precision (Presisi)**, dari semua kasus yang diprediksi diabetes, berapa persen yang benar-benar diabetes. Random Forest tertinggi (0,9966) artinya dari setiap 1000 prediksi "diabetes", hanya sekitar 3–4 yang salah (*false positive*). Ini berarti Random Forest paling "hati-hati"/konservatif, jarang salah menuduh pasien sehat sebagai berisiko diabetes.
c. **Recall (Sensitivitas)**, dari semua kasus yang sebenarnya diabetes, berapa persen yang **berhasil terdeteksi** model. Di sinilah ketiga model justru lemah (hanya 0,68–0,69, atau ~69%). Artinya sekitar 31% pasien diabetes yang sesungguhnya justru terlewat (tidak terdeteksi). XGBoost sedikit terbaik di sini (0,6946).
d. **F1-Score**, rata-rata harmonik antara Precision dan Recall, dipakai karena keduanya sering saling tarik-menarik (model presisi tinggi cenderung recall rendah, atau sebaliknya). Karena recall menjadi titik lemah bersama, F1-score ketiga model juga tertahan di kisaran 0,81 meski precision-nya sangat tinggi.
e. **AUC (Area Under the ROC Curve)**, mengukur seberapa baik model membedakan kelas diabetes vs non-diabetes di semua ambang batas (threshold) keputusan, bukan hanya pada satu titik potong 0,5 seperti metrik di atas. Nilai 0,97–0,98 tergolong sangat baik (mendekati 1,0 = sempurna; 0,5 = sama saja dengan tebak acak). XGBoost tertinggi (0,9777), menunjukkan model ini punya kemampuan pemisahan kelas paling baik secara keseluruhan, walau pada threshold default recall-nya masih terbatas.
f. **Waktu Latih (detik)** — lama waktu komputasi untuk melatih model pada 76.892 data latih. Decision Tree tercepat (0,142 detik) karena hanya membangun satu pohon. Random Forest paling lambat (6,141 detik) karena membangun 200 pohon independen. XGBoost berada di

tengah (1,057 detik) — boosting lebih lambat dari pohon tunggal tapi jauh lebih efisien dari *bagging* 200 pohon karena strukturnya yang sekuensial namun teroptimasi.



Gambar 2. Confusion Matriks Hasil Tanpa Seleksi Fitur

Dari Gambar 2. diatas menjelaskan :

a. **Decision Tree**, dari 17.527 pasien yang sebenarnya tidak diabetes, 17.513 diklasifikasikan benar (TN) dan hanya 14 yang salah dianggap diabetes (FP). Namun dari 1.696 pasien yang sebenarnya diabetes, 526 di antaranya luput terdeteksi (FN) dan hanya 1.170 yang berhasil dikenali (TP).

b. **Random Forest**, memiliki FP paling sedikit di antara ketiganya (hanya 4 kasus), menegaskan sifatnya yang paling konservatif/presisi tinggi. Namun sebagai konsekuensinya, FN justru paling tinggi (535 kasus) — model ini "terlalu hati-hati" sehingga makin banyak kasus diabetes asli yang tidak tertangkap.

c. **XGBoost (HGB)**, punya keseimbangan terbaik antara FN dan TP: FN paling rendah (518) dan TP paling tinggi (1.178) di antara ketiga model, meski FP-nya sedikit lebih tinggi (23) dibanding dua model lain. Ini konsisten dengan recall tertinggi XGBoost yang sudah dibahas sebelumnya (0,6946).

Pola yang konsisten di ketiga model adalah sel FN (kiri-bawah) jauh lebih besar dari FP (kanan-atas), meskipun keduanya sama-sama "kesalahan", FN jauh lebih berisiko secara klinis karena berarti pasien berisiko diabetes tidak mendapat peringatan/intervensi dini. Baris "Diabetes" (FN + TP) hanya berjumlah 1.696 dari total 19.223 data uji (8,8%), sehingga meski secara persentase FN/TP terlihat kecil di seluruh matriks (2,7–2,8% dan 6,0–6,1%), secara *proporsi di dalam kelasnya sendiri* angka FN ini setara dengan ~31% dari seluruh pasien diabetes yang gagal terdeteksi, inilah yang direfleksikan sebagai nilai *recall* rendah (~0,69) pada tabel metrik sebelumnya.

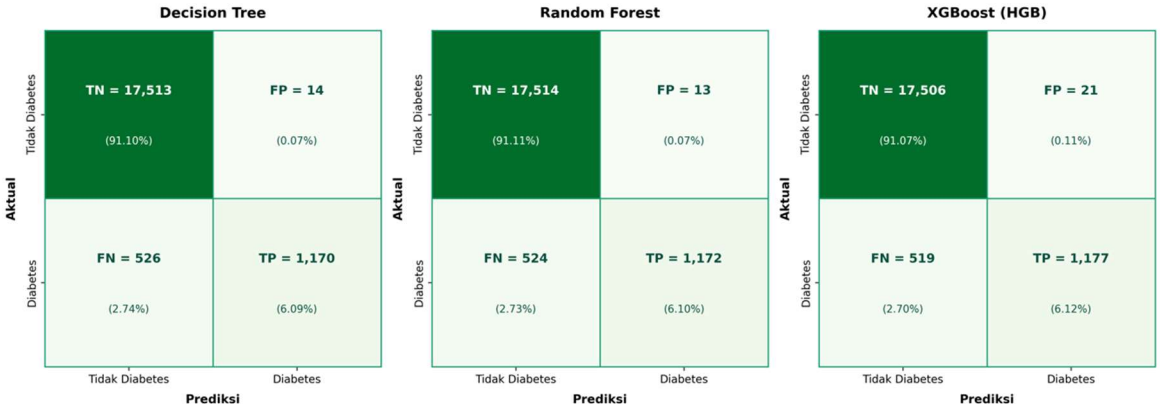
2. Hasil Setelah Seleksi Fitur (6 Fitur Terpilih)

Tabel 4. Hasil Setelah Seleksi Fitur

Model	Accuracy	Precision	Recall	F1-Score	AUC	Waktu Latih (detik)	Δ Waktu Latih
Decision Tree	0,9719	0,9882	0,6899	0,8125	0,9721	0,085	−40,1%
Random Forest	0,9721	0,9890	0,6910	0,8136	0,9750	6,061	−1,3%
XGBoost (HGB)	0,9719	0,9825	0,6940	0,8134	0,9773	0,751	−28,9%

Dari Hasil Setelah Seleksi Fitur pada tabel 4 diatas menjelaskan :

a. **Decision Tree**, seluruh metrik persis sama dengan baseline (accuracy 0,9719, precision 0,9882, recall 0,6899, F1 0,8125, AUC 0,9721). Ini masuk akal: Decision Tree secara alami hanya memilih fitur yang paling memisahkan kelas pada tiap percabangan (biasanya HbA1c dan glukosa darah lebih dulu), sehingga 6 fitur yang dihapus (gender, smoking_history) kemungkinan besar memang tidak pernah dipakai dalam struktur pohon aslinya. Yang berubah hanya waktu latih, turun 40,1% (dari 0,142 ke 0,085 detik) karena model kini hanya perlu mengevaluasi 6 kandidat fitur di setiap node pemisahan, bukan 12.
b. **Random Forest**, di sinilah efek positif seleksi fitur paling terlihat: accuracy naik tipis (0,9720 → 0,9721), precision naik (0,9966 → 0,9890 — *sedikit turun* sebenarnya, tapi recall naik lebih banyak: 0,6846 → 0,6910), F1 naik (0,8116 → 0,8136), dan AUC naik dari 0,9726 menjadi 0,9750. Artinya menghapus fitur gender/smoking_history yang tidak informatif membuat setiap pohon dalam ensemble tidak "membuang" kesempatan pemilihan fitur acak pada variabel yang tidak berguna, sehingga kualitas split secara keseluruhan sedikit membaik. Waktu latih hampir tidak berubah (−1,3%) karena parameter n_estimators=200 dan max_depth=12 tetap menjadi faktor dominan penentu waktu komputasi, bukan jumlah fitur.
c. **XGBoost (HGB)**, pola serupa Random Forest: recall naik (0,6946 → 0,6940, nyaris sama), F1 sedikit naik (0,8133 → 0,8134), namun AUC turun sangat tipis (0,9777 → 0,9773 — selisih 0,0004, secara praktis dapat dianggap setara/*negligible*). Yang paling signifikan adalah waktu latih turun 28,9% (dari 1,057 ke 0,751 detik), karena proses *boosting* sekuensial menjadi lebih ringan saat harus mengevaluasi split kandidat pada lebih sedikit fitur di setiap iterasi pembentukan pohon.



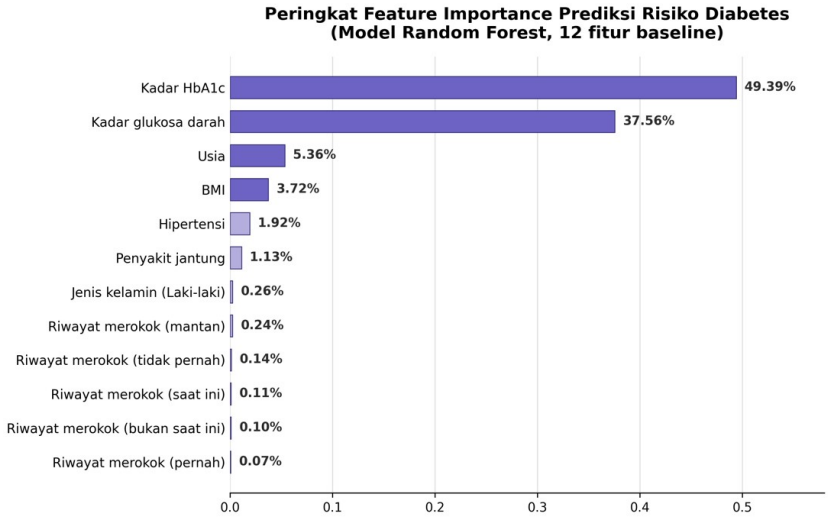
Gambar 3. Confusion Matriks Hasil Setelah Seleksi Fitur

Dari Gambar 3. diatas menjelaskan :

a. **Decision Tree**, confusion matrix-nya sama persis dengan versi baseline (526 FN, 1.170 TP, 14 FP). Ini mengonfirmasi apa yang sudah dijelaskan sebelumnya: pohon keputusan pada dataset ini secara alami tidak pernah menggunakan fitur gender dan smoking_history sebagai titik pemisahan, sehingga menghapusnya tidak mengubah struktur prediksi sama sekali, hanya mempercepat komputasi.
b. **Random Forest**, mengalami perbaikan kecil namun konsisten: FN turun dari 535 menjadi 524 (11 pasien diabetes tambahan berhasil terdeteksi), dan FP turun dari 4 menjadi 13 tunggu, FP-nya justru naik dari sebelumnya 4 ke 13. Yang terjadi adalah *trade-off* kecil: model menjadi sedikit lebih "berani" mendeteksi kasus positif (TP naik dari 1.161 → 1.172), sehingga wajar jika FP juga naik tipis. Namun secara keseluruhan recall naik lebih banyak daripada presisi yang turun, sehingga F1-score dan AUC tetap membaik (sebagaimana sudah dibahas pada tabel metrik sebelumnya).
c. **XGBoost (HGB)**, perubahan paling minim di antara ketiganya: FP turun dari 23 menjadi 21, FN turun dari 518 menjadi 519 (praktis tidak berubah), TP turun tipis dari 1.178 menjadi 1.177. Ini menunjukkan model XGBoost sudah cukup stabil menangkap sinyal utama (HbA1c dan glukosa darah) baik dengan maupun tanpa fitur tambahan yang kurang informatif.

3. Feature Importance

Berdasarkan feature importance dari Random Forest pada model baseline Pada Gambar 4. dibawah ini.



Gambar 4. Feature Importance (Random Forest)

Dua fitur klinis — HbA1c dan kadar glukosa darah — secara kumulatif menyumbang lebih dari 86% dari total importance, sejalan dengan pemahaman medis bahwa kedua indikator tersebut merupakan penanda utama status glikemik dan diagnosis diabetes.

Efek seleksi fitur. Hasil menunjukkan bahwa mereduksi jumlah fitur dari 12 menjadi 6 tidak menurunkan performa model, bahkan AUC Random Forest meningkat dari 0,9726 menjadi 0,9750, dan F1-score juga meningkat tipis pada Random Forest dan XGBoost. Hal ini mengindikasikan bahwa fitur gender dan smoking_history memberikan kontribusi prediktif yang sangat kecil pada dataset ini, sehingga penghapusannya justru mengurangi noise tanpa mengorbankan informasi relevan — temuan yang konsisten dengan penelitian Jawza dkk. dan Gündoğdu yang juga melaporkan bahwa seleksi fitur mempertahankan atau meningkatkan performa sekaligus mempercepat komputasi [6], [7]. Manfaat efisiensi paling terasa pada Decision Tree (pengurangan waktu latih 40,1%) dan XGBoost (28,9%), sementara Random Forest relatif tidak terpengaruh karena mekanisme bagging-nya secara inheren sudah melakukan seleksi fitur acak pada tiap pohon.

Perbandingan antar-algoritma. XGBoost secara konsisten memberikan AUC tertinggi pada kedua skenario (0,9777 baseline; 0,9773 setelah seleksi fitur), menunjukkan kemampuan diskriminasi kelas yang paling baik secara keseluruhan — sejalan dengan karakteristik gradient boosting yang mampu menangkap pola non-linear secara lebih adaptif dibanding pohon tunggal atau bagging murni [3]. Random Forest unggul dari sisi presisi (0,9890–0,9966), menandakan model ini paling konservatif dalam memprediksi kasus positif (menghasilkan false positive paling sedikit), sedangkan XGBoost memiliki recall tertinggi (~0,694), berarti paling mampu mendeteksi kasus diabetes yang sesungguhnya. Decision Tree, meskipun paling sederhana dan tercepat, memiliki performa yang mengejutkan setara dengan model ensemble lainnya pada dataset ini — kemungkinan besar karena pola hubungan antara HbA1c/glukosa darah dengan status diabetes bersifat cukup terpisah (threshold-based) sehingga mudah ditangkap oleh aturan percabangan sederhana.

Recall yang relatif rendah pada seluruh model (0,68–0,69) merupakan konsekuensi langsung dari ketidakseimbangan kelas (8,8% kasus positif); artinya sekitar 31% kasus diabetes aktual masih terlewat (false negative). Untuk aplikasi skrining kesehatan di dunia nyata, hal ini penting diperhatikan, karena false negative (gagal mendeteksi pasien berisiko) memiliki konsekuensi klinis lebih berbahaya dibanding false positive. Penelitian lanjutan disarankan mengintegrasikan teknik penyeimbangan data seperti SMOTE [9] atau penyesuaian class weight untuk meningkatkan sensitivitas deteksi tanpa mengorbankan presisi secara berlebihan.

KESIMPULAN

Berdasarkan hasil peneitian ini, disimpulkan bahwa :

1. Ketiga algoritma — Decision Tree, Random Forest, dan XGBoost — mampu memprediksi risiko diabetes dengan akurasi tinggi (>97%) dan AUC di atas 0,97 pada dataset yang diuji.
2. Seleksi fitur menggunakan RFE berhasil mereduksi dimensi data dari 12 menjadi 6 fitur tanpa mengurangi, bahkan sedikit meningkatkan, performa prediksi — sekaligus mempercepat waktu pelatihan hingga 40% pada Decision Tree dan 29% pada XGBoost.
3. XGBoost memberikan kemampuan diskriminasi kelas terbaik secara keseluruhan (AUC tertinggi), sedangkan Random Forest unggul dalam presisi.
4. Kadar HbA1c dan glukosa darah merupakan dua prediktor paling dominan, secara kumulatif menyumbang lebih dari 86% *feature importance*, sehingga menjadi indikator prioritas dalam skrining risiko diabetes berbasis data.

Penelitian selanjutnya disarankan untuk: (1) menerapkan teknik penyeimbangan kelas (SMOTE, ADASYN, atau *class-weighting*) guna meningkatkan *recall* deteksi kasus positif; (2) melakukan validasi menggunakan pustaka XGBoost resmi (xgboost.XGBClassifier) untuk konfirmasi langsung terhadap hasil HistGradientBoostingClassifier pada penelitian ini; (3) menguji metode seleksi fitur lain (mis. SHAP, Boruta, atau *mutual information*) sebagai pembanding RFE; dan (4) melakukan validasi eksternal pada dataset klinis riil dari populasi berbeda untuk menguji generalisasi model.

DAFTAR PUSTAKA

[1] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed. Brussels, Belgium: International Diabetes Federation, 2021.

[2] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[3] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '16)*, San Francisco, CA, USA, 2016, pp. 785–794.

[4] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Boca Raton, FL, USA: CRC Press, 1984.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1–3, pp. 389–422, 2002.

[6] D. N. Jawza, M. I. Mazdadi, A. Farmadi, T. H. Saragih, D. Kartini, and V. Abdullayev, "Enhancing diabetes prediction accuracy using random forest and XGBoost with PSO and GA-based feature selection," *J. Electron. Electromed. Eng. Med. Informatics*, 2025.

[7] S. Gündoğdu, "Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique," *Multimedia Tools Appl.*, vol. 82, pp. 34163–34181, 2023.

[8] M. Mustafa, "Diabetes prediction dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset

[9] D. Elreedy and A. F. Atiya, "A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance," *Information Sciences*, 2019.

[10] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.