JTERA (Jurnal Teknologi Rekayasa), Vol. 9, No. 2, Desember 2024, Hal. 121-130 p-ISSN 2548-737X
Terakreditasi "**Peringkat 3**" oleh Kemenristek/BRIN, Nomor SK: 85/M/KPT/2020 e-ISSN 2548-8678
DOI: 10.31544/jtera.v9.i1.2024.121-130

Comparison of Random Forest and XGBoost for Diabetes Classification with SHAP and LIME Interpretation

Mubaraqah[1], Annisa Nurul Puteri[2], A. Sumardin[3]
[1,2,3]Program Studi Teknik Informatika, Universitas Teknologi Akba Makassar
Jl. Perintis Kemerdekaan ,Tamalanrea, Kec. Tamalanrea, Kota Makassar, Indonesia
mubaraqah21@mhs.akba.ac.id

Abstrak

Diabetes Mellitus (DM) merupakan tantangan kesehatan global yang memerlukan pendekatan inovatif untuk deteksi dini dan pengelolaan yang efektif. Penelitian ini bertujuan membandingkan algoritma **Random Forest** dan **XGBoost** dalam klasifikasi jenis diabetes serta meningkatkan interpretasi model menggunakan teknik **Explainable AI (XAI**) seperti SHAP dan LIME. Metode penelitian mencakup pengolahan dataset publik berisi 70.000 entri dengan 34 fitur medis, pelatihan model menggunakan parameter yang dioptimalkan serta analisis interpretasi. Hasil penelitian menunjukkan bahwa XGBoost memiliki akurasi lebih tinggi (90,6%) dengan generalisasi yang baik sedangkan Random Forest lebih unggul dalam efisiensi waktu pelatihan. Analisis fitur mengidentifikasi faktor utama seperti Age, Blood Glucose Levels dan Weight Gain During Pregnancy yang mempengaruhi prediksi. Temuan ini memberikan panduan model yang akurat dan transparan untuk mendukung pengambilan keputusan medis.

Kata kunci: Machine Learning, XAI, Klasifikasi Diabetes, Evaluasi Model, Interpretasi Model

Abstract

*Diabetes Mellitus (DM) is a global health challenge requiring innovative approaches for early detection and effective management. This study aims to compare **Random Forest** and **XGBoost** algorithms in diabetes classification while enhancing model interpretability using **Explainable AI (XAI)** techniques such as SHAP and LIME. The methodology involves processing a public dataset containing 70,000 entries with 34 medical features, training models with optimized parameters, and performing interpretative analysis. Results show that XGBoost achieves higher accuracy (90.6%) with better generalization, while Random Forest excels in training efficiency. Feature analysis identifies key factors such as Age, Blood Glucose Levels, and Weight Gain During Pregnancy as significant contributors to predictions. These findings provide an accurate and transparent model guide to support medical decision-making.*

Keywords: *Machine Learning, XAI, Diabetes Classification, Model Evaluation, Model Interpretation*

I. PENDAHULUAN

Diabetes Mellitus (DM) merupakan salah satu masalah kesehatan global utama yang terus meningkat, memengaruhi lebih dari 400 juta orang di seluruh dunia dan angka ini diproyeksikan terus bertambah seiring dengan urbanisasi, perubahan gaya hidup serta peningkatan prevalensi obesitas [1]-[5]. DM dibagi menjadi dua tipe utama: Tipe 1 yang disebabkan oleh gangguan autoimun dan Tipe 2 yang lebih sering terjadi akibat resistensi insulin dan gaya hidup tidak sehat. Kedua tipe ini berkontribusi pada peningkatan risiko komplikasi serius seperti penyakit kardiovaskular, stroke dan gagal ginjal [6]-[8]. Deteksi dini dan pengelolaan yang tepat sangat penting untuk mengurangi dampak negatif penyakit ini terhadap kualitas hidup individu maupun sistem kesehatan masyarakat secara keseluruhan. Namun, dalam praktiknya, identifikasi dini diabetes sering kali menghadapi tantangan, termasuk keakuratan prediksi dan keterbatasan analisis faktor risiko yang dapat diandalkan. Oleh karena itu, diperlukan pendekatan inovatif berbasis teknologi untuk mendukung deteksi dini yang lebih efektif dan efisien [9]-[12].

Dalam beberapa tahun terakhir, berbagai pendekatan pembelajaran mesin *(Machine Learning/ML)* telah diterapkan untuk prediksi diabetes menghasilkan tingkat akurasi yang bervariasi tergantung pada metode yang digunakan Pendekatan berbasis ekstraksi fitur seperti *Principal Component*

Diterima: 28 Desember 2024; Direvisi: 31 Desember 2024; Disetujui: 02 Januari 2025
JTERA, Vol. 9, No. 2, Desember 2024

© 2024 JTERA (Jurnal Teknologi Rekayasa), Politeknik Sukabumi

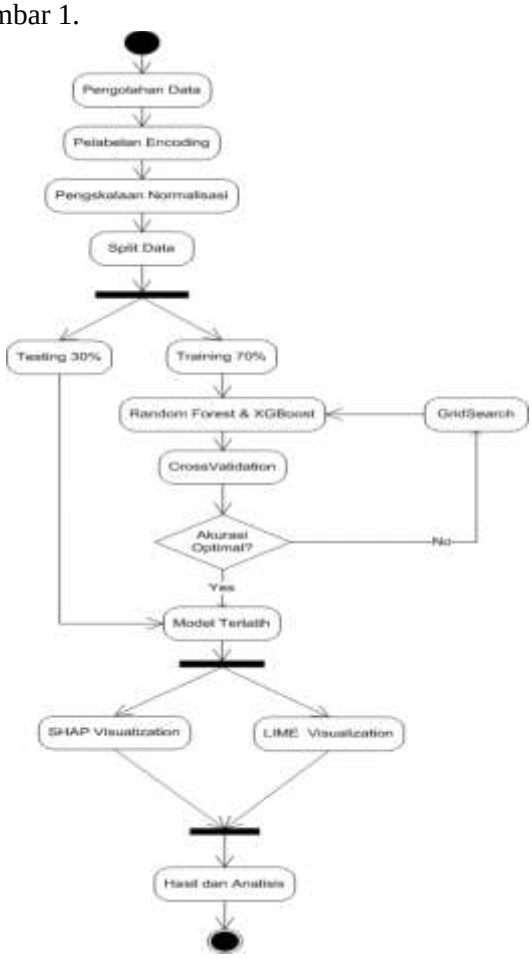
Analysis (PCA) dan filter *resampling*, bersama dengan algoritma seperti *K-Nearest Neighbor (KNN), Naive Bayes*, dan *Decision Tree*, telah menunjukkan hasil yang menjanjikan. Pengklasifikasi *Decision Tree* misalnya berhasil mencapai akurasi 94,4% [13]. Model ensemble seperti *AdaBoost* yang dilengkapi dengan proses pra proses yang cermat, bahkan mampu mencapai akurasi hingga 98%. Selain itu, algoritma *Gradient Boosting* mencapai akurasi 92,85% pada dataset *Pima Indian Diabetes* [15]. Namun, meskipun akurasi model telah meningkat, tantangan utama terletak pada interpretabilitasnya [16],[17]. Model seperti *Random Forest (RF)* dan *XGBoost (Extreme Gradient Boosting)* sering kali dianggap sebagai *black-box*, sehingga hasil prediksi sulit dipahami, khususnya dalam aplikasi klinis yang membutuhkan transparansi tinggi. Untuk menjawab tantangan ini, pendekatan *Explainable AI (XAI)* seperti *SHAP* dan *LIME* telah dikembangkan untuk meningkatkan transparansi model, memungkinkan analisis kontribusi fitur dan penjelasan lokal terhadap prediksi individu [18]-[21].

Penelitian ini bertujuan untuk membandingkan performa *Random Forest* dan *XGBoost* dalam klasifikasi diabetes, dengan fokus pada akurasi serta interpretabilitas menggunakan metode *XAI* seperti *SHAP* dan *LIME*. Dengan menggunakan dataset publik yang mencakup 70.000 entri, penelitian ini mengkaji bagaimana kedua algoritma ini dapat memberikan prediksi yang akurat dan transparan. Gap yang diidentifikasi adalah kurangnya penelitian yang mengintegrasikan metode interpretasi seperti *SHAP* dan *LIME* dengan algoritma *Random Forest* dan *XGBoost* dalam konteks klasifikasi diabetes. Selain itu, dibandingkan dengan algoritma lain seperti *KNN*, *Naive Bayes*, atau *SVM*, *Random Forest* dan *XGBoost* lebih sesuai untuk diintegrasikan dengan metode interpretasi *XAI* (*SHAP* dan *LIME*), karena kemampuan keduanya dalam menangani dataset berskala besar dan data dengan distribusi kompleks. *Random Forest* menawarkan stabilitas dan efisiensi waktu pelatihan melalui pendekatan *bagging*, sedangkan *XGBoost* unggul dalam akurasi tinggi dan generalisasi berkat pendekatan *boosting*. Dengan integrasi metode *XAI* ini, penelitian ini bertujuan tidak hanya mengisi kekosongan tersebut, tetapi juga memberikan rekomendasi model yang dapat diandalkan dan mudah diinterpretasikan untuk mendukung pengambilan keputusan klinis yang lebih baik.

II. METODE PENELITIAN

Penelitian ini bertujuan membandingkan performa *Random Forest* dan *XGBoost* dalam klasifikasi diabetes dengan menggunakan teknik interpretasi *SHAP* dan *LIME* untuk memberikan penjelasan yang transparan terhadap keputusan model. Penelitian dilakukan menggunakan pendekatan kuantitatif dengan desain eksperimen. Dataset yang digunakan berisi 70.000 entri dengan 34 kolom, di mana satu kolom berisi target berupa jenis diabetes (klasifikasi multi-kelas) dan 33 kolom lainnya merupakan fitur medis pasien seperti usia, tekanan darah, kadar gula, dan informasi kesehatan lainnya. Dataset ini tersedia dalam format *.xlsx.*

Tahapan penelitian meliputi pengumpulan data, preprocessing, pelatihan model, dan evaluasi kinerja berdasarkan metrik seperti *akurasi*, *presisi*, dan *F1-Score*. Selain itu, dilakukan analisis visual menggunakan grafik *SHAP* dan *LIME* untuk menjelaskan faktor-faktor yang mempengaruhi keputusan model. Perancangan desain penelitian yang mencakup langkah-langkah ini dapat dilihat pada Gambar 1.



Gambar 1. Perancangan Desain Penelitian

Proses pengolahan data dilakukan dalam beberapa tahap:

a. *Encoding* Kategorikal: Fitur kategorikal yang ada dalam *dataset* diubah menjadi numerik menggunakan *LabelEncoder* untuk memastikan bahwa model *machine learning* dapat memproses data dengan benar.
b. Normalisasi Data: Fitur numerik distandarisasi menggunakan *StandardScaler* untuk mengurangi bias yang diakibatkan oleh skala fitur yang berbeda. Normalisasi dilakukan dengan rumus *Z-score* sebagai berikut:

$$z = \frac{x-\mu}{\sigma}$$

Dimana:

- x adalah nilai data yang akan distandarisasi,
- μ adalah rata-rata dari *dataset*,
- σ adalah *deviasi* standar dari *dataset*.

c. Pemilahan Target dan Fitur : Kolom target yang menunjukkan jenis diabetes dipisahkan dari fitur yang menunjukkan faktor atau penyebab dari diabetes lainnya.

d. Pembagian Data : *Dataset* dibagi menjadi dua bagian, yaitu data pelatihan (70%) dan data pengujian (30%) menggunakan metode *train_test_split* dengan stratifikasi berdasarkan distribusi target.

Model yang digunakan dalam penelitian ini adalah *Random Forest* dan *XGBoost*. Kedua model ini dipilih karena kemampuannya dalam menangani masalah klasifikasi multi-kelas dan kemampuannya untuk mengatasi *overfitting*. Pada tahap ini, dilakukan pencarian *hyperparameter* optimal dengan menggunakan *Grid Search* dan *CrossValidation* untuk mendapatkan kombinasi parameter terbaik yang meningkatkan kinerja model.

a. *Random Forest*

Random Forest adalah model ensemble yang terdiri dari sekumpulan pohon keputusan yang dilatih pada subset data yang berbeda. Setiap pohon keputusan dalam *Random Forest* memberikan hasil prediksi dan hasil akhir diputuskan berdasarkan mayoritas suara (*voting*) dari semua pohon. Fungsi prediksi untuk setiap pohon keputusan dapat dinyatakan dalam rumus berikut:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

Dimana :

- $\hat{y}$ adalah prediksi hasil model,
- N adalah jumlah pohon keputusan dalam Random Forest,

- $f_i(x)$ adalah fungsi prediksi dari pohon ke-i untuk input x.

Pada proses pelatihan, *Random Forest* menggunakan teknik *bagging* (*Bootstrap Aggregating*) untuk mengurangi varians dan meningkatkan akurasi prediksi. Selain itu, *Grid Search* digunakan untuk mencari kombinasi parameter optimal, seperti jumlah pohon ($n_{estimators}$) kedalaman maksimum pohon (*max_depth*) serta parameter lainnya yang mempengaruhi performa model.

b. *XGBoost*

XGBoost (*Extreme Gradient Boosting*) adalah algoritma pembelajaran berbasis pohon yang mengimplementasikan teknik *Gradient Boosting*. Berbeda dengan *Random Forest* yang menggunakan pemungutan suara dari beberapa pohon, *XGBoost* membangun model secara bertahap, di mana setiap pohon baru berusaha mengoreksi kesalahan dari pohon sebelumnya. Fungsi prediksi untuk model *boosting* dapat dinyatakan dengan rumus berikut:

$$F(x)=\sum_{k=1}^{K} f_k(x)$$

Dimana :

- $F(x)$ adalah hasil prediksi untuk input x,

- $f_k(x)$ adalah fungsi pohon keputusan pada iterasi ke-k,

- K adalah jumlah iterasi atau *estimators*.

XGBoost mengoptimalkan fungsi loss menggunakan algoritma *gradient descent* untuk memperbaharui bobot pada setiap pohon, sehingga model menjadi lebih sensitif terhadap data pelatihan. Proses ini mengarah pada peningkatan performa prediksi dengan mengurangi kesalahan secara bertahap.

c. Implementasi Model

Untuk membangun model prediksi, 2 model digunakan, yaitu *Random Forest* dan *XGBoost*, dengan konfigurasi yang telah dioptimalkan menggunakan *GridSearchCV* untuk mendapatkan parameter terbaik. *Grid search* ini memastikan bahwa setiap model diuji dengan berbagai kombinasi parameter untuk menemukan konfigurasi yang paling optimal dalam hal kinerja.

Model pelatihan dilakukan dengan menggunakan *X_train* dan *y_train*, di mana *X_train* adalah fitur yang telah diproses dan *y_train* adalah target yang telah di*encode*. Setelah model dilatih, *X_test* dan *y_test* digunakan untuk mengukur kinerja model pada data yang belum pernah dilihat sebelumnya. Evaluasi model dilakukan dengan menghitung metrik-metrik berikut :

Akurasi: Menghitung persentase prediksi yang benar.

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Prediksi} \times 100\%$$

Laporan Klasifikasi: menggunakan *classification_report* untuk mendapatkan metrik seperti *precision* (P), *recall* (R), dan *F1-score* (F_1) untuk setiap kelas, yang dihitung sebagai berikut:

Precision : menunjukkan seberapa banyak prediksi yang benar-benar positif dari semua prediksi yang dilakukan sebagai positif.

$$P = \frac{TP}{TP + FP}$$

Recall : menunjukkan seberapa banyak kasus positif yang benar-benar berhasil diprediksi oleh model.

$$R = \frac{TP}{TP + FN}$$

F1-score : menggabungkan *precision* dan *recall* menjadi satu nilai tunggal untuk memberikan gambaran keseluruhan dari kinerja model.

$$F_1 = 2 \times \frac{P \times R}{P + R}$$

TP, FP, dan FN masing-masing adalah *true positive*, *false positive*, dan *false negative*.

Matriks konfusi (*Confusion Matrix*): digunakan untuk menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas, sehingga kita dapat melihat seberapa baik model dalam mengklasifikasikan data. *Matriks* ini membantu untuk menilai akurasi model secara lebih detail dengan membandingkan hasil prediksi dan label sebenarnya.

Setelah model dilatih dan dievaluasi, dilakukan interpretasi hasil prediksi dengan menggunakan teknik *XAI* untuk memberikan pemahaman yang lebih dalam mengenai keputusan model.

d. *SHAP* (*SHapley Additive Explanations*)

SHAP digunakan untuk menjelaskan kontribusi setiap fitur terhadap keputusan model. Dengan *SHAP*, nilai kontribusi setiap fitur dapat dilihat baik pada level *global* (mengidentifikasi fitur yang paling berpengaruh dalam model) maupun pada level lokal (menjelaskan keputusan model untuk data tertentu). Visualisasi *SHAP* digunakan untuk menunjukkan fitur-fitur yang memiliki dampak terbesar terhadap klasifikasi yang dihasilkan oleh kedua model.

e. *LIME* (*Local Interpretable Model-agnostic Explanations*)

LIME digunakan untuk memberikan penjelasan lokal bagi prediksi model. *LIME* bekerja dengan cara membuat model sederhana yang dapat dipahami di sekitar prediksi tertentu dan menjelaskan kontribusi setiap fitur terhadap hasil prediksi. Penjelasan ini memungkinkan pemahaman yang lebih baik mengenai bagaimana model memutuskan kelas untuk setiap entri data.

III. HASIL DAN PEMBAHASAN

Pada bagian ini dibahas perbandingan hasil implementasi model prediksi diabetes menggunakan algoritma Random Forest dan XGBoost serta interpretasi menggunakan teknik XAI, yaitu LIME dan SHAP. Evaluasi dilakukan untuk mengukur kinerja model dan memahami pengambilan keputusan berdasarkan fitur. Sistem dibangun menggunakan dataset berisi 70.000 entri dengan 34 fitur medis, termasuk satu kolom target yang merepresentasikan jenis diabetes. Hyperparameter model dioptimalkan menggunakan Grid Search dengan 3-fold CrossValidation untuk menemukan kombinasi parameter terbaik. Hasil Grid Search untuk Random Forest dan XGBoost disajikan pada tabel berikut.

Tabel 1. Hasil Grid Search untuk RF dan XGBoost

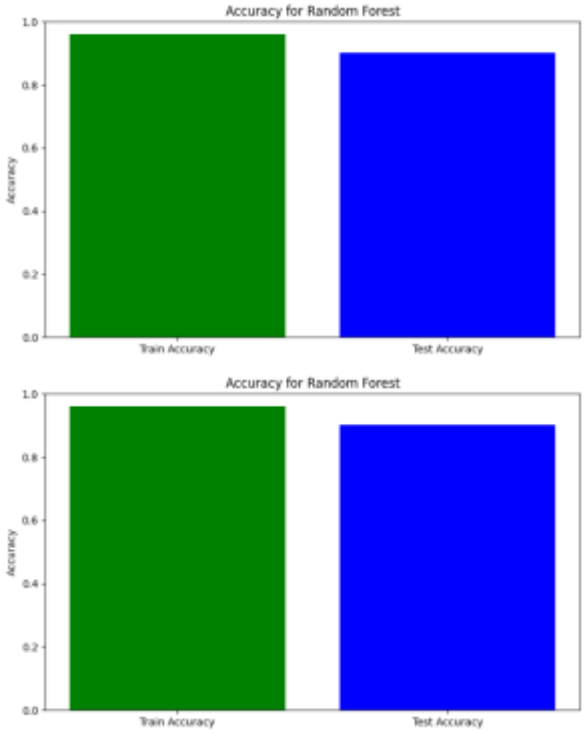
Model	Parameter	Akurasi(%)	*Fit Time* (detik)
RF	n_estimators=500, max_depth=20, min_samples_split=10, min_samples_leaf=1	90,3	16,2
XGB	n_estimators=200, learning_rate=0.1, gamma=1, max_depth=3, min_child_weight=1	90,6	136

Pada Tabel 1 menyajikan hasil pengoptimalan hyperparameter yang dilakukan pada model *Random*

Forest dan *XGBoost*. Pada *Random Forest*, kombinasi parameter terbaik menghasilkan akurasi 90,3% dengan waktu pelatihan 16,2 detik. Sementara itu, *XGBoost* menunjukkan akurasi yang sedikit lebih tinggi, yaitu 90,6%, meskipun waktu pelatihannya lebih lama, yakni 136 detik. Hasil ini menunjukkan bahwa meskipun *XGBoost* memiliki keunggulan dalam akurasi namun *Random Forest* dapat menjadi pilihan yang lebih efisien jika waktu pelatihan menjadi pertimbangan utama. Pemilihan model terbaik bergantung pada kebutuhan dan prioritas pengguna, baik dari segi performa akurasi maupun efisiensi waktu.

Dalam evaluasi ini, kedua model menunjukkan performa yang baik dalam mengklasifikasikan jenis diabetes. Namun, terdapat perbedaan signifikan dalam beberapa aspek kinerja antara *Random Forest* dan *XGBoost* yang dapat digunakan untuk menilai keunggulan masing-masing model.

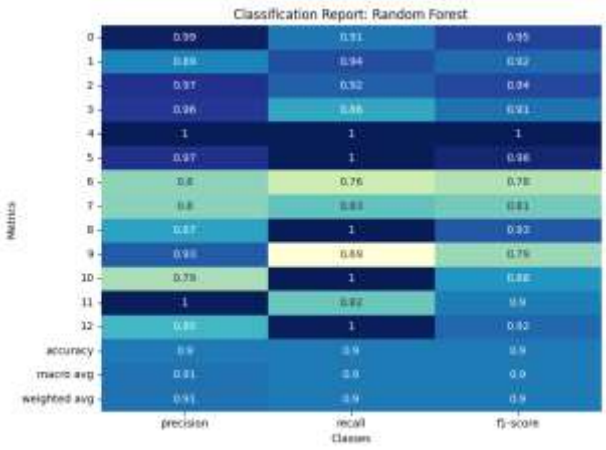
a. Perbandingan Akurasi



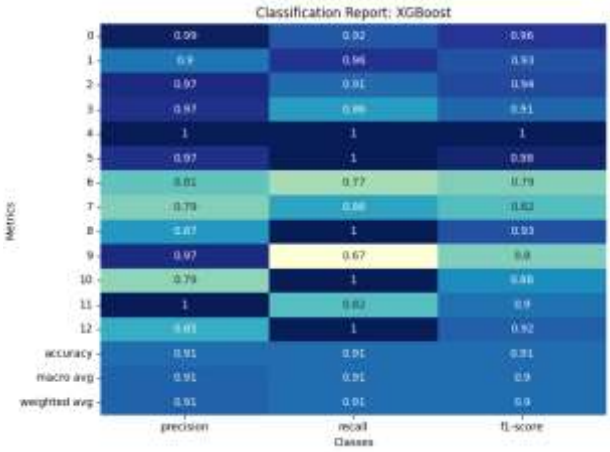
Gambar 2. Perbandingan Akurasi RF dan XGBoost

Pada gambar 2 dapat dilihat jika *Random Forest* memiliki akurasi pelatihan sebesar 96,13%, jauh lebih tinggi dibandingkan *XGBoost* yang hanya mencapai 90,85%. Hal ini menunjukkan bahwa *Random Forest* lebih mampu mempelajari pola dari data pelatihan. Namun, pada data pengujian, *XGBoost* mencatat akurasi yang sedikit lebih tinggi 90,58% dibandingkan *Random Forest* 90,22%. Kesenjangan antara akurasi pelatihan dan pengujian pada *XGBoost* lebih kecil dibandingkan *Random Forest*. Hal ini mengindikasikan bahwa *XGBoost* memiliki generalisasi yang lebih baik dengan risiko *overfitting* yang lebih rendah.

b. Laporan Klasifikasi



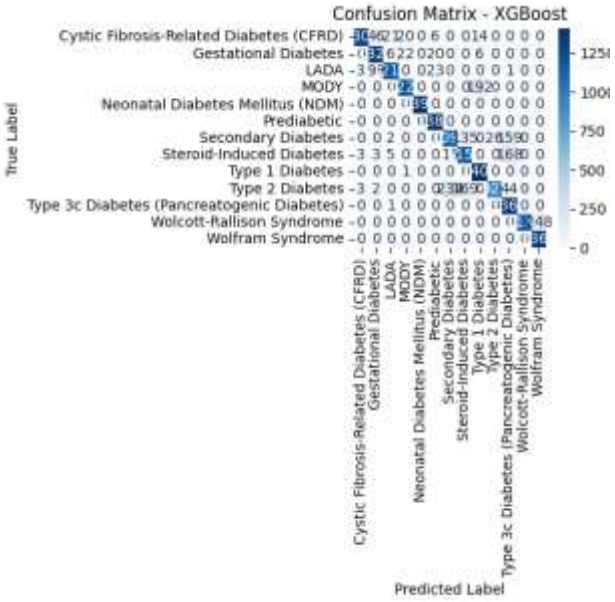
Gambar 3. Klasifikasi Random Forest



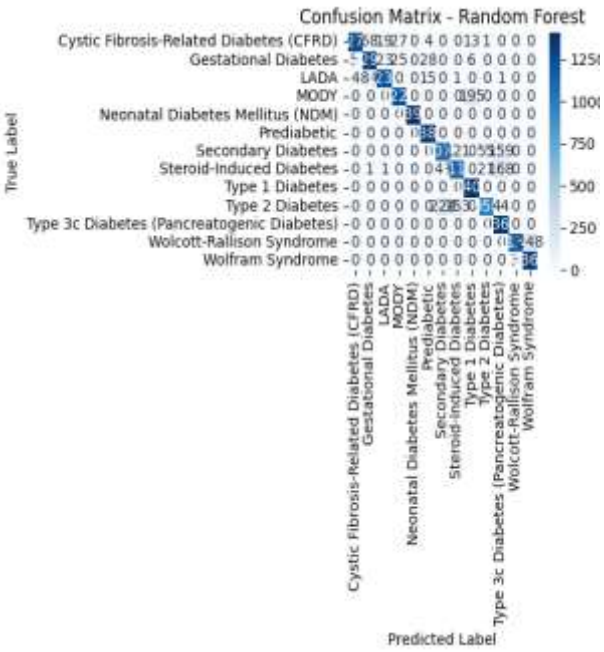
Gambar 4. Klasifikasi XGBoost

Berdasarkan gambar 3 dan 4 dapat dilihat pada *matrix precision*, *recall* dan *F1-score*, kedua model memiliki performa rata-rata yang hampir setara dengan nilai 0,90. Namun, jika dilihat secara detail pada kelas tertentu, *XGBoost* unggul dalam menangani kelas-kelas yang memiliki kompleksitas lebih tinggi. *XGBoost pada* kelas 6 dan 7 misalnya memiliki *precision* sebesar 0,81 dan 0,79 sedikit lebih baik dibandingkan *Random Forest* yang mencatat *precision* sebesar 0,80 dan 0,80. *Random Forest* unggul pada kelas dengan distribusi data yang konsisten, seperti kelas 4 dan 5, dimana *precision* dan recall mencapai nilai sempurna (1,00). Sementara, pada kelas dengan distribusi yang lebih bervariasi atau seimbang, *XGBoost* mampu mempertahankan performa yang lebih stabil.

c. *Confusion Matrix*



Gambar 5. Confusion Matrix XGBoost



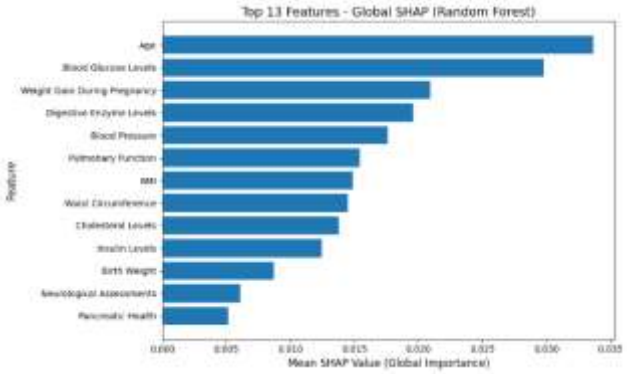
Gambar 6. Confusion Matrix Random Forest

Pada analisis *confusion matrix*, *Random Forest* cenderung membuat kesalahan lebih tinggi pada kelas tertentu, seperti kelas 9 dan 11, dengan distribusi prediksi yang lebih menyebar ke kelas lain. Sebaliknya, *XGBoost* mampu mempertahankan konsistensi prediksi yang lebih baik, terutama pada kelas dengan jumlah sampel kecil. *XGBoost* pada kelas 9 misalnya, mencatat lebih sedikit kesalahan prediksi dibandingkan *Random Forest* ini menunjukkan kemampuan model untuk menangani kelas dengan distribusi yang lebih kompleks.

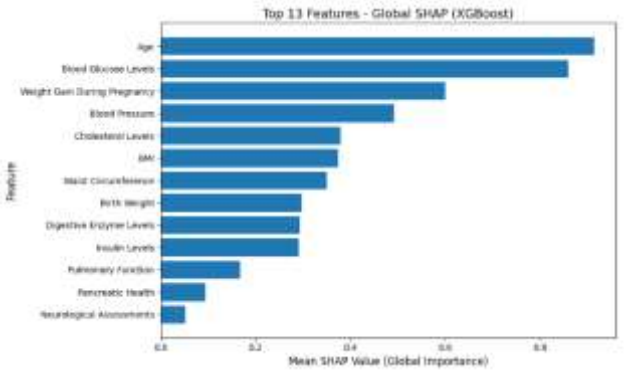
Secara keseluruhan, *XGBoost* menunjukkan keunggulan dalam generalisasi dan stabilitas performa pada kelas dengan kompleksitas lebih tinggi. Sementara itu, *Random Forest* cenderung lebih unggul dalam mempelajari pola pada data pelatihan dan pada kelas-kelas dengan distribusi data yang jelas. Pemilihan model yang terbaik dapat disesuaikan dengan kebutuhan seperti apakah fokus utama pada akurasi pelatihan atau kemampuan generalisasi pada data pengujian.

d. Interpretasi Model Menggunakan *XAI*

Pada penelitian ini, interpretasi model dilakukan menggunakan dua pendekatan *explainable AI (XAI)*, yaitu *SHAP (SHapley Additive exPlanations)* dan *LIME (Local Interpretable Model-Agnostic Explanations)*. Untuk analisis global menggunakan SHAP, data testing sebanyak 21.000 entri digunakan untuk memahami pengaruh fitur secara menyeluruh terhadap prediksi model. Sementara itu, LIME diterapkan pada data baru yang tidak termasuk dalam dataset training untuk melihat interpretasi pada tingkat lokal.

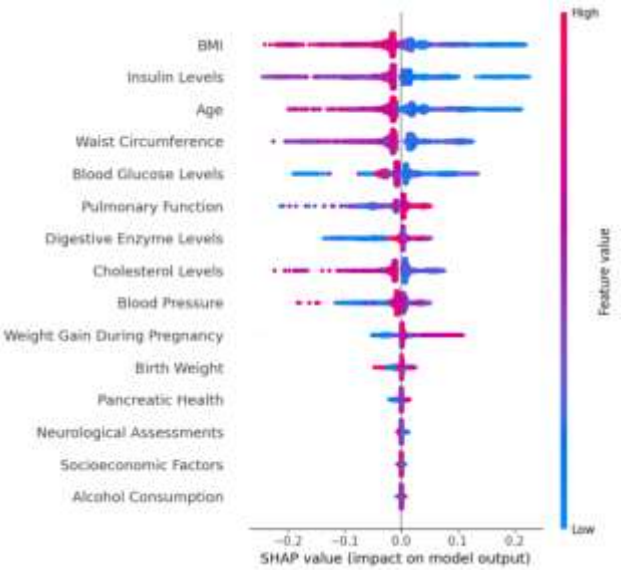


Gambar 7. Analisis Global SHAP untuk semua target (Random Forest)

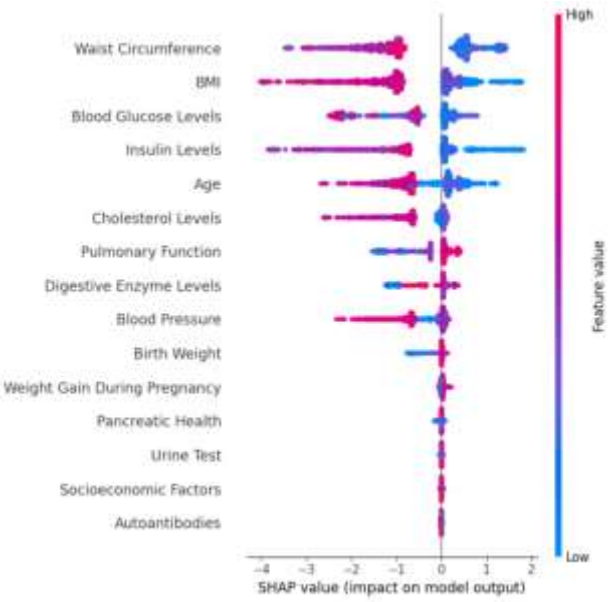


Gambar 8. Analisis Global SHAP untuk semua target (XGBoost)

Berdasarkan gambar 7 dan 8 dapat dilihat jika *Random Forest* ini menunjukkan bahwa fitur *Age* (usia), *Blood Glucose Levels* (kadar glukosa darah), dan *Weight Gain During Pregnancy* (kenaikan berat badan selama kehamilan) merupakan tiga fitur yang paling berpengaruh dalam prediksi. Fitur-fitur ini memberikan kontribusi rata-rata yang konsisten terhadap prediksi model, menunjukkan bahwa *Random Forest* sangat bergantung pada informasi klinis utama untuk menghasilkan prediksi. Lalu untuk *XGBoost* memberikan bobot yang lebih signifikan pada fitur-fitur utama, dengan *Age* dan *Blood Glucose Levels* tetap menjadi yang paling berpengaruh, namun dengan nilai kontribusi yang jauh lebih besar dibandingkan Random Forest. Selain itu, fitur seperti *Weight Gain During Pregnancy* dan *Blood Pressure* (tekanan darah) juga memainkan peran penting dalam keputusan model XGBoost, mencerminkan sensitivitas model ini terhadap parameter fisiologis yang lebih kompleks. Dari kedua model, dapat dilihat bahwa *Age* dan *Blood Glucose Levels* konsisten menjadi fitur andalan dalam memprediksi hasil, namun *XGBoost* cenderung lebih tajam dalam membedakan pengaruh fitur dengan bobot kontribusi yang lebih tinggi, khususnya pada fitur utama. Hal ini menunjukkan bahwa XGBoost mampu menangkap pola hubungan yang lebih kompleks dibandingkan Random Forest.



Gambar 9. Analisis Global SHAP per Target (Random Forest)



Gambar 10. Analisis Global SHAP per Target (XGBoost)

Analisis *SHAP* global juga dapat mengungkap fitur utama yang memengaruhi prediksi untuk satu target contohnya diabetes tipe 1 yang mana nanti menunjukkan perbedaan fokus antara kedua model. Pada *Random Forest* ini Fitur *BMI* (0,032), *Insulin Levels* (0,031) dan *Age* (0,029) menjadi yang paling berpengaruh. Model ini memanfaatkan kombinasi fitur metabolik seperti *Blood Glucose Levels* dan *Pulmonary Function* dengan kontribusi yang merata. Ini menunjukkan *Random Forest* lebih stabil dalam mempertimbangkan berbagai faktor. Lalu untuk *XGBoost* lebih menonjolkan *Waist Circumference* sebagai fitur utama dengan nilai *SHAP* tertinggi (1,042), diikuti *BMI* (0,876) dan *Blood Glucose Levels* (0,747). XGB lebih agresif dalam memberi bobot pada fitur-fitur utama, menandakan kemampuannya untuk menangkap hubungan kompleks dengan lebih tajam. Kedua model menunjukkan jika mereka sepakat bahwa fitur metabolik seperti *BMI* dan *Insulin Levels* sangat penting, tetapi *XGBoost* memberikan perspektif yang lebih mendalam dengan menyoroti pengaruh besar pada beberapa fitur utama. Visualisasi dari analisis ini dapat dilihat pada Gambar di atas.

Data baru yang tidak termasuk dalam *dataset training* maupun *testing* dianalisis menggunakan *LIME* untuk memahami prediksi model. Setelah memuat data, kolom *Target* dihapus (jika ada), fitur kategori di-*label encode*, dan data dinormalisasi dengan *Z-score*. Data ini kemudian diprediksi menggunakan model *Random Forest (RF)* dan *XGBoost (XGB)*.

Gambar 11. Analisis Lime untuk Random Forest

Gambar 12. Analisis Lime untuk XGBoost

Random Forest memprediksi *Cystic Fibrosis-Related Diabetes (CFRD)* dengan probabilitas 0,75. Fitur yang paling berpengaruh adalah *Age* (<= -0,85) dengan kontribusi terbesar (0,14), diikuti *Pulmonary Function* (0,07) dan *Weight Gain During Pregnancy* (0,05). Prediksi ini mengandalkan kombinasi fitur Age, Pulmonary Function, dan Weight Gain During Pregnancy.

XGBoost memprediksi juga CFRD, namun dengan probabilitas lebih tinggi, yaitu 0,97. Fitur utama adalah *Weight Gain During Pregnancy* (> 0,05) dan *Age* (<= -0,85), masing-masing berkontribusi 0,13. Model ini lebih fokus pada beberapa fitur utama, seperti *Pulmonary Function* dan *Insulin Levels*. Hasil ini menunjukkan bahwa kedua model sepakat dalam prediksi, dengan XGBoost lebih percaya diri dan fokus pada fitur utama. Gambar 11 dan 12 akan memperlihatkan bagaimana fitur mempengaruhi hasil prediksi ini.

Hasil penelitian ini menunjukkan peningkatan signifikan dibandingkan penelitian terdahulu. Penelitian sebelumnya hanya menggunakan dataset kecil dengan 768 entri dan 9 fitur, yang menghasilkan akurasi sebesar 70% untuk *Random Forest* dan 76% untuk *XGBoost* [22]. Sebaliknya, penelitian ini menggunakan dataset yang jauh lebih besar yang terdiri dari 70.000 entri dengan peningkatan jumlah data dan kelengkapan fitur medis memainkan peran penting dalam peningkatan performa model. Dataset yang lebih besar membantu model memahami pola yang lebih kompleks, sementara fitur medis yang lebih beragam memberikan informasi tambahan yang relevan untuk membuat prediksi yang lebih akurat. Hal ini sejalan dengan prinsip *machine learning* dimana kualitas dan kuantitas data yang digunakan memiliki dampak langsung pada hasil model.

Hasil ini juga menunjukkan pentingnya pengoptimalan *hyperparameter* yang dilakukan dalam penelitian ini. Dengan memanfaatkan *GridSearch* dan *CrossValidation*, model dapat disesuaikan untuk mencapai performa terbaiknya. Ini menjadi bukti bahwa pendekatan yang lebih menyeluruh dalam pemrosesan data dan pelatihan model dapat memberikan hasil yang jauh lebih unggul dibandingkan pendekatan yang sederhana.

Penelitian ini tidak hanya menunjukkan peningkatan dari segi akurasi tetapi juga memberikan wawasan bahwa penggunaan dataset yang lebih kaya dan teknik pengoptimalan yang tepat dapat menjadi kunci untuk menghasilkan model prediktif yang andal, terutama dalam bidang medis.

IV. KESIMPULAN

Penelitian ini membandingkan Random Forest dan XGBoost dalam klasifikasi diabetes serta mengukur interpretabilitasnya melalui teknik *XAI* yakni *SHAP dan LIME*. Hasilnya, *XGBoost* unggul dalam generalisasi dan akurasi (90,6%) dengan risiko *overfitting* yang lebih rendah, sementara *Random Forest* menawarkan efisiensi waktu pelatihan (16,2 detik) dan performa solid pada pola data konsisten. Teknik *XAI* mempermudah interpretasi klinis, memberikan transparansi dan membangun kepercayaan pada keputusan model. Untuk penelitian selanjutnya, pengujian pada data klinis dan integrasi metode interpretasi baru dapat memperluas manfaat model ini dalam pengambilan keputusan medis.

REFERENSI

[1] P. Saeedi, I. Petersohn, P. Salpea, et al., "Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition," *Diabetes Research and Clinical Practice*, vol. 157, p. 107843, 2019, doi: 10.1016/j.diabres.2019.107843.

[2] International Diabetes Federation, *IDF Diabetes Atlas*, 10th ed., 2021. [Online]. Available: https://diabetesatlas.org.

[3] Reuters, "*More than 800 million adults have diabetes globally, many untreated, study suggests*," Nov. 13, 2024. [Online]. Available: https://www.reuters.com/business/healthcare-pharmaceuticals/more-than-800-million-adults-have-diabetes-globally-many-untreated-study-2024-11-13/.

[4] World Health Organization, "*Diabetes*," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes.

[5] D. J. Magliano, R. M. Islam, E. L. M. Barr, et al., "*Diabetes mellitus, the fastest growing global public health concern*," The Lancet, vol. 403, no. 10399, pp. 1234-1245, 2024, doi: 10.1016/S0140-6736(24)00345-2.

[6] Ciputra Hospital, "*Komplikasi Diabetes Mellitus*," 2022. [Online]. Available: https://ciputrahospital.com/komplikasi-diabetes-mellitus/.

[7] A. Suherman, H. Harun, and D. Deni, "*Dampak Diabetes Melitus Tipe 2 pada Fungsi Ginjal: Faktor Risiko, Gejala, dan Pentingnya Manajemen yang Tepat*," Jurnal Keperawatan Thamrin, vol. 1, no. 1, pp. 15-16, 2023. [Online]. Available: https://eresources.thamrin.ac.id/1629/2/Winda%20Wardianti_S1%20Keperawatan_2024%20-%20Bab%201.pdf.

[8] W. Wardianti, *"Bab 1 Pendahuluan: Diabetes Melitus Tipe 2 dan Komplikasinya,"* Jurnal Keperawatan Thamrin, vol. 1, no. 1, pp. 1-16, 2024. [Online]. Available: https://eresources.thamrin.ac.id/1629/2/Winda%20Wardianti_S1%20Keperawatan_2024%20-%20Bab%201.pdf.

[9] H. Yin, B. Mukadam, X. Dai, and N. K. Jha, "DiabDeep: Pervasive Diabetes Diagnosis based on Wearable Medical Sensors and Efficient Neural Networks," *arXiv preprint arXiv:1910.04925, 2019*. [Online]. Available: https://arxiv.org/abs/1910.04925.

[10] Z. Zhang, K. A. Ahmed, M. R. Hasan, T. Gedeon, and M. Z. Hossain, "A Deep Learning Approach to Diabetes Diagnosis," *arXiv preprint arXiv:2403.07483, 2024.* [Online]. Available: https://arxiv.org/abs/2403.07483.

[11] S. Hardani, "Diagnosa Penyakit Diabetes dengan Metode Forward Chaining," *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, vol. 5, no. 2, pp. 231–236, 2020, doi: 10.33480/jitk.v5i2.1132.

[12] Kementerian Kesehatan Republik Indonesia, "*RSCM Luncurkan Tes Genomik Pengobatan Presisi Untuk Diabetes, Kolesterol Tinggi, dan Nutrisi*," Sehat Negeriku, Dec. 23, 2024. [Online]. Available: https://sehatnegeriku.kemkes.go.id/baca/umum/20241223/4247160/rscm-luncurkan-tes-genomik-pengobatan-presisi-untuk-diabetes-kolesterol-tinggi-nutrisi/.

[13] S. Saru and S. Subashree, "Analysis and prediction of diabetes using machine learning," *International Journal of Emerging Technology and Innovative Engineering*, vol. 5, no. 4, 2019.

[14] I. Tasin, T. U. Nabil, S. Islam, and R. Khan, "Diabetes prediction using machine learning and explainable AI techniques," *Healthcare Technology Letters*, vol. 10, no. 1-2, pp. 1-10, 2023.

[15] S. M. Ganie, P. K. D. Pramanik, M. Bashir Malik, S. Mallik, and H. Qin, "An ensemble learning approach for diabetes prediction using boosting techniques," *Frontiers in Genetics*, vol. 14, p. 1252159, 2023.

[16] Xue, J., Min, F., & Ma, F. (2020, November). Research on diabetes prediction method based on machine learning. In Journal of Physics: Conference Series (Vol. 1684, No. 1, p. 012062). IOP Publishing.

[17] Vijayan, V. V., & Anjali, C. (2015, December). Prediction and diagnosis of diabetes mellitus—A machine learning approach. In 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) (pp. 122-127). IEEE.

[18] Zhou, H., Xin, Y., & Li, S. (2023). A diabetes prediction model based on Boruta feature selection and ensemble learning. BMC bioinformatics, 24(1), 224.

[19] Zohair, M., Chandra, R., Tiwari, S., & Agarwal, S. (2024). A model fusion approach for severity prediction of diabetes with respect to binary and multiclass classification. International Journal of Information Technology, 16(3), 1955-1965.

[20] Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. J Big Data 6: 54.

[21] Doğru, A., Buyrukoğlu, S., & Arı, M. (2023). Model pembelajaran super ensemble hibrida untuk prediksi risiko diabetes tahap awal. Teknik & Komputasi Medis & Biologi , 61 (3), 785-797.

[22] Salsabil, M., Azizah, N. L., & Eviyanti, A. (2024). Implementasi Data Mining Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode Random Forest Dan Xgboost. Jurnal Ilmiah Komputasi, 23(1), 51-58.