JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Komparasi Algoritma XGBoost dan Logistic Regression Berbasis Explainable AI (SHAP) untuk Deteksi Hipertensi

Risky Radison Nasution, KurniaBudi, Dodo Zaenal Abidin*

Fakultas Ilmu Komputer, Program Studi Magister Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia
Email: [1]riskyradisonnasution@gmail.com, [2]kurniabudi@unama.ac.id, [3,*]dodozaenalabidin@gmail.com
Email Penulis Korespondensi: [1]riskyradisonnasution@gmail.com
Submitted **30-04-2026;** Accepted **17-06-2026;** Published **30-06-2026**

Abstrak

Hipertensi merupakan tantangan kesehatan global yang signifikan karena sifatnya yang sering kali tidak menunjukkan gejala hingga terjadi komplikasi serius. Penelitian ini bertujuan untuk melakukan komparasi mendalam antara algoritma XGBoost dan Logistic Regression dalam deteksi dini hipertensi dengan menitikberatkan pada aspek Explainability menggunakan metode SHAP Permutation Explainer. Masalah utama yang diangkat adalah adanya kecenderungan manipulasi dataset melalui teknik oversampling pada penelitian terdahulu yang berisiko mengubah distribusi alami fitur klinis serta adanya "bias explainer" akibat penggunaan metode penjelasan yang tidak seragam. Metode penelitian ini menggunakan data asli dari Behavioral Risk Factor Surveillance System (BRFSS) tahun 2015 guna memastikan hasil yang reliabel secara klinis. Hasil penelitian menunjukkan bahwa model XGBoost secara konsisten mengungguli Logistic Regression pada seluruh parameter dengan XGBoost mencapai AUC 0,805 dan akurasi 73,4% dibandingkan Logistic Regression dengan AUC 0,802 dan akurasi 73,1%. Interpretasi menggunakan SHAP mengungkapkan bahwa fitur BMI dan Usia merupakan prediktor paling dominan dalam kedua model. Kontribusi utama dari penelitian ini adalah penyediaan kerangka kerja prediksi menggunakan distibusi data asli yang menghasilkan prediksi lebih objektif dibandingkan manipulasi sintetis, sehingga dapat menjadi referensi kredibel bagi praktisi medis untuk memahami faktor risiko hipertensi secara transparan dan akuntabel.

Kata Kunci: Hipertensi; Machine Learning; XGBoost; Explainable AI; SHAP

Abstract

Hypertension represents a significant global health challenge due to its frequently asymptomatic nature until serious complications arise. This study aims to conduct an in-depth comparison between XGBoost and Logistic Regression algorithms for the early detection of hypertension, with a specific focus on explainability utilizing the SHAP Permutation Explainer method. The primary issue addressed is the tendency of previous research to manipulate datasets through oversampling techniques, which risks altering the natural distribution of clinical features, alongside the presence of "explainer bias" resulting from the use of non-uniform explanation methods. This research utilizes original data from the 2015 Behavioral Risk Factor Surveillance System (BRFSS) to ensure clinically reliable results. The results demonstrate that the XGBoost model consistently outperforms Logistic Regression across all evaluation metrics, with XGBoost achieving an AUC of 0.805 and an accuracy of 73.4%, compared to Logistic Regression's AUC of 0.802 and accuracy of 73.1%. SHAP-based interpretation reveals that BMI and Age are the most dominant predictors in both models. The primary contribution of this study is the provision of a predictive framework leveraging the original data distribution, yielding more objective predictions compared to synthetic manipulation, thereby serving as a credible reference for medical practitioners to understand hypertension risk factors in a transparent and accountable manner.

Keywords: Hypertension; Machine Learning; XGBoost; Explainable AI; SHAP

1. PENDAHULUAN

Hipertensi atau tekanan darah tinggi merupakan tantangan kesehatan global yang signifikan dan sering dijuluki sebagai "silent killer" karena sifatnya yang sering kali tidak menunjukkan gejala hingga terjadi komplikasi serius[1]. Penyakit ini memicu risiko kesehatan kronis seperti penyakit jantung koroner, stroke, dan gagal ginjal yang berdampak pada penurunan kualitas hidup serta peningkatan beban biaya Kesehatan[2]. Secara epidemiologi, prevalensi hipertensi terus menunjukkan tren peningkatan yang mengkhawatirkan di berbagai belahan dunia. Proyeksi data menunjukkan bahwa jumlah penderita hipertensi diperkirakan akan mencapai 1,5 miliar orang pada tahun 2025, yang mencerminkan kebutuhan mendesak akan sistem deteksi dini yang lebih efisienso[3]. Metode deteksi konvensional yang mengandalkan pengukuran manual dan evaluasi klinis rutin mulai menghadapi keterbatasan terutama dalam hal waktu pemrosesan data dan jangkauan skrining massal. Oleh karena itu, integrasi teknologi *machine learning* menjadi sangat urgensi untuk membantu tenaga medis dalam mengenali pola risiko tersembunyi pada data multidimensional yang kompleks secara cepat dan akurat[4].

Pemanfaatan kecerdasan buatan dalam memprediksi risiko hipertensi telah menjadi subjek penelitian yang intensif dalam beberapa tahun terakhir. Beberapa penelitian telah mengeksplorasi efektivitas berbagai algoritma *machine learning* misalnya pada Penelitian oleh AlKaabi dkk. [5] yang membandingkan *Logistic Regression* dengan model seperti *Random Forest* dan *Support Vector Machine*, serta Chen dkk. [6] yang membuktikan keunggulan *XGBoost* pada tingkat perawatan primer namun belum membandingkan stabilitas fiturnya dengan model linier tradisional. Terkait penanganan data medis, studi seperti Montagna dkk. [7] cenderung menggunakan teknik *oversampling* seperti SMOTE untuk mengatasi *imbalanced data* yang secara empiris berisiko memunculkan bias pada distribusi fitur asli. Pentingnya perbandingan performa antar arsitektur model juga ditekankan oleh Si dkk. [8] yang menerapkan *XGBoost* dan *Logistic Regression* pada pasien lanjut usia. Penelitian tersebut menggarisbawahi bahwa model yang berbeda dapat memberikan prioritas fitur yang berbeda pula yang memperkuat perlunya analisis komparatif yang ketat. Terakhir, penelitian oleh Bisong dkk. [9]

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

mulai memperkenalkan konsep *Explainable AI* (XAI), namun masih memiliki kelemahan berupa penggunaan metode penjelasan yang tidak seragam pada model dengan arsitektur fundamental yang berbeda.

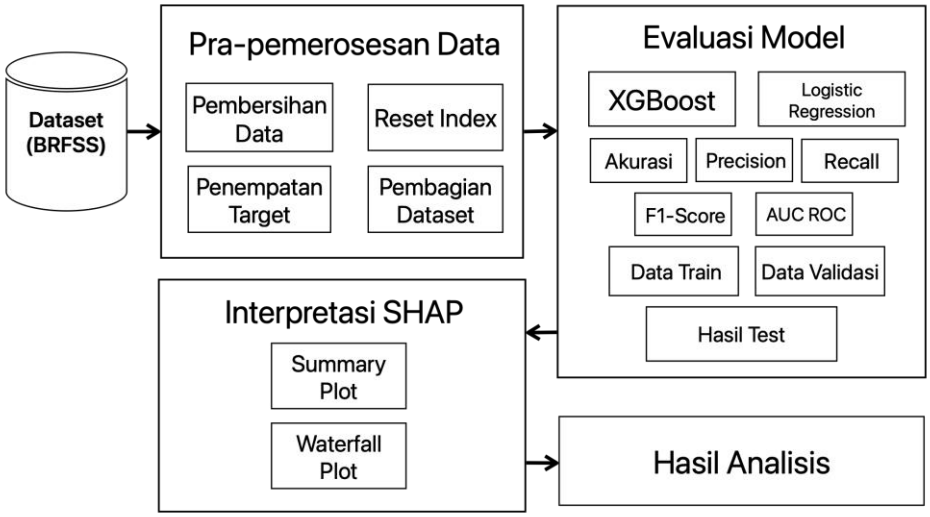
Berdasarkan tinjauan literatur yang telah dipaparkan, muncul sebuah celah penelitian (*GAP Analysis*) yang sangat krusial untuk diselesaikan. Sebagian besar penelitian terdahulu memiliki kecenderungan untuk memanipulasi dataset asli melalui teknik *oversampling* atau *undersampling* demi mengejar performa metrik seperti *Accuracy* atau *F1-Score*. Secara teoritis, tindakan ini berisiko mengubah distribusi alami dari fitur klinis pasien, sehingga penjelasan yang dihasilkan model mungkin tidak lagi mencerminkan realitas medis di lapangan. Selain itu, terdapat masalah "*bias explainer*" di mana penelitian sebelumnya seringkali menggunakan metode penjelasan yang berbeda untuk model yang berbeda misalnya menggunakan koefisien untuk model linier dan *feature importance* berbasis Gini impurity untuk model pohon. Pendekatan yang tidak seragam ini menciptakan ketidakadilan dalam membandingkan "logika" antara model *glass-box* seperti *Logistic Regression* yang transparan secara struktural dan model *black-box* seperti *XGBoost* yang memiliki struktur non-linier kompleks. Hingga saat ini, belum ditemukan studi yang secara spesifik melakukan komparasi *head-to-head* antara *XGBoost* dan *Logistic Regression* pada data asli tanpa manipulasi *oversampling* dengan menggunakan satu lensa penjelasan yang seragam dan bersifat *model-agnostic* yaitu *SHAP Permutation Explainer*.

Tujuan utama dari penelitian ini adalah untuk melakukan komparasi mendalam antara algoritma *XGBoost* dan *Logistic Regression* dalam deteksi dini hipertensi dengan menitikberatkan pada aspek *Explainability*. Dengan menerapkan *SHAP Permutation Explainer* penelitian ini bertujuan untuk memvalidasi secara ilmiah apakah fitur-fitur penting yang dideteksi oleh model *machine learning* kompleks selaras dengan temuan pada model statistik tradisional ketika keduanya diukur menggunakan parameter penjelasan yang identik. Harapan yang ingin dicapai melalui penelitian ini adalah menghasilkan kerangka kerja prediksi yang tidak hanya unggul secara performa teknis, tetapi juga memiliki tingkat transparansi dan akuntabilitas yang tinggi melalui *feature importance* yang konsisten. Hasil dari penelitian ini diharapkan dapat menjadi referensi kredibel bagi praktisi medis untuk memahami faktor risiko hipertensi secara lebih objektif, serta memberikan keyakinan bahwa keputusan yang diambil oleh model *artificial intelligence* dapat dipertanggungjawabkan secara klinis tanpa adanya artefak dari proses manipulasi data sintetis.

2. METODOLOGI PENELITIAN

2.1 Alur Eksperimen

Penelitian ini dijalankan melalui serangkaian tahapan sistematis yang dirancang untuk memastikan integritas data medis tetap terjaga tanpa intervensi manipulasi sintetis. Alur eksperimen secara menyeluruh disajikan pada Gambar 1:



Gambar 1. Alur Eksperimen

Berdasarkan Gambar 1, tahapan penelitian dimulai dengan mempersiapkan data mentah yang diperoleh dari Dataset BRFSS. Data tersebut kemudian masuk ke tahap Pra-pemrosesan Data yang meliputi pembersihan data, pengesetan ulang indeks (*reset index*), penempatan variabel target dan pembagian dataset. Selanjutnya, kedua bagian data ini digunakan pada tahapan Evaluasi Model untuk melatih dan menguji algoritma XGBoost dan Logistic Regression serta mengukur performanya melalui metrik Akurasi, Presisi, Recall, F1-Score, dan AUC ROC. Guna memastikan transparansi model, hasil evaluasi tersebut dibedah pada tahap Interpretasi Model SHAP menggunakan visualisasi *Summary Plot* dan *Waterfall Plot*. Eksperimen ini diakhiri dengan tahapan Analisis Hasil untuk memvalidasi konsistensi logika prediksi secara klinis. Secara rinci, penjelasan dari masing-masing blok tahapan pada Gambar 1 dipaparkan sebagai berikut:

a. Dataset BRFSS

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penelitian ini menggunakan data dari *Behavioral Risk Factor Surveillance System* (BRFSS) tahun 2015 yang dikumpulkan oleh *Centers for Disease Control and Prevention* (CDC). Dataset terdiri dari 253.680 sampel dengan 21 fitur dan 1 target biner (HighBP). Fitur mencakup variabel klinis seperti BMI, HighChol, Diabetes_012, dll. Gaya hidup misalnya Smoker, PhysActivity, DiffWalk, dll. dan sosiodemografi misalnya Age, Income, HvyAlcoholConsump, dll. Variabel target menunjukkan status diagnosis hipertensi (1 = Ya, 0 = Tidak).

b. Pra-pemrosesan Data

Pada tahap awal, data BRFSS diproses agar layak digunakan dalam pemodelan *machine learning* melalui pembersihan data dengan menghapus nilai hilang pada variabel target serta menyaring data tidak valid pada variabel biner dan ordinal untuk meminimalkan noise dan bias, kemudian dilakukan pengaturan ulang indeks dataset agar konsisten dan mudah ditelusuri, serta penempatan variabel target (HighBP) secara tepat guna mendukung proses klasifikasi pada tahap pelatihan dan evaluasi. Setelah dilakukan pembersihan data, pengaturan ulang indeks, dan juga penempatan target variabel selanjutnya dibagi menjadi data pelatihan (*train*) dan data pengujian (*test*) dengan skema 80:20 untuk menguji generalisasi model pada data baru. Skema pembagian ini mengadopsi rasio data latih sebesar 0,8 yang digunakan dalam penelitian oleh AlKaabi dkk. [5] di mana proporsi tersebut terbukti efektif dalam menjaga stabilitas performa model.

c. Evaluasi Model

Evaluasi performa dilakukan untuk mengukur sejauh mana model mampu mengklasifikasikan risiko hipertensi secara akurat. Metrik yang digunakan meliputi Akurasi, Presisi, *Recall*, *F1-Score*, dan *Area Under Curve* (AUC). Seluruh metrik dihitung berdasarkan hasil pengujian pada 20% data uji yang diisolasi. Penerapan *Stratified K-Fold Cross Validation* sangat krusial dalam data medis untuk menangani ketidakseimbangan data dengan menjaga distribusi kelas target tetap konsisten di setiap *fold*, sekaligus mengoptimalkan *bias-variance trade-off* guna meningkatkan kemampuan generalisasi pada data baru (*unseen data*). Melalui mekanisme ini, setiap poin data memiliki kesempatan yang sama untuk divalidasi sehingga mampu meminimalkan fluktuasi prediksi dan menghasilkan penilaian performa yang lebih objektif, andal, serta reliabel untuk lingkungan klinis[10], [11].

d. Interpretasi Model SHAP

Transparansi keputusan model dicapai dengan menerapkan SHAP (*Shapley Additive Explanations*) menggunakan pendekatan *Permutation Explainer*. Penggunaan *Permutation Explainer* dalam alur ini berkontribusi langsung dalam meminimalkan bias interpretasi, mendorong transparansi hasil diagnostik, serta menyediakan kerangka kerja integrasi AI yang etis dan bertanggung jawab dalam praktik klinis nyata[12]. Metode ini dipilih karena bersifat *model-agnostic*, sehingga dapat memberikan standar penilaian yang seragam saat membandingkan kontribusi fitur antara model linear (*Logistic Regression*) dan model kompleks (*XGBoost*). SHAP akan menghitung nilai atribusi untuk setiap fitur yang menunjukkan seberapa besar pengaruh variabel tersebut dalam meningkatkan atau menurunkan probabilitas prediksi hipertensi.

e. Analisis Hasil

Analisis dilakukan dengan membandingkan hasil dari kedua model melalui dua tingkatan interpretasi. Secara global, dilakukan analisis terhadap *Summary Plot* untuk melihat urutan kepentingan fitur pada seluruh dataset. Secara lokal, dilakukan bedah kasus individu menggunakan *Waterfall Plot* untuk memahami bagaimana variabel klinis tertentu mempengaruhi prediksi pada pasien spesifik. Fokus utama analisis ini adalah memvalidasi konsistensi logika model tanpa adanya intervensi data sintetis (*oversampling*) guna memastikan hasil yang reliabel secara klinis.

2.2 Konfigurasi Parameter Model

Penelitian ini tidak menerapkan proses optimasi parameter. Seluruh model klasifikasi dikonfigurasi menggunakan parameter bawaan yang disediakan oleh pustaka *Scikit-Learn* dan *XGBoost*. Pendekatan ini dipilih untuk memperoleh performa dasar yang objektif serta memastikan hasil eksperimen mudah direproduksi. Guna menjaga konsistensi hasil pada setiap tahapan eksperimen digunakan nilai parameter random_state = 1 secara seragam. Rincian spesifik mengenai konfigurasi parameter yang diterapkan pada masing-masing algoritma dapat dilihat pada Tabel 1.

Tabel 1. Konfigurasi Parameter Algoritma dan Metode

Algoritma	Konfigurasi Parameter
Logistic Regression	penalty='l2', C=1.0, solver='lbfgs', max_iter=100, random_state=1
XGBoost	n_estimators=100, max_depth=6, learning_rate=0.3, subsample=1.0, colsample_bytree=1.0, random_state=1

2.3 Machine Learning

Machine Learning (ML) adalah proses pembelajaran otomatis di mana algoritma belajar langsung dari data untuk menemukan representasi input dan memprediksi output tanpa pengkodean eksplisit[13]. Terdapat tiga jenis utama Machine Learning yaitu *Supervised Learning* yang menggunakan data berlabel, *Unsupervised Learning* untuk menemukan pola pada data tak berlabel, dan *Reinforcement Learning* yang melatih agen melalui interaksi lingkungan

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

berdasarkan sistem *reward*[14]. Dalam sektor kesehatan, ML berperan penting dalam analitik prediktif dan pengobatan yang dipersonalisasi melalui pemrosesan data medis yang kompleks untuk meningkatkan hasil perawatan pasien serta mempercepat pengembangan pengobatan [15].

2.4 Logistic Regression

Logistic Regression adalah metode statistik untuk memodelkan hubungan antara satu atau lebih variabel independen terhadap variabel dependen dikotomis atau biner[16], [17]. Metode ini berfungsi untuk memperkirakan probabilitas suatu peristiwa pada data kategoris serta memungkinkan interpretasi koefisien melalui rasio odds, sehingga sangat berguna dalam memprediksi hasil di bidang kedokteran, ilmu sosial, dan pemasaran [18]. Formulasi matematis fundamental yang digunakan untuk memetakan probabilitas biner (0 atau 1) didefinisikan pada Persamaan (1) [19]:

$$P(Y = 1|X) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n)}} \quad (1)$$

Di mana $P(Y = 1|X)$ merepresentasikan probabilitas seseorang menderita hipertensi berdasarkan vektor fitur input X. Nilai e adalah bilangan Euler, sedangkan β_0 bertindak sebagai konstanta atau intersepsi model. Komponen β_1, β_2, ..., β_n mewakili koefisien bobot regresi linier yang mencerminkan arah dan kekuatan kontribusi masing-masing fitur klinis X_1. Keunggulan utama dari struktur matematis ini adalah kemampuannya untuk diekstraksi ke dalam nilai *Odds* OR $= e^{\beta_i}$ sehingga membuat algoritma ini dikategorikan sebagai model *glass-box* yang sangat transparan dan populer dalam analisis epidemiologi medis tradisional [19].

2.5 XGBoost

XGBoost merupakan algoritma *ensemble learning* berbasis *gradient boosting* yang menggabungkan beberapa model lemah untuk meningkatkan akurasi, efisiensi pada data besar, serta ketahanan terhadap *overfitting* melalui optimasi *hyperparameter*[20]. Sebagai salah satu implementasi *boosting* yang populer, algoritma ini memiliki keserbagunaan tinggi pada berbagai jenis data dan terbukti sukses dalam aplikasi visi komputer hingga bahasa alami[21]. Secara teknis, XGBoost membangun model berbasis pohon keputusan secara aditif untuk meminimalkan fungsi kerugian, skalabel, serta memiliki mekanisme efisien dalam menangani data yang hilang[22]. Guna menghasilkan akurasi yang tinggi sekaligus mempertahankan kemampuan generalisasi model, XGBoost meminimalkan fungsi objektif formal yang mengintegrasikan fungsi kerugian terdiferensiasi (*loss function*) dan suku regularisasi pada iterasi ke- t seperti yang dirumuskan pada Persamaan (2) [23]:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (2)$$

Di mana $\mathcal{L}$ melambangkan fungsi kerugian (*loss function*) terdiferensiasi yang mengukur selisih antara target aktual y_i dengan hasil prediksi kumulatif pada iterasi sebelumnya ditambahkan fungsi pohon baru $\hat{y}_i^{(t-1)}$ditambah proyeksi fungsi dari pohon baru $f_t(x_i)$. Komponen $\Omega(f_t)$ merupakan fungsi regularisasi formal yang dirumuskan sebagai $\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$. Suku regularisasi ini mengontrol kompleksitas struktural pohon dengan memberikan penalti terhadap jumlah daun (T) dan bobot skor daun (w). Penambahan fungsi regularisasi intrinsik inilah yang memberikan keunggulan bagi XGBoost berupa ketahanan yang sangat tinggi terhadap kendala *overfitting* serta kemampuan menangani pola data klinis non-linier yang kompleks secara efisien [23], [24].

2.6 Explainable AI

Explainable Artificial Intelligence (XAI) merupakan kumpulan teknik dan strategi yang dirancang untuk memberikan penjelasan transparan terhadap keputusan model pembelajaran mesin guna meningkatkan akuntabilitas dan kepercayaan pengguna . Salah satu metode XAI yang dominan adalah *ShaPley Additive Explanations* (SHAP) yang menggunakan pendekatan teori permainan untuk mengonversi keputusan model yang kompleks menjadi skor kontribusi fitur yang mudah dipahami, sehingga membantu membuka sifat "kotak hitam" pada domain yang sensitif. Meskipun menawarkan keunggulan sistematis dalam mengukur kontribusi fitur, SHAP memiliki keterbatasan berupa kompleksitas komputasi yang tinggi serta risiko interpretasi yang tidak akurat pada data yang kolinear, sehingga pengguna perlu memahami keterbatasan tersebut untuk menghindari penyalahgunaan pada aplikasi krusial [25]. SHAP mengadopsi prinsip *Shapley Values* dari teori permainan kooperatif (*cooperative game theory*) untuk mengonversi model prediktif menjadi model penjelasan aditif, di mana setiap fitur diposisikan sebagai "pemain" yang berkontribusi terhadap "skor permainan" (hasil prediksi) [20]. Formulasi dasar nilai Shapley untuk mengukur nilai atribusi kontribusi fitur i dinyatakan pada Persamaan (3) [26]:

$$\phi_i(x) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f_x(S \cup \{i\}) - f_x(S)] \quad (3)$$

Di mana $\phi_i(x)$ melambangkan kontribusi marjinal atau nilai atribusi dari fitur i terhadap hasil prediksi final pada sampel x. Simbol F melambangkan himpunan seluruh fitur dalam dataset, sedangkan S merupakan subset fitur mengecil yang dipilih dari seluruh kombinasi fitur yang tidak mengandung fitur i. Selisih $[f_x(S \cup \{i\}) - f_x(S)]$mengukur deviasi marginal dari penambahan fitur i ke dalam subset evaluasi. Meskipun nilai Shapley memberikan jaminan aksiomatik yang adil, perhitungan eksak secara kombinatorial memerlukan beban komputasi yang sangat masif $2^{|F|}$. Oleh karena

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

itu, penelitian ini secara spesifik mengimplementasikan pendekatan SHAP *Permutation Explainer*. Justifikasi ilmiah pemilihan *Permutation Explainer* dibandingkan jenis pendekatan SHAP spesifik seperti *TreeExplainer* yang eksklusif untuk arsitektur pohon XGBoost atau *LinearExplainer* yang eksklusif untuk Logistic Regression didasarkan pada sifatnya yang sepenuhnya model-agnostic [26].

Dengan karakteristik *model-agnostic*, *Permutation Explainer* mengukur kontribusi fitur murni dengan melakukan permutasi nilai fitur pada data uji secara berulang tanpa bergantung pada struktur internal algoritma dasar model. Pendekatan ini menyajikan satu lensa standar penilaian yang adil, setara, dan seragam ketika membandingkan logika penalaran antara arsitektur statistik linear (*glass-box*) pada *Logistic Regression* dan arsitektur non-linear kompleks (*black-box*) pada *XGBoost*.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan analisis mendalam mengenai performa model *XGBoost* dan *Logistic Regression* dalam mendeteksi hipertensi, serta interpretasi logikanya menggunakan *SHAP Permutation Explainer*. Fokus utama analisis ini adalah memvalidasi hasil prediksi pada distribusi data asli tanpa penggunaan teknik *oversampling* untuk menjaga orisinalitas fitur klinis.

3.1 Hasil Pelatihan dan Validasi (Cross Validation)

Langkah pertama dalam pengujian ini adalah menilai stabilitas model menggunakan *Stratified K-Fold 10-fold Cross Validation*. Evaluasi performa model pada data uji independen yang tidak dilibatkan dalam proses pelatihan dapat dilihat secara rinci pada Tabel 2 berikut.

Tabel 2. Hasil Cross Validation Train dan Validasi

Model	Rata-Rata Fold	Accuracy	Precision	Recall	F1	AUC
Logistic Regression	Train	0,724	0,697	0,631	0,662	0,795
	Validasi	0,723	0,697	0,630	0,662	0,795
XGBoost	Train	0,746	0,713	0,683	0,698	0,824
	Validasi	0,729	0,693	0,663	0,677	0,802

Berdasarkan hasil pada Tabel 2, kedua model menunjukkan performa yang konsisten antara data pelatihan dan validasi. XGBoost menunjukkan skor metrik yang secara konsisten lebih tinggi dibandingkan *Logistic Regression*. Hal yang paling krusial adalah kecilnya selisih (*gap*) antara akurasi *train* dan *validation* yang mengindikasikan bahwa kedua model memiliki kemampuan generalisasi yang baik dan terbebas dari masalah *overfitting*. Penggunaan data asli (tanpa *oversampling*) terbukti mampu menghasilkan model yang stabil meskipun terdapat ketidakseimbangan kelas pada dataset.

3.2 Hasil Pengujian Akhir (Test Set)

Setelah fase Pelatihan dan validasi, model diuji menggunakan 20% data independen yang tidak terlibat dalam proses pelatihan. Tahap ini krusial untuk mengukur kemampuan generalisasi model terhadap data baru yang mencerminkan kondisi riil di lapangan. Ringkasan hasil pengujian akhir dari kedua model disajikan pada Tabel 3 berikut.

Tabel 3. Hasil Test Penelitan

Model	Accuracy	Precision	Recall	F1	AUC
Logistic Regression	0,731	0,706	0,638	0,670	0,802
XGBoost	0,734	0,699	0,668	0,683	0,805

Berdasarkan data yang tercantum pada Tabel 3, model XGBoost secara konsisten mengungguli *Logistic Regression* pada seluruh parameter evaluasi. Nilai AUC sebesar 0,805 menunjukkan bahwa XGBoost memiliki diskriminasi yang "sangat baik" dalam memisahkan kelas penderita hipertensi dan non-hipertensi. Keunggulan ini disebabkan oleh mekanisme *gradient boosting* yang mampu menangkap hubungan non-linear dan interaksi kompleks antar variabel Kesehatan seperti kaitan antara usia, BMI, dan riwayat penyakit kronis yang seringkali tidak mampu dipetakan secara sempurna oleh model linear seperti *Logistic Regression*.

Hal yang menarik untuk dicatat adalah nilai *Recall* pada kedua model berada di kisaran 0,638 hingga 0,668. Fenomena ini merupakan konsekuensi logis dari keputusan penelitian untuk tidak menggunakan teknik oversampling (seperti SMOTE). Meskipun nilai *Recall* tidak setinggi pada penelitian yang memanipulasi data sintetis, hasil ini memberikan potret kemampuan deteksi yang lebih jujur dan objektif. Dengan mempertahankan distribusi asli dataset BRFSS, model terhindar dari risiko bias informasi klinis sehingga prediksi yang dihasilkan tetap reliabel untuk digunakan sebagai instrumen pendukung keputusan bagi tenaga medis dalam melakukan skrining awal hipertensi.

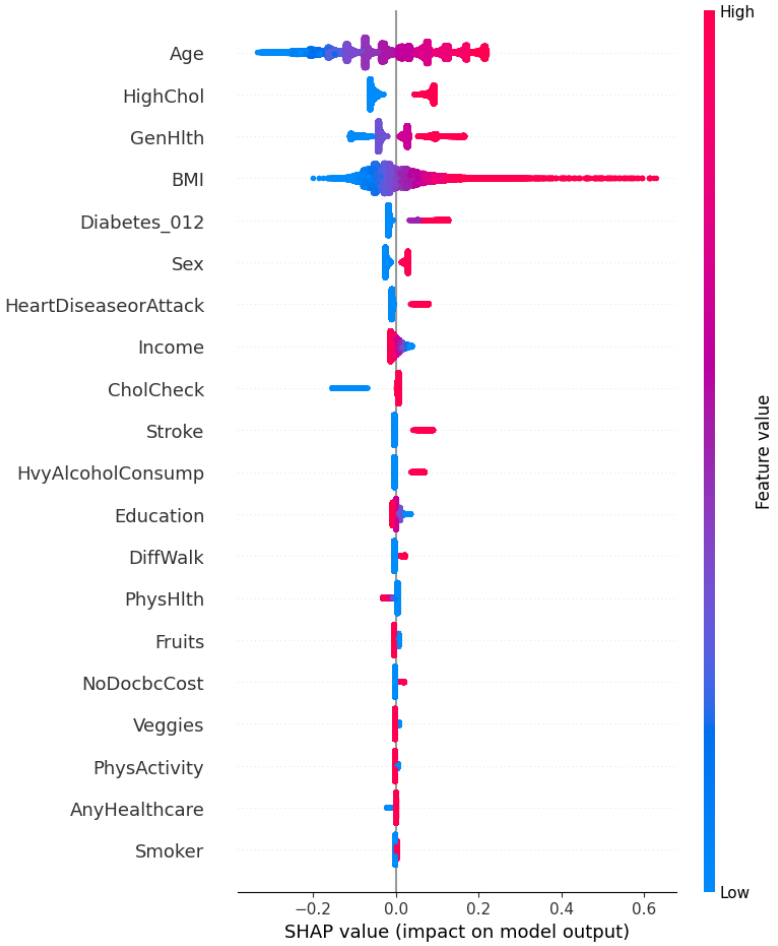
3.3 Analisis Interpretasi Model dengan SHAP Permutation Explainer

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Setelah mengevaluasi performa prediktif model, tahapan krusial berikutnya adalah membedah mekanisme pengambilan keputusan model guna memastikan transparansi dan akuntabilitas klinis. Pada bagian ini, digunakan SHAP (*Shapley Additive Explanations*) dengan metode *Permutation Explainer* untuk memberikan interpretasi yang bersifat *model-agnostic*. Pendekatan ini memungkinkan perbandingan yang adil antara struktur non-linear pada XGBoost dan struktur linear pada *Logistic Regression* melalui atribusi nilai kontribusi fitur yang konsisten. Penjelasan ini disajikan dalam dua perspektif utama yaitu signifikansi fitur secara menyeluruh (global) dan dinamika variabel pada tingkat pasien secara spesifik (lokal).

3.3.1 Analisis Global (Summary Plot)

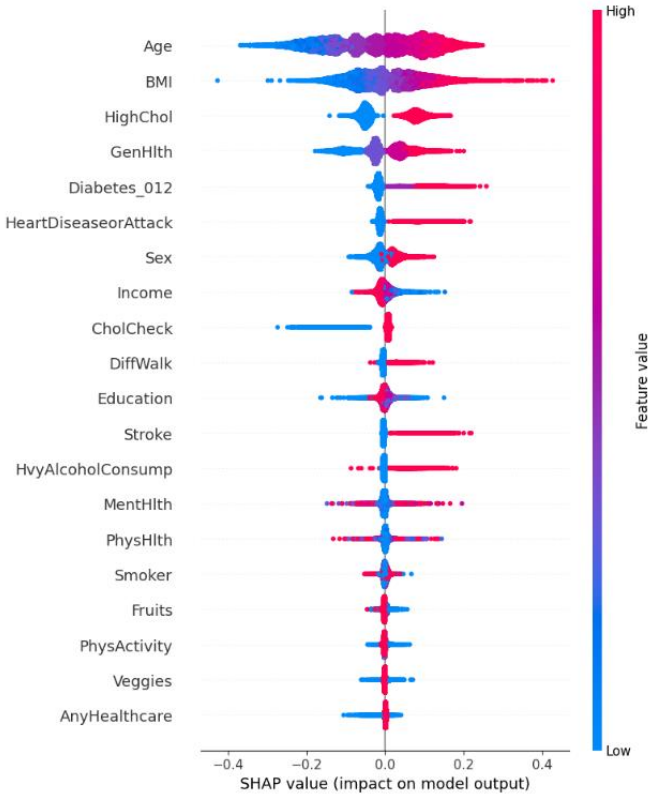
Analisis global memberikan gambaran mengenai fitur-variabel mana yang paling mendominasi keputusan model secara keseluruhan pada dataset BRFSS 2015. Dengan menggunakan *SHAP Summary Plot*, tingkat kepentingan fitur diurutkan berdasarkan nilai absolut rata-rata SHAP. Pendekatan ini memungkinkan peneliti untuk memverifikasi apakah model telah mempelajari pola klinis yang benar atau terjebak pada *noise* dataset. Visualisasi hasil analisis ini disajikan pada Gambar 2 untuk model Logistic Regression dan Gambar 3 untuk model XGBoost.



Gambar 2. Summary Plot Logistic Regression

Interpretasi Summary Plot pada Gambar 2 memperlihatkan pola pengaruh yang linear dan simetris, di mana BMI dan Usia menjadi prediktor utama (positif) terhadap peningkatan risiko hipertensi. Titik-titik merah yang merepresentasikan nilai fitur tinggi seperti kadar kolesterol atau diabetes secara konsisten bergeser ke arah kanan sumbu nol menunjukkan kontribusi yang proporsional terhadap risiko. Hasil ini membuktikan bahwa meskipun bersifat linear, Logistic Regression mampu mempertahankan logika medis yang stabil dan transparan. Untuk Interpertasi Summary Plot XGBoost terletak pada gambar 3 dibawah.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 3. Summary Plot XGBoost

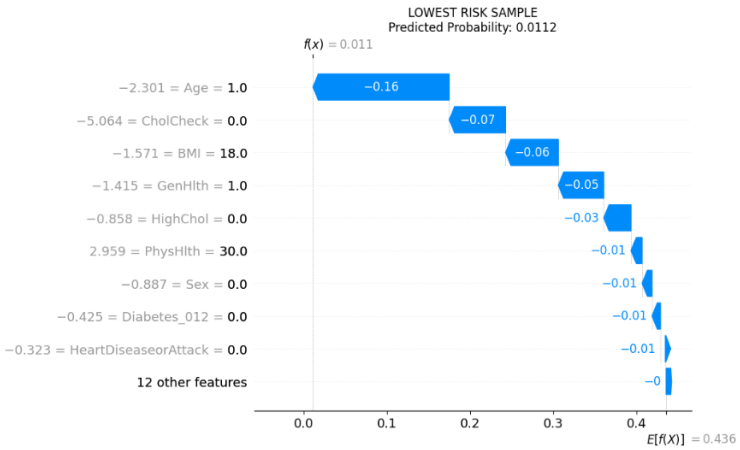
Berbeda dengan model sebelumnya, Gambar 3 menunjukkan kemampuan XGBoost dalam menangkap hubungan non-linear yang lebih kompleks melalui sebaran titik yang lebih dinamis. Meskipun BMI dan Usia tetap menjadi prediktor paling dominan, sebaran nilai yang lebih luas pada plot ini mencerminkan adanya interaksi antar fitur seperti kombinasi usia lanjut dan indeks massa tubuh yang memberikan beban kontribusi unik bagi setiap individu. Hal ini mengonfirmasi efektivitas XGBoost dalam memetakan pola risiko yang mendalam pada distribusi data asli tanpa memerlukan intervensi data sintetis.

3.3.2 Analisis Lokal (Waterfall Plot)

Analisis lokal bertujuan untuk memberikan transparansi pada tingkat individu dengan menguraikan bagaimana setiap fitur klinis berkontribusi terhadap perubahan probabilitas prediksi dari nilai dasar (*base value*). Melalui visualisasi *Waterfall Plot*, tenaga medis dapat memahami alasan spesifik mengapa seorang pasien diklasifikasikan ke dalam kelompok risiko tertentu. Hasil analisis lokal untuk sampel dengan risiko terendah dan tertinggi pada kedua model disajikan melalui Gambar 4 hingga Gambar 7.

a. Probabilitas Terendah

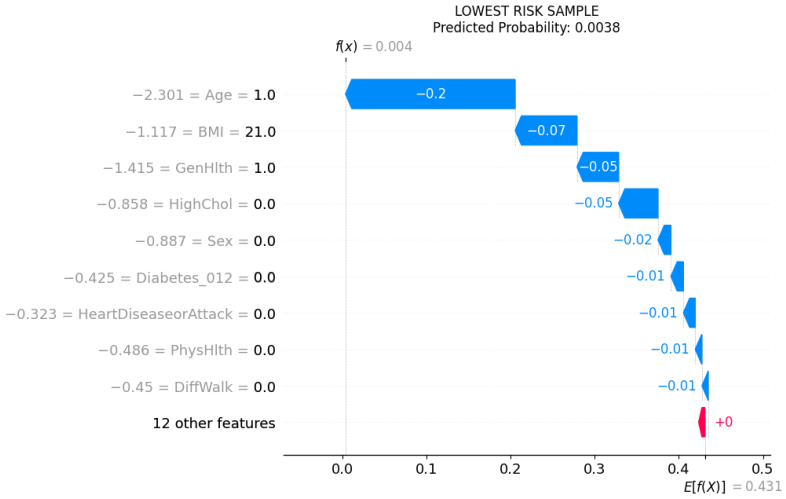
Untuk mengamati bagaimana model membedah kasus dengan profil kesehatan optimal, visualisasi *Waterfall Plot* untuk model Logistic Regression ditampilkan pada Gambar 4 dibawah.



Gambar 4. Waterfall Plot Terendah Logistic Regression

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Untuk memahami mekanisme prediksi pada tingkat individu, Gambar 4 menampilkan *Waterfall Plot* model Logistic Regression bagi sampel dengan risiko terendah. Plot ini mengilustrasikan bagaimana probabilitas akhir sebesar 0,0112 diperoleh dari nilai dasar (*expected value*) sebesar 0,436. Penurunan drastis ini dipicu oleh dominasi pengaruh negatif (panah biru) dari berbagai profil klinis pasien. Kontribusi utama penurunan risiko pada sampel ini berasal dari faktor usia muda (kontribusi -0,16), nilai BMI ideal 18.0 (-0,06), serta persepsi kesehatan umum yang optimal (-0,05). Kombinasi faktor tersebut, ditambah dengan ketiadaan riwayat kolesterol dan penyakit kronis lainnya secara akumulatif menggeser probabilitas prediksi mendekati angka nol. Hal ini membuktikan konsistensi model dalam memetakan indikator kesehatan optimal terhadap penurunan risiko yang secara spesifik berbeda dengan dinamika interpretasi pada model XGBoost yang disajikan pada Gambar 5 dibawah.

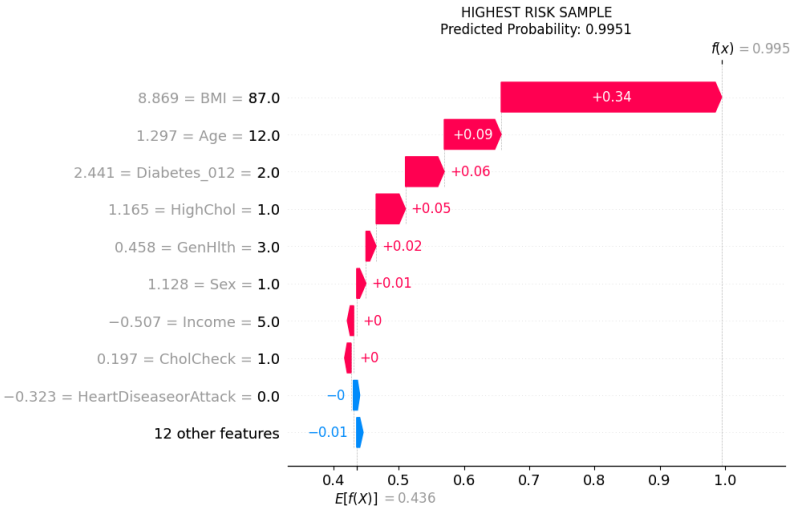


Gambar 5. Waterfall Plot Terendah XGBoost

Untuk membedah mekanisme pengambilan keputusan pada model XGBoost, Gambar 5 menyajikan *Waterfall Plot* bagi sampel individu dengan risiko terendah. Plot ini menunjukkan penurunan probabilitas dari nilai dasar sebesar 0,431 menjadi hanya 0,003, yang divisualisasikan melalui dominasi panah biru yang mengarah ke kiri sebagai indikator pengaruh negatif terhadap risiko. Faktor dominan yang menekan risiko pada sampel ini adalah usia muda dengan kontribusi -0,2. Selain itu, profil kesehatan optimal seperti BMI 21.0 (-0,07), status kesehatan umum prima, serta ketiadaan riwayat kolesterol tinggi, diabetes, dan hambatan fisik (masing-masing -0,05) secara kolektif menekan probabilitas risiko ke level yang sangat rendah. Analisis ini mengonfirmasi kemampuan XGBoost dalam mengenali kombinasi faktor pelindung klinis secara akurat dalam menetapkan prediksi individu.

b. Probabilitas Tertinggi

Untuk membandingkan kinerja model dalam mendeteksi profil risiko tertinggi, *Waterfall Plot* bagi model Logistic Regression dan XGBoost masing-masing disajikan pada Gambar 6 dan Gambar 7.

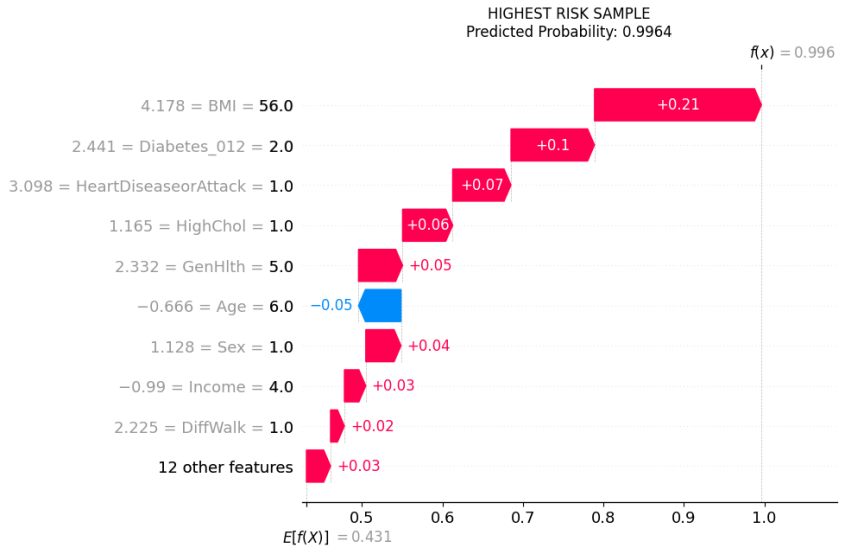


Gambar 6. Waterfall Plot Tertinggi Logistic Regression

Interpretasi lokal melalui *Waterfall Plot* pada Gambar 6 menunjukkan lonjakan probabilitas dari 0,436 menjadi 0,9951 pada model Logistic Regression. Peningkatan drastis ini divisualisasikan melalui dominasi batang merah yang mengarah ke kanan yang menandakan akumulasi faktor risiko. Pemicu utama lonjakan risiko adalah nilai BMI ekstrem

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9748
Hal 1054-1063
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

sebesar 87,0 (+0,34), diikuti oleh faktor usia (+0,09), riwayat diabetes (+0,06), serta buruknya persepsi kesehatan umum dan kadar kolesterol. Secara kolektif, akumulasi variabel klinis tersebut mengunci prediksi model pada tingkat risiko yang hampir mutlak yang kemudian akan dibandingkan dengan performa XGBoost pada Gambar 7 dibawah.



Gambar 7. Waterfall Plot Tertinggi XGBoost

Untuk mengevaluasi performa model pada profil risiko tertinggi, Gambar 7 diataas menyajikan *Waterfall Plot* bagi model XGBoost. Plot ini mengilustrasikan lonjakan probabilitas dari nilai dasar 0,431 menjadi 0,9964 yang didominasi oleh akumulasi faktor risiko klinis. Pemicu utama lonjakan tersebut adalah nilai BMI ekstrem sebesar 56.0 (+0,21), diikuti oleh riwayat diabetes (+0,1), serta penyakit jantung (+0,07). Meskipun variabel usia memberikan efek penurun kecil, kombinasi berbagai indikator kesehatan yang buruk secara kolektif mendorong prediksi risiko hingga mendekati angka mutlak.

3.4 Analisis Perbandingan Hasil dengan Studi Terkait

Temuan dalam penelitian ini memiliki pola performa yang sejalan dengan studi oleh Bisong dkk. [9] yang juga melaporkan bahwa algoritma *ensemble learning* seperti XGBoost memberikan keunggulan diskriminatif dibandingkan metode linear tradisional dalam memprediksi risiko penyakit tidak menular. Namun, jika dibandingkan dengan AlKaabi dkk. [5] yang menggunakan dataset Qatar Biobank, model penelitian ini menunjukkan nilai AUC yang kompetitif (0,805) meskipun tanpa bantuan teknik *oversampling* atau manipulasi data sintetis. Perbedaan performa ini kemungkinan besar disebabkan oleh karakteristik distribusi data BRFSS 2015 yang digunakan dalam penelitian ini yang mencerminkan variabilitas populasi yang lebih luas dibandingkan dataset berbasis biobank yang cenderung lebih homogen. Selain itu, konsistensi model dalam menangkap fitur klinis utama seperti usia dan BMI menegaskan bahwa pemilihan fitur yang tepat pada data asli seringkali lebih krusial dibandingkan upaya artifisial untuk meningkatkan metrik sensitivitas melalui data sintetis, sebagaimana juga disinggung dalam analisis oleh Montagna dkk. [7] mengenai perlunya metode deteksi yang tetap menjaga objektivitas klinis di lapangan.

4. KESIMPULAN

Penelitian ini berhasil menyimpulkan bahwa deteksi dini risiko hipertensi menggunakan data asli tanpa intervensi *oversampling* mampu menghasilkan performa klasifikasi yang reliabel dengan nilai AUC konsisten di atas 0,80 pada kedua model yang diuji. Secara metrik, model non-linear XGBoost menunjukkan keunggulan tipis dengan nilai AUC sebesar 0,805 jika dibandingkan dengan model linear Logistic Regression yang memperoleh nilai AUC 0,802. Keunggulan ini membuktikan bahwa mekanisme *gradient boosting* efektif dalam menangkap interaksi fitur yang kompleks serta hubungan non-linear antar variabel kesehatan, seperti korelasi mendalam antara usia, BMI, dan riwayat penyakit kronis pasien. Di sisi lain, penerapan *Explainable AI* (XAI) melalui metode SHAP *Permutation Explainer* berhasil membongkar aspek "kotak hitam" (*black box*) model dan mengungkapkan konsistensi logika yang kuat di antara kedua arsitektur tersebut. Melalui analisis global *Summary Plot*, variabel BMI dan Usia secara konsisten terbukti menjadi prediktor paling dominan dalam mendorong probabilitas risiko hipertensi, baik pada pendekatan linear maupun non-linear. Dengan demikian, penggunaan data asli dalam penelitian ini terbukti memberikan hasil prediksi yang lebih jujur, objektif, dan memiliki reliabilitas klinis yang lebih tinggi bagi praktisi medis untuk kebutuhan instrumen skrining awal, meskipun nilai *Recall* yang dihasilkan masih berada pada kisaran moderat. Secara keseluruhan, kerangka kerja pemodelan yang transparan ini berhasil menjawab tantangan akuntabilitas implementasi kecerdasan buatan di sektor kesehatan. Meskipun memberikan kontribusi yang kuat, penelitian ini memiliki keterbatasan tertentu yang perlu diperhatikan untuk pengembangan di masa depan. Salah satu batasan utama adalah ketergantungan pada dataset sekunder BRFSS yang

bersifat survei mandiri, sehingga potensi bias pelaporan dari responden tetap ada. Selain itu, kompleksitas komputasi yang tinggi saat menggunakan metode SHAP pada dataset berskala besar dapat menjadi hambatan jika diterapkan pada sistem *real-time*. Penelitian selanjutnya disarankan untuk mengintegrasikan dataset dari populasi yang lebih beragam atau data klinis *real-time* dari fasilitas kesehatan untuk meningkatkan generalisasi model. Selain itu, eksplorasi terhadap teknik optimasi *hyperparameter* yang lebih luas atau penggabungan beberapa metode penjelasan (XAI) lainnya dapat dilakukan guna memperkuat akuntabilitas dan kepercayaan praktl;isi medis terhadap sistem pendukung keputusan berbasis kecerdasan buatan ini.

REFERENCES

[1] M. Puji Lestari and A. M. Rusydan, "Penyuluhan Hipertensi dan DAGUSIBU Obat Antihipertensi," *J. Innov. Community Empower.*, vol. 6, no. 2, pp. 79–85, Sep. 2024, doi: 10.30989/jice.v6i2.1400.
[2] R. M. Touyz, "*Hypertension* 2022 Update: Focusing on the Future," *Hypertension*, vol. 79, no. 8, pp. 1559–1562, Aug. 2022, doi: 10.1161/HYPERTENSIONAHA.122.19564.
[3] Y. A. Mashuri, N. Ng, and A. Santosa, "Socioeconomic disparities in the burden of hypertension among Indonesian adults - a multilevel analysis," *Global Health Action*, vol. 15, no. 1, Dec. 2022, doi: 10.1080/16549716.2022.2129131.
[4] E. Martinez-Ríos, L. Montesinos, M. Alfaro-Ponce, and L. Pecchia, "A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data," *Biomedical Signal Processing and Control*, vol. 68, p. 102813, Jul. 2021, doi: 10.1016/j.bspc.2021.102813.
[5] L. A. AlKaabi, L. S. Ahmed, M. F. Al Attiyah, and M. E. Abdel-Rahman, "Predicting hypertension using machine learning: Findings from Qatar Biobank Study," *PLoS ONE*, vol. 15, no. 10, p. e0240370, Oct. 2020, doi: 10.1371/journal.pone.0240370.
[6] N. Chen *et al.*, "Evaluating the risk of hypertension in residents in primary care in Shanghai, China with machine learning algorithms," *Front. Public Health*, vol. 10, p. 984621, Oct. 2022, doi: 10.3389/fpubh.2022.984621.
[7] S. Montagna *et al.*, "Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data," *J Med Syst*, vol. 47, no. 1, Dec. 2022, doi: 10.1007/s10916-022-01900-5.
[8] F. Si, Q. Liu, and J. Yu, "A prediction study on the occurrence risk of heart disease in older hypertensive patients based on machine learning," *BMC Geriatr*, vol. 25, no. 1, p. 27, Jan. 2025, doi: 10.1186/s12877-025-05679-1.
[9] E. Bisong, N. Jibril, P. Premnath, E. Buligwa, G. Oboh, and A. Chukwuma, "Predicting high blood pressure using machine learning models in low- and middle-income countries," *BMC Med Inform Decis Mak*, vol. 24, no. 1, p. 234, Aug. 2024, doi: 10.1186/s12911-024-02634-9.
[10] S. Widodo, H. Brawijaya, and S. Samudi, "Stratified K-fold cross validation optimization on machine learning for prediction," *SinkrOn*, vol. 7, no. 4, pp. 2407–2414, Oct. 2022, doi: 10.33395/sinkron.v7i4.11792.
[11] T. Abedin, H. Xu, and S. Uddin, "The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models," *Sci Rep*, vol. 16, no. 1, p. 6084, Jan. 2026, doi: 10.1038/s41598-026-37247-x.
[12] U. Mitra, P. Sarkar, J. Mondal, and J. Kundu, "Enhancing Interpretability in Diabetics Prediction: A Comparative Study of SHAP, LIME and Permutation Feature Importance," in *2025 AI-Driven Smart Healthcare for Society 5.0*, Feb. 2025, pp. 1–6. doi: 10.1109/IEEECONF64992.2025.10962890.
[13] G. Di Franco and M. Santurro, "Machine learning, artificial neural networks and social research," *Qual Quant*, vol. 55, no. 3, pp. 1007–1025, Jun. 2021, doi: 10.1007/s11135-020-01037-y.
[14] F. Emmert-Streib and M. Dehmer, "Taxonomy of machine learning paradigms: A data-centric perspective," *WIREs Data Min & Knowl*, vol. 12, no. 5, p. e1470, Sep. 2022, doi: 10.1002/widm.1470.
[15] J. Kumari, E. Kumar, and D. Kumar, "A Structured Analysis to study the Role of Machine Learning and Deep Learning in The Healthcare Sector with Big Data Analytics," *Arch Computat Methods Eng*, vol. 30, no. 6, pp. 3673–3701, Jul. 2023, doi: 10.1007/s11831-023-09915-y.
[16] D. W. H. Jr, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. John Wiley & Sons, 2013.
[17] P. Trevisan *et al.*, "Sparse Logistic Regression for RR Lyrae versus Binaries Classification," *ApJ*, vol. 950, no. 2, p. 103, Jun. 2023, doi: 10.3847/1538-4357/accf8f.
[18] J. F. Vera, "Distance-based logistic model for cross-classified categorical data," *Brit J Math & Statis*, vol. 75, no. 3, pp. 466–492, Nov. 2022, doi: 10.1111/bmsp.12264.
[19] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate data analysis*, Eighth edition. Australia Brazil Mexico South Africa Singapore United Kingdom United States: Cengage, 2019.
[20] S. Luo and K. Shaukat, Eds., *Computational Methods for Medical and Cyber Security*. MDPI, 2022. doi: 10.3390/books978-3-0365-5115-9.
[21] R. E. Schapire and Y. Freund, "Boosting: Foundations and Algorithms".
[22] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
[23] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, Agustus 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[24] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif Intell Rev*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.
[25] A. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," 2023, doi: 10.48550/ARXIV.2305.02012.
[26] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A Unified Framework for Machine Learning Interpretability," Sep. 19, 2019, *arXiv*: arXiv:1909.09223. doi: 10.48550/arXiv.1909.09223.