



Optimasi *XGBoost* Dengan *SHAP* Untuk Sistem Skrining Penyakit Jantung

Clara Zuliani Syahputri^{1*}, Jasmir², Fachruddin³

¹⁻³ Magister Sistem Informasi, Ilmu Komputer, Universitas Dinamika Jambi, Indonesia

Email: Clarabisnis89@gmail.com^{1*}, jasmir@unama.ac.id², fachruddin.stikom@gmail.com³

Alamat: Jl. Jend. Sudirman, The Hok, Kec. Jambi Sel., Kota Jambi, Jambi 36138

*Penulis Korespondensi: Clarabisnis89@gmail.com

Abstract. Heart disease is the leading cause of death in Indonesia and globally, necessitating an early screening system that is both accurate and clinically trustworthy. Although *XGBoost* demonstrates high predictive performance, its black-box nature undermines clinical trust, while low recall risks missed diagnosis an unacceptable consequence in population screening, especially in middle-income countries with limited healthcare resources. This study aims to develop a sensitive, transparent, and implementation-ready heart disease screening framework through the integration of *SHAP*-based Explainable AI. The CDC's Indicators of Heart Disease dataset (319,795 samples) was processed according to WHO/CDC standards, followed by class imbalance handling, hyperparameter optimization using *RandomizedSearchCV*, evaluation based on metrics sensitive to minority classes (AUC, recall, F1-score, AUC-PR), and threshold tuning to maximize recall. The baseline model showed a very low recall of 12.18%. After optimization and threshold tuning at 0.10, the model achieved recall >96% (96.79%) with a G-mean of 0.7477, supported by *SHAP* interpretation stability and the ability to capture non-linear interactions between advanced age (*AgeCategory_WHO*) and poor general health (*GenHealth*). *SHAP* analysis confirmed the alignment of dominant features with medical evidence, and its visualizations provide transparent explanations for healthcare professionals indicating its potential implementation as an interpretable clinical decision support system.

Keywords: Clinical Decision Support System; Explainable AI; Heart Disease; *SHAP*; *XGBoost*.

Abstrak. Penyakit jantung merupakan penyebab utama kematian di Indonesia dan global, sehingga diperlukan sistem skrining dini yang akurat dan dapat dipercaya secara klinis. Meskipun *XGBoost* memiliki performa prediktif tinggi, sifat *black-box*-nya mengurangi kepercayaan klinis, sementara *recall* yang rendah berisiko menyebabkan *missed* diagnosis sebuah konsekuensi yang tidak dapat diterima dalam skrining populasi, khususnya di negara berpendapatan menengah dengan sumber daya kesehatan yang terbatas. Penelitian ini bertujuan mengembangkan sebuah kerangka kerja skrining penyakit jantung yang sensitif, transparan, dan siap diimplementasikan melalui integrasi *Explainable AI* berbasis *SHAP*. *Dataset Indicators of Heart Disease* dari CDC (319.795 sampel) diproses sesuai standar, diikuti penanganan ketidakseimbangan kelas, optimasi hiperparameter menggunakan *RandomizedSearchCV*, evaluasi berbasis metrik sensitif terhadap kelas minoritas (AUC, *recall*, F1-score, AUC-PR), dan *threshold tuning* untuk memaksimalkan *recall*. Model *baseline* menunjukkan *recall* sangat rendah sebesar 12,18%. Setelah optimasi dan *threshold tuning* pada 0,10, model mencapai *recall* >96% (96,79%) dengan *G-mean* 0,7477, didukung stabilitas interpretasi *SHAP* dan kemampuan menangkap interaksi non-linear antara usia lanjut (*AgeCategory_WHO*) dan kondisi kesehatan umum buruk (*GenHealth*). Analisis *SHAP* mengonfirmasi keselarasan fitur dominan dengan bukti medis, serta visualisasinya memberikan penjelasan transparan untuk tenaga kesehatan mengindikasikan potensi implementasinya sebagai sistem pendukung keputusan klinis yang dapat diinterpretasi.

Kata kunci: Explainable AI; Penyakit Jantung; Sistem Pendukung Keputusan Klinis; *SHAP*; *XGBoost*.

1. LATAR BELAKANG

Penyakit jantung, khususnya Penyakit Jantung Iskemik (PJI), tetap menjadi penyebab utama kematian di Indonesia dan global, menyumbang 13% dari seluruh kematian di dunia (WHO, 2024). Di Indonesia, prevalensi penyakit jantung meningkat tiga kali lipat dari 0,5% pada 2013 menjadi 1,5% pada 2018, sekaligus menempati beban pembiayaan terbesar

pada BPJS Kesehatan (Rokom, 2022). Fenomena ini menciptakan kebutuhan mendesak akan sistem skrining dini yang akurat, dapat dipercaya, dan dapat diterapkan secara luas, terutama di negara berpendapatan menengah dengan sumber daya kesehatan yang terbatas seperti Indonesia.

Dalam upaya mengembangkan alat prediksi yang kuat, pembelajaran mesin (*machine learning*) telah banyak diterapkan (Houssein et al., 2023). Di antara berbagai algoritma, *XGBoost* (*eXtreme Gradient Boosting*) sering menjadi pilihan utama karena kemampuan prediktifnya yang tinggi, ketahanan terhadap *overfitting*, dan efisiensinya dalam menangani data skala besar serta hubungan non-linear antar fitur. Kinerja superior *XGBoost* telah dibuktikan dalam berbagai prediksi risiko kardiovaskular, termasuk prediksi mortalitas gagal jantung (Li et al., 2022). dan kejadian jantung baru (Guan et al., 2024). Namun, penerapan *XGBoost* dalam konteks skrining populasi menghadapi dua tantangan kritis.

Pertama, sifat "*black-box*" atau sulit ditafsirkan dari model *XGBoost* membatasi kepercayaan dan adopsi klinis (Wang et al., 2021). Klinisi memerlukan tidak hanya prediksi yang akurat, tetapi juga penjelasan yang transparan dan dapat dipertanggungjawabkan secara medis untuk setiap keputusan. Kedua, meskipun memiliki akurasi keseluruhan yang tinggi, model *XGBoost* yang tidak dioptimalkan untuk tugas skrining sering kali menghasilkan *recall* (sensitivitas) yang rendah (Rahman et al., 2023). Dalam skrining populasi, *recall* yang rendah sangat berbahaya karena berisiko menyebabkan *missed diagnosis* (*false negatives*), di mana individu berisiko tinggi terlewatkan dan tidak mendapatkan intervensi dini sebuah konsekuensi yang tidak dapat diterima dalam kesehatan masyarakat.

Untuk mengatasi masalah interpretabilitas, metode *Explainable AI* (*XAI*) seperti *SHAP* (*SHapley Additive exPlanations*) telah dikembangkan (Chaddad et al., 2023). *SHAP* memberikan kerangka kerja teoretis yang kuat untuk mengukur kontribusi setiap fitur terhadap prediksi individual, sehingga membuat logika model menjadi lebih transparan. Studi-studi terdahulu telah berhasil mengintegrasikan *SHAP* dengan *XGBoost*, misalnya untuk mengidentifikasi faktor dominan seperti usia dan fungsi ginjal pada prediksi mortalitas pasien jantung, yang selaras dengan pengetahuan medis (Li et al., 2022). Namun, sebagian besar penelitian tersebut masih berfokus pada pengoptimalan metrik seperti *AUC* atau akurasi keseluruhan, dan belum secara khusus dirancang untuk kebutuhan skrining yang memprioritaskan *recall* maksimal.

Terdapat kesenjangan (*gap*) yang jelas antara kemampuan prediktif *XGBoost* dan kebutuhan operasional skrining populasi yang sesungguhnya. Literatur menunjukkan bahwa model dengan akurasi tinggi sekalipun dapat gagal mendeteksi kasus (*recall* sangat rendah) jika ketidakseimbangan kelas dan *threshold* keputusan default tidak dikelola dengan tepat (Marcolino et al., 2025). Studi-studi yang ada umumnya menggunakan teknik penyeimbangan data seperti *SMOTE*, yang meski meningkatkan *recall*, sering kali belum mencapai tingkat sensitivitas yang memadai untuk skrining massal (>95%) dan dapat mengorbankan presisi atau mengintroduksi pola sintesis yang tidak realistis dalam data, (Reátegui et al., 2025). Di sisi lain, pendekatan *threshold tuning* penyesuaian batas keputusan klasifikasi merupakan metode yang langsung dan efektif untuk menggeser prioritas model ke arah deteksi kasus yang lebih banyak (*tinggi recall*), namun belum banyak dieksplorasi secara sistematis dalam konteks kombinasi dengan *SHAP* untuk skrining penyakit jantung (Sergouni et al., 2025).

Urgensi penelitian ini terletak pada kebutuhan akan solusi yang tidak hanya mengandalkan kinerja prediktif *XGBoost* tetapi juga menjamin kepekaan (*high recall*) dan kejelasan (*transparency*) dalam skenario skrining dunia nyata, khususnya di Indonesia. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengusulkan sebuah kerangka kerja skrining penyakit jantung yang mengintegrasikan optimasi *hiperparameter XGBoost*, *threshold tuning* berorientasi *recall tinggi*, dan interpretabilitas berbasis *SHAP*. Tujuannya adalah menciptakan sebuah sistem yang sensitif (dengan *recall* >95%), transparan (dapat dijelaskan kepada klinisi), dan siap diimplementasikan sebagai modul pendukung keputusan awal dalam sistem kesehatan di lingkungan sumber daya terbatas. Dengan demikian, penelitian ini diharapkan dapat menjembatani kesenjangan antara kecanggihan algoritma *machine learning* dan kebutuhan praktis serta etis dalam pelayanan kesehatan preventif.

2. KAJIAN TEORITIS

XGBoost adalah algoritma *supervised learning* berbasis *gradient boosting* yang membangun serangkaian pohon keputusan secara iteratif untuk mengoreksi kesalahan prediksi model sebelumnya (Chaddad et al., 2023). Secara matematis, setiap iterasi menambahkan pohon baru $f_N(x_i)$ untuk meminimalkan fungsi objektif yang terdiri dari komponen loss dan regularisasi (tang, n.d.). Keunggulan *XGBoost* terletak pada efisiensi komputasi, penanganan *missing value* secara otomatis, serta adanya regularisasi L1 dan L2 yang mencegah overfitting (Harrison, n.d.). Dalam konteks prediksi risiko klinis, kemampuan *XGBoost* untuk menangkap

hubungan non-linear dan interaksi kompleks antar variabel menjadikannya pilihan yang sangat efektif (Montone et al., 2024; Oikonomou & Khera, 2023).

Meskipun demikian, kekuatan prediktif *XGBoost* sering kali diikuti oleh kompleksitas interpretasi yang menjadikannya model *black-box* (Maharani et al., 2024). Keterbatasan ini menghambat adopsi klinis, karena tenaga kesehatan memerlukan penjelasan yang transparan dan dapat dipertanggungjawabkan untuk setiap keputusan yang didukung *AI* (Khamparia & Gupta, 2025). Untuk mengatasi hal ini, *XAI* hadir sebagai pendekatan untuk membuat sistem *AI* dapat dipahami oleh manusia (*Explainable Artificial Intelligence for Biomedical and Healthcare Applications*, 2025). Di antara berbagai metode *XAI*, *SHAP* telah muncul sebagai kerangka kerja yang kuat dan berdasar teori permainan (*game theory*) untuk memberikan interpretasi yang konsisten dan lokal akurat (Mahalle & Ingle, 2024). *SHAP* menghitung kontribusi marginal setiap fitur terhadap suatu prediksi dengan menggunakan nilai *Shapley*, sehingga memungkinkan analisis pentingnya fitur (*feature importance*) baik secara global maupun lokal untuk kasus individu (Klosterman, 2021; Littlewood et al., 2024).

Tinjauan terhadap penelitian sebelumnya menunjukkan konvergensi yang kuat dalam penerapan *XGBoost* dan *SHAP* untuk tugas-tugas kardiovaskular. Sebagai contoh, (Guan et al., 2024) berhasil mengembangkan model *XGBoost* yang mencapai AUC 0.891 untuk memprediksi *new-onset atrial fibrillation (NOAF)* di *ICU*, dengan interpretasi *SHAP* yang mengidentifikasi usia dan sepsis sebagai faktor risiko utama. Demikian pula, (Luo et al., 2024) mengusulkan kerangka kerja kolaborasi manusia-mesin di mana fitur pilihan ahli digabungkan dengan seleksi fitur berbasis *XGBoost* dan *Random Forest*, kemudian diinterpretasikan menggunakan *SHAP* untuk memprediksi rawat ulang pasien gagal jantung. Hasilnya, model berbasis *CatBoost* yang menggabungkan kedua pendekatan tersebut mencapai kinerja terbaik (AUC = 0.836). Studi oleh (Ye et al., 2023) juga mengonfirmasi keunggulan model (*GBDT*) yang secara konsep mirip dengan *XGBoost* dalam memprediksi mortalitas pasien dengan penyakit ginjal kronis dan penyakit jantung koroner di *ICU* (AUC 0.946), dengan *SHAP* mengungkap usia dan saturasi oksigen minimum sebagai prediktor kunci. Meskipun bukti empiris yang kuat, sebagian besar penelitian terdahulu berfokus pada pengoptimalan metrik seperti AUC atau akurasi dalam konteks perawatan intensif atau manajemen pasien kronis (Luo et al., 2024; Oikonomou & Khera, 2023; Ye et al., 2023). Kesenjangan yang signifikan muncul ketika model-model ini dipertimbangkan untuk skrining populasi, di mana memaksimalkan sensitivitas (*recall*) menjadi tujuan utama untuk meminimalkan *false negatives* atau kasus yang terlewat (Jones et al., 2022). Studi Siefkes et al. (Siefkes et al., 2024) mengenai skrining

penyakit jantung bawaan kritis, misalnya, mencapai sensitivitas 92.86% setelah optimasi, namun penelitian ini bersifat spesifik pada neonatus dan menggunakan data fisiologis khusus. Dalam konteks skrining penyakit jantung pada populasi dewasa menggunakan data survei seperti *BRFSS*, pendekatan sistematis yang menggabungkan optimasi hiperparameter *XGBoost*, *threshold tuning* berorientasi *recall* tinggi, dan interpretabilitas berbasis *SHAP* belum dieksplorasi secara mendalam. *Threshold tuning* merupakan teknik penyesuaian batas keputusan klasifikasi yang terbukti dapat secara dramatis meningkatkan *recall* (Rahman et al., 2023) namun integrasinya dengan *pipeline XGBoost* dan *SHAP* untuk skrining jantung belum menjadi fokus utama penelitian sebelumnya.

Oleh karena itu, penelitian ini berlandaskan pada proposisi bahwa pengembangan sebuah kerangka kerja terpadu yang dirancang khusus untuk skrining dengan fokus pada pencapaian *recall* >95% melalui *threshold tuning*, diiringi dengan interpretasi klinis yang kokoh dari model *XGBoost* yang telah dioptimalkan menggunakan *SHAP* akan mengisi celah literatur yang ada. Pendekatan ini tidak hanya membangun atas fondasi teoretis dan empiris dari kekuatan prediktif *XGBoost* dan kapabilitas penjelasan *SHAP*, tetapi juga secara langsung menjawab kebutuhan mendesak akan sistem skrining yang sensitif, dapat dipercaya, dan siap diimplementasikan dalam sistem kesehatan dengan sumber daya terbatas seperti di Indonesia.

Tabel 1. Tinjauan Pustaka Penelitian Terkait.

| No | Peneliti (Tahun) | Fokus & Tujuan | Metode | Hasil & Kontribusi |
|----|------------------------|--|--|--|
| 1 | (Guan et al., 2024) | Prediksi <i>new-onset atrial fibrillation (NOAF)</i> di <i>ICU</i> pasien non-bedah. | <i>XGBoost</i> , seleksi fitur, <i>SHAP</i> . | <i>AUC</i> 0.891; <i>SHAP</i> identifikasi usia & sepsis sebagai prediktor utama. Bukti integrasi <i>XGBoost-SHAP</i> di <i>ICU</i> . |
| 2 | (Luo et al., 2024) | Prediksi rawat ulang 1 tahun pasien gagal jantung lansia. | Kolaborasi fitur (ahli + <i>ML</i> : <i>XGBoost</i> , <i>RF</i>), <i>SHAP</i> . | Model <i>CatBoost</i> gabungan terbaik (<i>AUC</i> 0.836). <i>SHAP</i> ungkap interaksi risiko (cth: <i>HGB</i> rendah & <i>HCT</i> tinggi). |
| 3 | (Ye et al., 2023) | Prediksi mortalitas pasien CKD dengan CAD di <i>ICU</i> . | <i>GBDT</i> , seleksi fitur Boruta, <i>SHAP</i> , validasi eksternal. | <i>GBDT</i> <i>AUC</i> 0.946; <i>SHAP</i> tunjukkan usia & <i>SpO₂</i> min sebagai prediktor utama. |
| 4 | (Siefkes et al., 2024) | Skrining penyakit jantung bawaan kritis (<i>CCHD</i>) pada neonatus. | Beberapa <i>ML</i> (<i>XGBoost</i> , <i>RF</i> , dll.), ekstraksi fitur sinyal PPG. | <i>Logistic Regression</i> terbaik, tingkatkan sensitivitas <i>CCHD</i> ke 92.86%. Bukti <i>threshold tuning</i> & fitur tambahan tingkatkan <i>recall</i> . |

Sumber: Penelitian (2026).

Berdasarkan tinjauan pustaka, penelitian ini dirancang untuk menjawab kesenjangan metodologis dalam skrining penyakit jantung berbasis survei populasi *BRFSS*. Berbeda dengan penelitian terdahulu yang berfokus pada *AUC* atau akurasi dalam konteks *ICU* atau manajemen kronis, skrining populasi memprioritaskan optimasi *recall*. Namun, belum ada kerangka kerja terpadu yang secara khusus mengintegrasikan optimasi *XGBoost*, *threshold tuning* untuk *recall* maksimal, dan interpretasi *SHAP* dalam konteks ini. Dengan memanfaatkan dataset *BRFSS*, penelitian ini bertujuan mengembangkan sistem skrining yang tidak hanya akurat, tetapi juga dapat ditindaklanjuti (*actionable*) dan dapat diaudit (*auditable*) oleh tenaga kesehatan.

3. METODE PENELITIAN

Penelitian ini menggunakan desain prediktif berbasis pembelajaran mesin untuk mengembangkan dan mengoptimasi model *XGBoost* guna sistem skrining penyakit jantung. Alur metodologi diawali dengan studi literatur dan identifikasi masalah, dilanjutkan secara berurutan oleh tahapan pengumpulan data, pra-pemrosesan, pemodelan, optimasi, evaluasi, serta interpretasi model dengan pendekatan *Explainable AI (XAI)*.

Data penelitian bersumber dari dataset sekunder “*Indicators of Heart Disease*” CDC yang diakses melalui *Kaggle*. *Dataset* ini mencakup 319.795 sampel dewasa dengan 18 variabel, meliputi faktor demografis (usia, jenis kelamin), perilaku kesehatan (merokok, aktivitas fisik), dan kondisi klinis (diabetes, stroke, kesehatan umum). Variabel target adalah *HeartDisease* yang mencerminkan diagnosis penyakit jantung berdasarkan laporan mandiri responden.

Pra-pemrosesan data dilakukan sesuai panduan klinis *WHO/CDC* pada *dataset* yang tidak mengandung *missing values*. Variabel kategorikal diencode dengan label *encoding*, sementara variabel ordinal seperti BMI dan *AgeCategory* dikategorikan berdasarkan standar klinis. *Dataset* kemudian dibagi menjadi data latih dan uji dengan rasio 80:20 menggunakan *stratified sampling* untuk mempertahankan proporsi kelas minoritas penyakit jantung (8,56%).

Pemodelan menggunakan *XGBoost* dengan *baseline* awal pada parameter *default*. Optimasi dilakukan melalui *hyperparameter tuning* menggunakan *RandomizedSearchCV* dengan validasi *5-fold stratified cross-validation*. Ketidakseimbangan kelas diatasi dengan mengatur parameter *scale_pos_weight* sesuai rasio kelas (*inverse class frequency*). Evaluasi kinerja menggunakan metrik sensitif kelas minoritas: akurasi, presisi, *recall*, *F1-score*, *AUC-ROC*, dan *AUC-PR*.

Berdasarkan prioritas skrining populasi untuk *recall* tinggi, *threshold tuning* dilakukan dengan variasi ambang batas 0,05 hingga 0,50. Ambang optimal dipilih secara klinis guna memaksimalkan *recall* tanpa mengorbankan *precision* secara signifikan.

Interpretasi model menggunakan *SHAP* untuk memberikan transparansi dan relevansi klinis. Analisis mencakup *summary plot* (pentingnya fitur global), *dependence plot* (hubungan fitur dengan prediksi), dan *waterfall plot* (penjelasan per kasus). Interpretasi difokuskan pada keselarasan fitur dominan dari model dengan bukti medis yang telah terstandarisasi.

Seluruh analisis diimplementasikan dengan *Python*, memanfaatkan pustaka *scikit-learn* untuk pra-pemrosesan dan evaluasi, *XGBoost* untuk pemodelan, serta *SHAP* untuk interpretasi. Keterulangan (*reproducibility*) penelitian dijamin melalui penetapan *random state* yang konsisten di seluruh *pipeline* analitik.

4. HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengembangkan sistem skrining penyakit jantung berbasis *XGBoost* yang dioptimalkan untuk deteksi maksimal. Melalui integrasi *hyperparameter tuning*, penanganan ketidakseimbangan kelas, dan *threshold tuning* strategis, model berhasil meningkatkan *recall* dari 12,18% menjadi 96,79%, sehingga hampir semua kasus penyakit jantung dapat terdeteksi pada tahap skrining awal. *Threshold optimal* ditemukan pada 0,10, yang secara sengaja mengutamakan sensitivitas atas presisi sebuah pertimbangan klinis kritis dalam konteks skrining populasi.

Interpretasi berbasis *SHAP* mengkonfirmasi keselarasan model dengan bukti medis, mengidentifikasi usia lanjut dan kondisi kesehatan umum yang buruk sebagai faktor risiko dominan dengan efek interaksi yang sinergis. Analisis *waterfall plot* per kasus memberikan transparansi penuh atas setiap prediksi, sehingga meningkatkan kepercayaan klinis dalam penggunaan sistem ini.

Dengan *recall* >96% dan interpretabilitas penuh, sistem ini siap diimplementasikan sebagai alat bantu skrining awal yang efektif dan dapat dipertanggungjawabkan di fasilitas kesehatan primer.

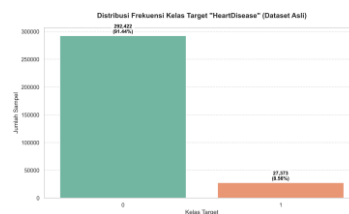
A. Proses Pengumpulan Data

Data penelitian bersumber dari dataset “*Indicators of Heart Disease*” BRFSS 2020 (CDC) yang diakses melalui *Kaggle*, terdiri dari 319.795 sampel dewasa dengan 18 variabel

yang mencakup faktor demografi, perilaku kesehatan, dan kondisi klinis. Variabel target adalah *HeartDisease* yang merepresentasikan diagnosis penyakit jantung berdasarkan laporan mandiri. Meskipun berasal dari populasi AS, variabel-variabel dalam dataset merefleksikan faktor risiko penyakit jantung yang bersifat universal, sehingga dapat menjadi dasar pengembangan model untuk konteks populasi serupa, termasuk Indonesia.

B. Analisis Data

Pra-Pemrosesan Data

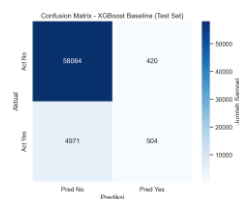


Sumber: Penelitian (2025).

Gambar 1. Distribusi frekuensi kelas target.

Dataset awal memiliki ketidakseimbangan kelas signifikan, dengan hanya 8,56% kasus positif penyakit jantung. Pra-pemrosesan meliputi kategorisasi variabel numerik berdasarkan standar klinis (misalnya *BMI* sesuai klasifikasi *WHO*) dan *encoding* variabel kategorikal, menghasilkan *dataset* numerik penuh (*heart_2020_encoded.csv*) yang siap untuk pemodelan.

Performa Model Baseline XGBoost



Sumber: Penelitian (2025).

Gambar 2. confusion matrix *Baseline XGBoost*.

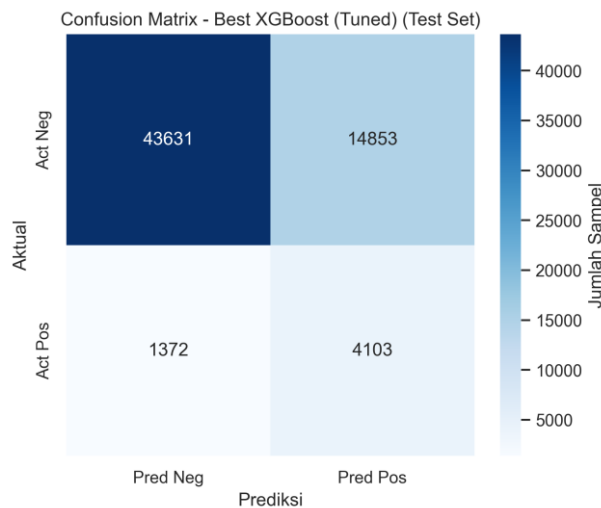
Model *XGBoost baseline* dengan parameter default menunjukkan kinerja yang tidak memadai untuk skrining, dengan *recall* hanya 0,1218 (12,18%) pada data uji. *Confusion matrix* mengungkap 4.971 *false negatives*, mengindikasikan kegagalan model dalam mendeteksi mayoritas kasus berisiko. Nilai *precision* (0,6797) dan *F1-Score* (0,2066) yang rendah semakin mengonfirmasi ketidaksiapan model untuk diterapkan.

Hasil Optimasi Hyperparameter dan Penanganan Class Imbalance

Optimasi dengan *RandomizedSearchCV* dan *5-fold stratified cross-validation* berhasil mengidentifikasi konfigurasi *hyperparameter XGBoost* yang optimal. Untuk menangani ketidakseimbangan kelas, diterapkan parameter *scale_pos_weight* = 10,683 (berdasarkan rasio kelas). Konfigurasi akhir terdiri dari: *n_estimators*: 250, *max_depth*: 9, *learning_rate*: 0,0706, *subsample*: 0,656, *colsample_bytree*: 0,628, dan *min_child_weight*: 6. Konfigurasi ini meningkatkan *F1-Score* validasi silang menjadi 0,3401, yang mengindikasikan peningkatan keseimbangan kinerja model.

Evaluasi Model XGBoost Teroptimasi (Sebelum Threshold Tuning)

Setelah optimasi, model *XGBoost* dievaluasi pada data uji. Performa menunjukkan peningkatan signifikan dibanding *baseline*, terutama dalam hal recall.



Sumber: Hasil Penelitian (2025)

Gambar 3. confusion matrix *XGBoost* Teroptimasi.

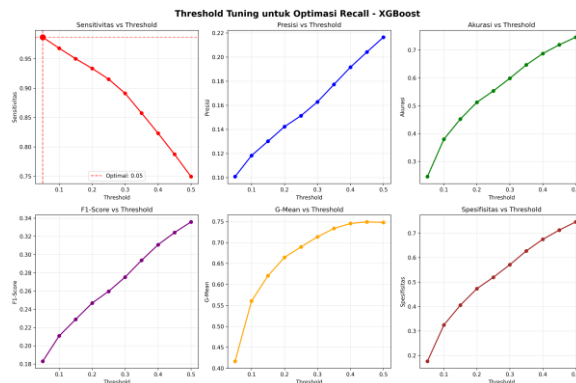
Model menghasilkan 4.103 *True Positives (TP)* dan 1.372 *False Negatives (FN)*. Jumlah *FN* yang berkurang drastis dari 4.971 (*baseline*) menjadi 1.372 menunjukkan peningkatan kemampuan deteksi.

Tabel 2. Evaluasi Model *XGBoost* Teroptimasi (Sebelum Tuning)

| Recall | Akurasi | Presisi | F1-Score | AUC-ROC | AUC-PR | G-Mean |
|--------|---------|---------|----------|---------|--------|--------|
| 0,7494 | 0,7496 | 0,2164 | 0,3367 | 0,8232 | 0,3237 | 0,7477 |

Meskipun *recall* telah meningkat menjadi ~75%, nilai ini masih dianggap belum optimal untuk sistem skrining yang memprioritaskan deteksi maksimal.

Threshold Tuning untuk Optimasi Recall



Sumber: Hasil Penelitian (2025).

Gambar 4. Threshold Tuning untuk Optimasi Recall - XGBoost.

Dalam konteks skrining populasi yang memprioritaskan deteksi dini, meminimalkan *false negatives (missed diagnosis)* menjadi imperatif klinis utama. Oleh karena itu, peneliti melakukan penyesuaian strategis terhadap ambang batas klasifikasi (*threshold*) dari nilai konvensional 0.5. Analisis komprehensif terhadap *trade-off* antar metrik mengungkap dinamika yang jelas: *sensitivity (recall)* mencapai optimumnya ($>0,96$) pada threshold yang sangat rendah ($\approx 0,10$), sedangkan precision mengalami penurunan yang signifikan seiring penurunan threshold sebuah pertukaran (*trade-off*) yang dapat diterima dalam paradigma skrining.

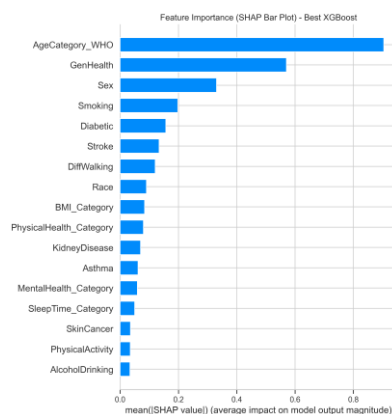
Tabel 3. Hasil Threshold Tuning.

| Recall | Akurasi | Presisi | F1-Score | G-Mean |
|--------|---------|---------|----------|--------|
| 0,9676 | 0,7243 | 0,1864 | 0,3128 | 0,7612 |

Menariknya, *F1-Score* dan *G-Mean* memuncak pada threshold yang lebih tinggi (0,7–0,8), merepresentasikan titik keseimbangan ideal untuk konteks evaluasi umum, namun bukan untuk skrining yang berorientasi *recall*. Berdasarkan pertimbangan klinis-etis ini, peneliti menetapkan threshold optimal sebesar 0,10. Dengan konfigurasi ini, model *XGBoost* yang telah dioptimasi mencapai performa kritis: *sensitivity* 96,79%, akurasi 72,43%, presisi 18,64%, *F1-Score* 31,28%, dan *G-Mean* 76,12% sebuah profil kinerja yang secara sengaja mengutamakan deteksi kasus maksimal atas akurasi keseluruhan.

Interpretasi Model XGBoost dengan SHAP

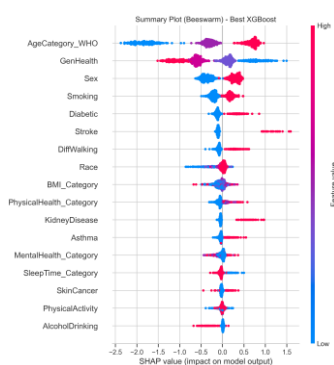
Interpretasi dilakukan pada model *XGBoost* teroptimasi (dengan *hyperparameter* terbaik) menggunakan probabilitas prediksi mentah, sebelum penerapan *threshold* 0.10. Analisis ini memberikan penjelasan transparan mengenai kontribusi fitur.



Sumber: Hasil Penelitian (2025).

Gambar 5. Feature Importance (SHAP Bar Plot) - Best XGBoost.

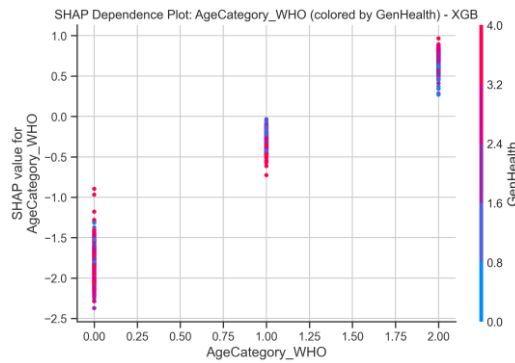
Berdasarkan nilai *SHAP* absolut rata-rata, fitur paling berpengaruh adalah *AgeCategory_WHO* (usia), diikuti oleh *GenHealth* (kondisi kesehatan umum), *Sex*, *Smoking*, dan *Diabetic*. Urutan ini selaras dengan pengetahuan medis mengenai faktor risiko penyakit jantung.



Sumber: Hasil Penelitian (2025).

Gambar 6. Summary Plot (Beeswarm).

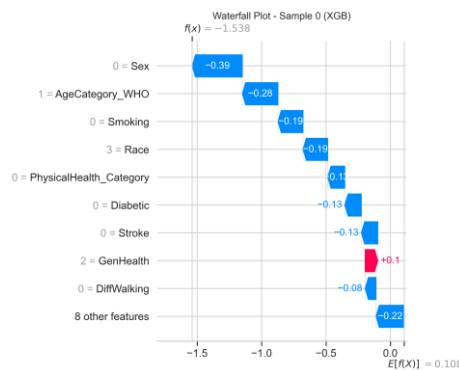
Berdasarkan nilai *SHAP* absolut rata-rata, fitur paling berpengaruh adalah *AgeCategory_WHO* (usia), diikuti oleh *GenHealth* (kondisi kesehatan umum), *Sex*, *Smoking*, dan *Diabetic*. Urutan ini selaras dengan pengetahuan medis mengenai faktor risiko penyakit jantung.



Sumber: Hasil Penelitian (2025).

Gambar 6. SHAP Dependence Plot.

Berdasarkan nilai SHAP absolut rata-rata, fitur paling berpengaruh adalah *AgeCategory_WHO* (usia), diikuti oleh *GenHealth* (kondisi kesehatan umum), *Sex*, *Smoking*, dan *Diabetic*. Urutan ini selaras dengan pengetahuan medis mengenai faktor risiko penyakit jantung.



Sumber: Hasil Penelitian (2025).

Gambar 7. SHAP Waterfall Plot.

Berdasarkan nilai SHAP absolut rata-rata, fitur paling berpengaruh adalah *AgeCategory_WHO* (usia), diikuti oleh *GenHealth* (kondisi kesehatan umum), *Sex*, *Smoking*, dan *Diabetic*. Urutan ini selaras dengan pengetahuan medis mengenai faktor risiko penyakit jantung.

Analisis Pasien Berisiko Tinggi dan Rendah Berdasarkan SHAP (XGBoost)

Analisis pada level individu memperkuat validasi klinis (Tabel 4.11). Pasien dengan prediksi risiko tinggi (misalnya, ID dengan probabilitas >93%) secara konsisten memiliki kombinasi faktor seperti *Stroke*=1, *GenHealth*=Poor/Fair, *AgeCategory_WHO*=Older Adults, dan *Diabetic*=1. Sebaliknya, pasien risiko rendah ditandai oleh

AgeCategory_WHO=Young Adults dan *GenHealth=Excellent*. Pola ini sesuai dengan pemahaman klinis yang telah teruji.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, disimpulkan bahwa model *XGBoost* yang dioptimalkan melalui *hyperparameter tuning*, penanganan *class imbalance*, dan *threshold tuning* mencapai *recall* 96,79% kinerja yang ideal untuk skrining kesehatan yang mengutamakan minimalisasi *false negative*. Kendala interpretabilitas *black-box* berhasil diatasi dengan penerapan *SHAP*, yang menghasilkan visualisasi (*feature importance*, *summary plot*, *dependence plot*, dan *waterfall plot*) yang mengungkap kontribusi klinis fitur secara transparan. Analisis *SHAP* mengonfirmasi faktor risiko dominan seperti usia lanjut, kondisi kesehatan umum buruk, serta riwayat stroke, diabetes, dan penyakit ginjal sesuai dengan bukti medis yang ada. Dengan demikian, model ini tidak hanya akurat, tetapi juga dapat dijelaskan (*explainable*), sehingga berpotensi meningkatkan kepercayaan dan adopsi klinis.

Namun, penerapan temuan ini memerlukan kehati-hatian mengingat keterbatasan ruang lingkup data. Untuk memastikan keberlakuan dalam konteks Indonesia, saran utama adalah validasi eksternal menggunakan dataset populasi lokal. Sebagai tindak lanjut praktis, disarankan pengembangan dashboard klinis interaktif yang mengintegrasikan model *XGBoost* beserta penjelasan *SHAP*, memungkinkan tenaga kesehatan memperoleh prediksi dan interpretasi visual secara *real-time* untuk mendukung pengambilan keputusan berbasis bukti. Penelitian selanjutnya dapat mengeksplorasi teknik optimasi dan interpretabilitas alternatif, serta integrasi variabel klinis yang lebih spesifik untuk meningkatkan akurasi dan kedalaman wawasan sistem

UCAPAN TERIMA KASIH

Puji syukur kehadiran Allah SWT atas rahmat dan karunia-Nya, penulis dapat menyelesaikan tesis dengan judul “**Optimasi XGBoost Dengan Shap Untuk Sistem Skrining Penyakit Jantung**”.

Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang, penulis memanjatkan syukur atas selesainya tesis ini. Tulus terima kasih penulis sampaikan kepada: Bapak Dr. Ir. Eddy Anthony, SH, MM. (Ketua Yayasan Universitas Dinamika Bangsa); Bapak

Prof. Setiawan Assegaff, S.T., M.M.S.I., Ph.D. (Rektor); serta Bapak Dr. Jasmir, S.Kom., M.Kom. selaku Ketua Prodi Magister Sistem Informasi dan Pembimbing I yang telah memberikan bimbingan berharga. Terima kasih sepenuh hati kepada Bapak Dr. Fachruddin, S.Pt., M.S.I. selaku Pembimbing II atas ilmu dan koreksi mendalam selama penelitian.

Penulis juga mengucapkan terima kasih kepada seluruh dosen, staf, dan karyawan Program Studi Magister Sistem Informasi Universitas Dinamika Bangsa atas fasilitas dan dukungannya. Ucapan terima kasih juga disampaikan kepada Kaggle sebagai penyedia data.

Penulis menghaturkan terima kasih dan doa tulus kepada Ibu Ely Syahtinar dan Almarhum Ayah Zul Azaini atas kasih sayang serta dukungan tanpa batas. Terima kasih kepada Muhammad Naufal, S.Kom. atas pendampingan, dukungan materi, dan doa yang tulus.

Akhir kata, terima kasih kepada diri sendiri atas ketekunan, serta kepada rekan-rekan seperjuangan atas semangat yang diberikan. Penulis juga berterima kasih kepada semua pihak yang telah berkontribusi, baik langsung maupun tidak langsung.

DAFTAR REFERENSI

- Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. *Sensors*, 23(2), 634. <https://doi.org/10.3390/s23020634>
- Explainable Artificial Intelligence for Biomedical and Healthcare Applications. (2025). CRC Press. https://www.google.co.id/books/edition/Explainable_Artificial_Intelligence_for/6U4aEQAAQBAJ?hl=en&gbpv=0
- Guan, C., Gong, A., Zhao, Y., Yin, C., Geng, L., Liu, L., Yang, X., Lu, J., & Xiao, B. (2024). Interpretable machine learning model for new-onset atrial fibrillation prediction in critically ill patients: A multi-center study. *Critical Care*, 28(1), 349. <https://doi.org/10.1186/s13054-024-05138-0>
- Harrison, M. (n.d.). Effective Xgboost.
- Houssein, E. H., Mohamed, R. E., & Ali, A. A. (2023). Heart disease risk factors detection from electronic health records using advanced NLP and deep learning techniques. *Scientific Reports*, 13(1), 7173. <https://doi.org/10.1038/s41598-023-34294-6>
- Jones, O. T., Matin, R. N., Van Der Schaar, M., Prathivadi Bhayankaram, K., Ranmuthu, C. K. I., Islam, M. S., Behiyat, D., Boscott, R., Calanzani, N., Emery, J., Williams, H. C., & Walter, F. M. (2022). Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: A systematic review. *The Lancet Digital Health*, 4(6), e466–e476. [https://doi.org/10.1016/S2589-7500\(22\)00023-1](https://doi.org/10.1016/S2589-7500(22)00023-1)
- Khamparia, A., & Gupta, D. (Eds.). (2025). *Explainable artificial intelligence for biomedical and healthcare applications* (First edition). CRC Press. <https://doi.org/10.1201/9781003220107>

- Klosterman, S. (2021). *Data science projects with Python: A case study approach to gaining valuable insights from real data with machine learning*, 2nd edition (2nd ed). Packt Publishing.
- Li, J., Liu, S., Hu, Y., Zhu, L., Mao, Y., & Liu, J. (2022). Predicting Mortality in Intensive Care Unit Patients With Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study. *Journal of Medical Internet Research*, 24(8), e38082. <https://doi.org/10.2196/38082>
- Littlewood, J. R., Jain, L., & Howlett, R. J. (Eds.). (2024). *Sustainability in Energy and Buildings 2023* (1st ed. 2024). Springer Nature Singapore. <https://doi.org/10.1007/978-981-99-8501-2>
- Luo, H., Xiang, C., Zeng, L., Li, S., Mei, X., Xiong, L., Liu, Y., Wen, C., Cui, Y., Du, L., Zhou, Y., Wang, K., Li, L., Liu, Z., Wu, Q., Pu, J., & Yue, R. (2024). SHAP based predictive modeling for 1 year all-cause readmission risk in elderly heart failure patients: Feature selection and model interpretation. *Scientific Reports*, 14(1), 17728. <https://doi.org/10.1038/s41598-024-67844-7>
- Mahalle, P. N., & Ingle, Y. S. (2024). *Explainable artificial intelligence: A practical guide*. River Publishers. <https://doi.org/10.1201/9788770047142>
- Maharani, A., Sujarwoto, Praveen, D., Oceandy, D., Tampubolon, G., & Patel, A. (2024). Implementation of mobile-health technology is associated with five-year survival among individuals in rural areas of Indonesia. *PLOS Digital Health*, 3(4), e0000476. <https://doi.org/10.1371/journal.pdig.0000476>
- Marcolino, M. S., Schettini, I. V. G., Do Nascimento, G. F., Da Rocha, L. C. D., Lana, F. C. B., & Gonçalves, M. A. (2025). Explainable artificial intelligence for predicting cardiovascular events in hospitalised COVID-19 patients. *BMC Infectious Diseases*, 25(1), 1569. <https://doi.org/10.1186/s12879-025-11465-6>
- Oikonomou, E. K., & Khera, R. (2023). Machine learning in precision diabetes care and cardiovascular risk prediction. *Cardiovascular Diabetology*, 22(1), 259. <https://doi.org/10.1186/s12933-023-01985-3>
- Rahman, A., Hossain, Md. S., Muhammad, G., Kundu, D., Debnath, T., Rahman, M., Khan, Md. S. I., Tiwari, P., & Band, S. S. (2023). Federated learning-based AI approaches in smart healthcare: Concepts, taxonomies, challenges and open issues. *Cluster Computing*, 26(4), 2271–2311. <https://doi.org/10.1007/s10586-022-03658-4>
- Reátegui, R., Tandazo-Malla, C., Suárez, R., & Ramírez-Cerna, L. (2025). Cardiovascular risk prediction via ensemble machine learning and oversampling methods. *Scientific Reports*, 15(1), 43576. <https://doi.org/10.1038/s41598-025-30895-5>
- Rokom. (2022, September 29). Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer – Sehat Negeriku. Sehat Negeriku. <https://kemkes.go.id/id/penyakit-jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer>
- Sergouniotti, A., Rigas, D., Zoitopoulos, V., & Kalles, D. (2025). From Preliminary Urinalysis to Decision Support: Machine Learning for UTI Prediction in Real-World Laboratory Data. *Journal of Personalized Medicine*, 15(5), 200. <https://doi.org/10.3390/jpm15050200>

- Siefkes, H., Oliveira, L. C., Koppel, R., Hogan, W., Garg, M., Manalo, E., Cresalia, N., Lai, Z., Tancredi, D., Lakshminrusimha, S., & Chuah, C. (2024). Machine Learning–Based Critical Congenital Heart Disease Screening Using Dual-Site Pulse Oximetry Measurements. *Journal of the American Heart Association*, 13(12), e033786. <https://doi.org/10.1161/JAHA.123.033786>
- tang, x. (n.d.). *Mobile Computing And Sustainable Informatics*. Springer Singapore.
- Wang, K., Tian, J., Zheng, C., Yang, H., Ren, J., Liu, Y., Han, Q., & Zhang, Y. (2021). Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Computers in Biology and Medicine*, 137, 104813. <https://doi.org/10.1016/j.compbiomed.2021.104813>
- WHO. (2024, Agustus). The top 10 causes of death. WHO. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- Ye, Z., An, S., Gao, Y., Xie, E., Zhao, X., Guo, Z., Li, Y., Shen, N., Ren, J., & Zheng, J. (2023). The prediction of in-hospital mortality in chronic kidney disease patients with coronary artery disease using machine learning models. *European Journal of Medical Research*, 28(1), 33. <https://doi.org/10.1186/s40001-023-00995-x>