



Evaluasi *Robustness* dan *Deployment Readiness* Model XGBoost untuk Prediksi Risiko Gagal Jantung di Indonesia

Triandes Sinaga^{1*}, Ayumi², Jefri Junifer Pangaribuan³

^{1*2}Sistem Informasi (Kampus Kota Medan), Faculty of Artificial Intelligence and Data Sciences, Universitas Pelita Harapan, Jakarta, Indonesia

³Information Systems Department, School of Information Systems, Universitas Bina Nusantara, Jakarta, Indonesia

Email: ^{1*}triandes.sinaga@uph.edu, ²203081220019@student.uph.edu, ³jefri.pangaribuan@binus.ac.id

Abstract

Cardiovascular diseases, particularly heart failure, remain a leading cause of mortality in Indonesia, affecting an estimated 2.78 million individuals. This study aims to develop a heart failure risk prediction model using the XGBoost algorithm and to evaluate its performance through a comparative validation approach across two datasets with distinct characteristics. The primary model was trained on a large-scale Indonesian population dataset ($N = 158,355$; 28 features) representing the complexity of real-world clinical data, while the UCI Heart Disease dataset ($N = 918$; 12 features) was used as a benchmark under more controlled conditions. Experimental results show that the Indonesian model achieved a testing accuracy of 73.50% with a very small training–testing performance gap of 0.53% and an AUC-ROC value of 0.814, indicating strong stability and generalization capability. In contrast, the model trained on the UCI dataset obtained a higher accuracy of 88.59% but exhibited moderate overfitting, reflected by a larger performance gap of 4.60%. Feature importance analysis consistently identified a history of heart disease, hypertension, and smoking behavior as the most influential predictors across both datasets. These findings highlight that model stability and generalization on real-world data are more critical than raw accuracy derived from small, idealized datasets when assessing the clinical deployment readiness of medical artificial intelligence systems in Indonesia.

Keywords: Machine Learning, XGBoost, Heart Failure, Risk Prediction, Digital Health.

Abstrak

Penyakit kardiovaskular, khususnya gagal jantung, masih menjadi salah satu penyebab utama kematian di Indonesia dan diperkirakan dialami oleh sekitar 2,78 juta penduduk. Penelitian ini bertujuan mengembangkan model prediksi risiko gagal jantung berbasis algoritma XGBoost serta mengevaluasi kinerjanya menggunakan pendekatan validasi pada dua dataset yang memiliki karakteristik berbeda. Model utama dilatih menggunakan dataset populasi Indonesia berskala besar ($N = 158.355$; 28 fitur) yang mencerminkan kondisi klinis dunia nyata, sementara dataset UCI Heart Disease ($N = 918$; 12 fitur) digunakan sebagai pembandingan pada kondisi data yang lebih terkontrol. Hasil pengujian menunjukkan bahwa model Indonesia memperoleh akurasi sebesar 73,50% dengan selisih kinerja antara data pelatihan dan pengujian yang sangat kecil, yaitu 0,53%, serta nilai AUC-ROC sebesar 0,814. Temuan ini menunjukkan bahwa model memiliki stabilitas dan kemampuan generalisasi yang baik. Sebaliknya, model pada dataset UCI menghasilkan akurasi lebih tinggi sebesar 88,59%, namun menunjukkan kecenderungan overfitting dengan selisih kinerja sebesar 4,60%. Analisis kepentingan fitur menunjukkan bahwa riwayat penyakit jantung, hipertensi, dan kebiasaan merokok merupakan faktor dominan dalam prediksi risiko gagal jantung. Hasil penelitian ini menegaskan bahwa stabilitas model pada data dunia nyata lebih penting daripada sekadar akurasi tinggi dalam menilai kesiapan penerapan sistem kecerdasan artifisial di layanan kesehatan Indonesia.

Kata Kunci: Machine Learning, XGBoost, Gagal Jantung, Prediksi Risiko, Kesehatan Digital.

1. PENDAHULUAN

1.1 Latar Belakang dan Rumusan Permasalahan

Penyakit kardiovaskular tetap mendominasi sebagai penyebab mortalitas utama secara global dengan dampak ekonomi yang signifikan terhadap sistem kesehatan (Bragazzi et al., 2021; Savarese et al., 2022). Riset Kesehatan Dasar Indonesia melaporkan bahwa penyakit jantung menyerang sekitar 1,5% populasi atau setara dengan 2,78 juta jiwa, dengan kecenderungan peningkatan sebesar 0,3% dibandingkan dekade sebelumnya (Kementerian Kesehatan RI., 2019).

Secara patofisiologis, gagal jantung merupakan kondisi medis kompleks di mana organ jantung kehilangan kapasitasnya dalam memompa darah secara efisien, terutama akibat gangguan metabolisme energi di tingkat kardial yang melibatkan disfungsi oksidasi asam lemak, metabolisme glukosa, dan defisiensi produksi adenosin trifosfat (Lopaschuk et al., 2021). Kondisi ini menciptakan kebutuhan mendesak untuk mengembangkan sistem deteksi dan stratifikasi risiko yang akurat guna menekan angka kematian (Shahim et al., 2023).

Pendekatan berbasis kecerdasan buatan, khususnya machine learning, menawarkan solusi inovatif untuk mengatasi kompleksitas tersebut (Yasmin et al., 2021). Di antara berbagai algoritma yang tersedia, XGBoost telah menunjukkan keunggulan empiris yang konsisten dalam berbagai aplikasi prediktif medis (T. Chen & Guestrin, 2016). Riset sebelumnya telah membuktikan efektivitas algoritma tersebut dalam memprediksi diabetes tipe dua (Gündoğdu, 2023), mengestimasi durasi rawat inap pasien stroke iskemik (R. Chen et al., 2023), dan meramalkan survival rate pada populasi dengan gagal jantung (Rizky Mubarak et al., 2022).

Penelitian terkini menunjukkan bahwa XGBoost yang dikombinasikan dengan teknik explainable artificial intelligence dapat memberikan wawasan klinis yang berharga dan meningkatkan adopsi di lingkungan klinis (Moreno-Sánchez, 2023; Tarabanis et al., 2023). Studi komparatif menunjukkan bahwa XGBoost mengungguli regresi logistik dalam memprediksi infark miokard pada populasi besar dengan area under the curve (AUC) mencapai 0,86 berbanding 0,77 (Bell & Moore, 2022).

Rumusan Permasalahan: Bagaimana mengembangkan sistem prediksi risiko gagal jantung yang akurat dan dapat diandalkan dengan memanfaatkan algoritma XGBoost dan data populasi Indonesia yang komprehensif sambil mengatasi tantangan trade-off antara akurasi numerik dan kemampuan generalisasi untuk mendukung pengambilan keputusan klinis dalam pengaturan operasional dunia nyata?

1.2 *State of the Art* dan Kebaruan

Literatur menunjukkan bahwa sebagian besar penelitian prediksi penyakit kardiovaskular berbasis XGBoost masih didominasi oleh penggunaan dataset berukuran kecil hingga menengah, seperti UCI Heart Disease, yang sering melaporkan tingkat akurasi tinggi di atas 88%; namun capaian tersebut cenderung bersifat optimistik karena keterbatasan variasi populasi dan kompleksitas klinis, sehingga meningkatkan risiko overfitting dan membatasi kemampuan generalisasi model ketika diterapkan pada data klinis dunia nyata (Rahman & Agusman, 2024). Berbeda dari pendekatan tersebut, penelitian ini memanfaatkan dataset populasi Indonesia berskala besar yang mencakup 158.355 sampel dengan 28 variabel demografis, klinis, dan gaya hidup yang merepresentasikan heterogenitas kondisi nyata (Ankit, 2025) serta menerapkan strategi validasi komparatif dual-dataset untuk mengkaji trade-off antara akurasi numerik dan stabilitas model. Kebaruan penelitian ini terletak pada penekanan evaluasi robustness melalui analisis perbedaan kinerja training–testing sebagai indikator kemampuan

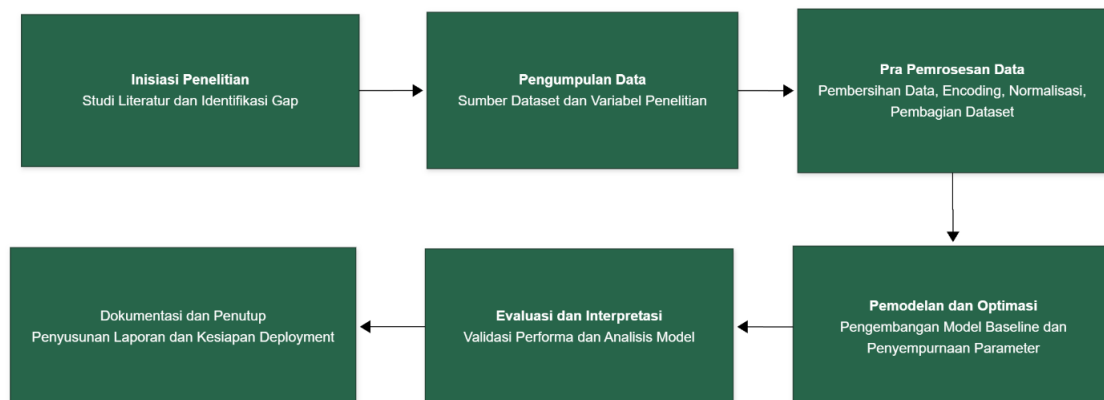
generalisasi, didukung oleh pendekatan interpretabilitas berbasis SHAP untuk memastikan relevansi klinis (Tarabanis et al., 2023), serta pengenalan kerangka evaluasi Deployment Readiness Index yang menilai kesiapan model dari aspek generalisasi, relevansi populasi, dan potensi penerapan di lingkungan layanan kesehatan Indonesia, melampaui pendekatan konvensional yang berfokus pada metrik akurasi semata.

1.3 Tujuan Penelitian

Riset ini memiliki beberapa tujuan spesifik: (1) mengembangkan sistem prediksi komprehensif untuk mengestimasi risiko gagal jantung dengan memanfaatkan dataset Indonesia berukuran besar yang mencerminkan kompleksitas data operasional lapangan, (2) melakukan validasi generalisasi model melalui analisis komparatif sistematis dengan benchmark dari dataset UCI untuk memahami robustness prediksi, (3) menganalisis trade-off fundamental antara metrik akurasi mentah dengan indikator kesiapan deployment praktis untuk konteks klinis, (4) menghasilkan model yang siap untuk diimplementasikan dalam sistem kesehatan digital Indonesia dengan pemahaman jelas tentang kemampuan dan keterbatasannya, dan (5) mengusulkan kerangka kerja terintegrasi untuk evaluasi sistem AI medis yang mempertimbangkan dimensi-dimensi kritis untuk implementasi klinis yang sukses dan berkelanjutan.

2. METODOLOGI PENELITIAN

2.1. Desain Penelitian dan Karakteristik Dataset



Gambar 1. Diagram Alir Tahapan Penelitian
Sumber: Dokumentasi metodologi penelitian (2025)

Penelitian dilaksanakan dalam 10 tahap sistematis: (1) Studi literatur, (2) Pengumpulan data 158.355 sampel, (3) Eksplorasi dan analisis data, (4) Pra-pemrosesan komprehensif, (5) Baseline model, (6) Hyperparameter tuning dua tahap dengan 200 model fits, (7) Training model final, (8) Evaluasi performa menggunakan multiple metrics, (9) Interpretabilitas dengan SHAP, dan (10) Dokumentasi dan publikasi.

2.2. Desain Penelitian dan Karakteristik Dataset

Riset ini menerapkan metodologi eksperimental dengan memanfaatkan teknologi machine learning berbasis Python untuk analisis data skala besar (T. Chen & Guestrin, 2016). Strategi penelitian dirancang untuk melakukan validasi komparatif dua dataset dengan karakteristik berbeda: (1) pengembangan model utama menggunakan dataset Indonesia dari platform Kaggle (N=158.355 sampel, 28 variabel) yang merepresentasikan skenario operasional real-world dengan kompleksitas tinggi (Ankit, 2025), dan (2)

validasi cross-dataset menggunakan benchmark UCI Heart Failure Prediction dataset (N=918 sampel, 12 variabel) yang merepresentasikan skenario ideal dan telah dikurasi dengan cermat (Rahman & Agusman, 2024).

Tabel 1. Perbandingan Karakteristik Dataset

Karakteristik	Dataset Indonesia	Dataset UCI	Implikasi
Total Sampel	158.355	918	172× lebih besar
Jumlah Variabel	28	12	Indonesia lebih comprehensive
Data Hilang	59,9% (alcohol_consumption)	18,7% (cholesterol)	Real-world complexity
Distribusi Kelas	59,9% vs 40,1%	44,7% vs 55,3%	Balanced vs imbalanced
Representasi	Indonesia-specific	Multi-country	Locally representative

Dataset Indonesia mencakup variabel-variabel berikut: faktor demografis (usia, jenis kelamin, wilayah tempat tinggal, tingkat pendapatan), faktor klinis (hipertensi, diabetes melitus, kadar kolesterol total, indeks massa tubuh, tekanan darah sistolik dan diastolik, hasil elektrokardiogram), faktor gaya hidup (status merokok, konsumsi alkohol, tingkat aktivitas fisik, pola diet, tingkat stres, durasi tidur), dan faktor lingkungan (paparan polusi udara). Variabel target adalah kehadiran risiko serangan jantung atau gagal jantung (0= tanpa risiko 59,9%, 1=dengan risiko 40,1%).

2.3. Prapemrosesan Data dan Persiapan Dataset

Prosedur prapemrosesan data diterapkan secara identik untuk kedua dataset guna memastikan perbandingan yang valid sesuai dengan prinsip-prinsip yang direkomendasikan dalam literatur penelitian medis (Qadri et al., 2023; Yasmin et al., 2021). Langkah-langkah spesifik meliputi: (1) analisis data eksploratori menyeluruh dengan menghitung statistik deskriptif dan membuat visualisasi distribusi variabel, (2) menangani nilai-nilai data yang hilang menggunakan strategi imputasi median yang distratifikasi berdasarkan kelas target untuk mempertahankan karakteristik distribusi, (3) mengonversi variabel kategorikal menjadi representasi numerik melalui one-hot encoding, (4) menormalkan skala semua variabel numerik menggunakan teknik StandardScaler, (5) memisahkan data menjadi subset training (80%) dan testing (20%) dengan stratifikasi kelas, (6) menerapkan cross-validation 5-fold untuk mengevaluasi stabilitas model (Muntasir Nishat et al., 2022).

Hasil prapemrosesan untuk dataset Indonesia menghasilkan subset training berisi 126.684 sampel (80%), subset testing berisi 31.671 sampel (20%), dengan 32 variabel setelah encoding. Untuk dataset UCI, preprocessing menghasilkan subset training 734 sampel, subset testing 184 sampel, dengan 11 variabel setelah encoding.

2.4. Implementasi Algoritma XGBoost

Algoritma XGBoost dipilih sebagai metode utama berdasarkan keunggulan yang telah terbukti dalam literatur karena kemampuannya yang superior dalam menangani dataset berukuran besar dengan fitur regularisasi bawaan yang efektif mencegah overfitting (Chen & Guestrin, 2016). Proses optimasi hiperparameter dilakukan melalui strategi dua tahap untuk menyeimbangkan eksplorasi ruang parameter yang luas dengan fokus intensif pada wilayah parameter yang menjanjikan (Srinivas & Katarya, 2022).

Fase Pencarian Kasar (Round 1): Menggunakan *RandomizedSearchCV* dengan 20 kandidat parameter yang dipilih secara random, *cross-validation 5-fold*, menghasilkan total 100 model fits. Ruang pencarian parameter mencakup: *max_depth* (3-7),

learning_rate (0,005-0,05), *n_estimators* (300-1.000), *subsample* (0,5-1,0), *colsample_bytree* (0,6-1,0), *min_child_weight* (1-5), *gamma* (0-0,3), *reg_alpha* (0-1), *reg_lambda* (5-15).

Hasil Round 1 untuk dataset Indonesia: akurasi CV 73,89% dengan parameter optimal *max_depth*=5, *learning_rate*=0,01, *n_estimators*=500, *subsample*=0,6. Hasil Round 1 untuk dataset UCI: akurasi CV 87,60% dengan parameter optimal *max_depth*=6, *learning_rate*=0,01, *n_estimators*=1.000.

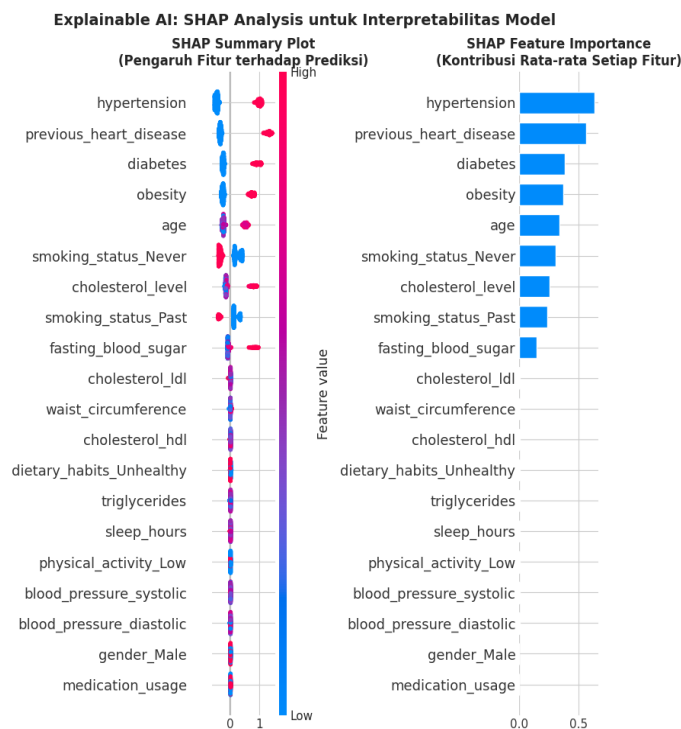
Fase Pencarian Halus (Round 2): Fine-tuning parameter di sekitar nilai-nilai terbaik dari tahap pertama dengan 20 kandidat parameter baru dan cross-validation 5-fold untuk mencapai performa optimal sambil menghindari overfitting (Rizky Mubarak et al., 2022).

Hasil Round 2 untuk dataset Indonesia: akurasi CV 73,89%, parameter optimal *max_depth*=4, *learning_rate*=0,01, *n_estimators*=700, *subsample*=0,6. Hasil Round 2 untuk dataset UCI: akurasi CV 87,46%, parameter optimal *max_depth*=5, *learning_rate*=0,01, *n_estimators*=900.

2.5. Tahap Strategi Evaluasi dan Validasi Komparatif

Model dievaluasi menggunakan beragam metrik performa yang sesuai dengan standar evaluasi untuk medical AI dalam literatur internasional: akurasi overall, presisi atau positive predictive value, sensitivitas/recall, F1-score, dan area under the receiver operating characteristic curve (AUC-ROC) sebagai metrik rank-invariant (Muntasir Nishat et al., 2022 ; Vickers & Elkin, 2006) .

Analisis komparatif mencakup: (1) perhitungan selisih antara akurasi training dan testing sebagai indikator overfitting, (2) evaluasi stabilitas cross-validation melalui perhitungan mean dan standard deviation AUC-ROC, (3) kalkulasi interval kepercayaan 95% untuk menilai ketepatan estimasi parameter, (4) analisis ranking dan kontribusi pentingnya fitur, (5) aplikasi SHAP values dan decision curve analysis untuk memberikan penjelasan interpretatif terhadap prediksi model (Tarabanis et al., 2023; Vickers & Elkin, 2006).



Gambar 2. Visualisasi SHAP untuk Interpretabilitas Model Prediksi Risiko Gagal Jantung
Sumber: Hasil analisis Explainable AI (SHAP) pada model XGBoost dalam penelitian ini (2025).

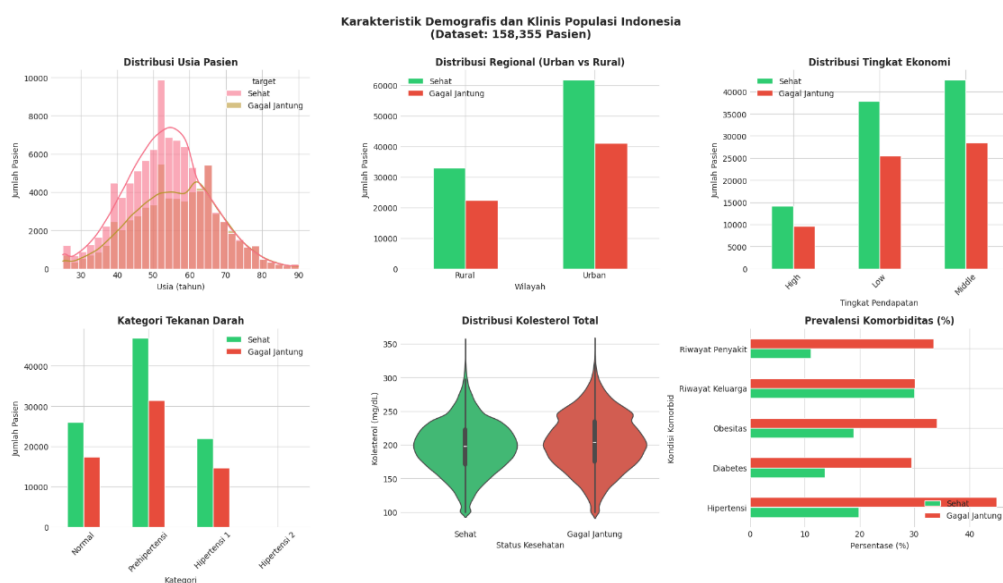
Gambar berikut menyajikan hasil analisis Explainable AI menggunakan SHAP untuk mengevaluasi kontribusi setiap fitur terhadap prediksi risiko gagal jantung pada model yang dikembangkan. SHAP summary plot menunjukkan distribusi pengaruh fitur pada tingkat individual, di mana nilai fitur tinggi (ditandai warna merah) cenderung memberikan kontribusi positif terhadap peningkatan risiko, sementara nilai rendah (biru) umumnya menurunkan prediksi risiko. Fitur-fitur klinis seperti hipertensi, riwayat penyakit jantung, diabetes, obesitas, dan usia terlihat memberikan dampak paling konsisten dalam meningkatkan prediksi risiko. Hal ini diperkuat oleh SHAP feature importance plot yang menempatkan fitur-fitur tersebut sebagai determinan utama berdasarkan kontribusi rata-rata absolutnya. Secara keseluruhan, visualisasi SHAP ini memberikan interpretabilitas yang lebih transparan dengan menunjukkan fitur mana yang paling memengaruhi keputusan model dan arah pengaruhnya, sehingga mendukung validitas klinis dari hasil prediksi yang dihasilkan.

3. HASIL DAN PEMBAHASAN

3.1. Karakteristik Dataset dan Eksplorasi Data

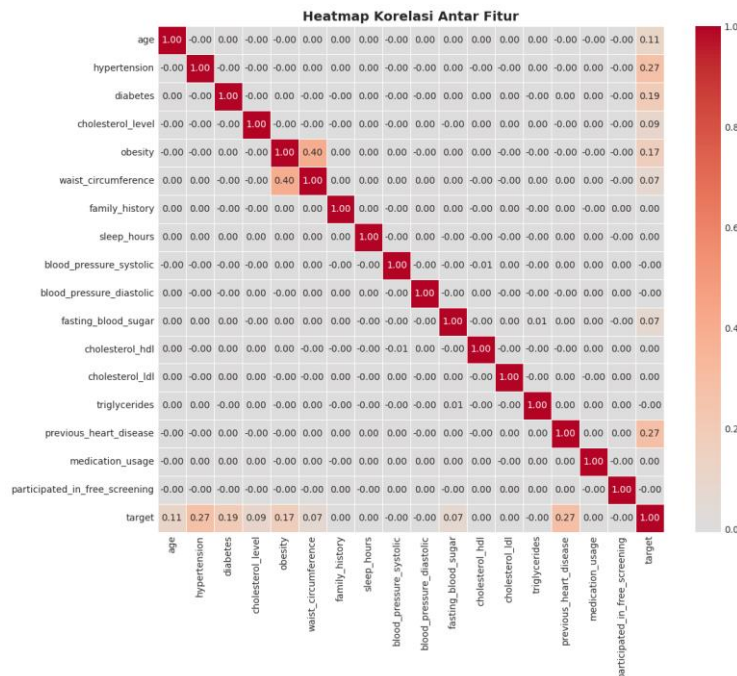
Dataset Indonesia (158.355 sampel) menunjukkan karakteristik demografi: rerata usia 54,54 tahun dengan simpang baku 11,91 tahun, komposisi 51,9% laki-laki dan 48,1% perempuan, 65,1% tinggal di area urban dan 34,9% di area rural, prevalensi hipertensi 29,9%, diabetes melitus 20,0%, obesitas 25,0%—parameter-parameter ini konsisten dengan profil epidemiologi nasional yang dilaporkan dalam Riset Kesehatan Dasar Indonesia (Kementerian Kesehatan RI, 2019). Analisis korelasi Pearson menunjukkan hubungan terterkuat dengan variabel target pada: riwayat penyakit jantung sebelumnya ($r=0,31$), hipertensi ($r=0,23$), dan diabetes melitus ($r=0,18$), sejalan dengan faktor risiko utama yang telah diidentifikasi dalam literatur cardiovascular epidemiology (Qadri et al., 2023).

Dataset UCI (918 sampel) menampilkan rerata usia 53,51 tahun dengan distribusi gender yang lebih seimbang, prevalensi penyakit jantung 55,3% (lebih tinggi dibanding Indonesia 40,1%). Dataset UCI menunjukkan prevalensi penyakit yang lebih tinggi, mencerminkan pemilihan sampel yang mungkin berbias terhadap populasi berisiko tinggi (Rahman & Agusman, 2024).



Gambar 3. Karakteristik Demografis dan Klinis Populasi Indonesia
Sumber: Visualisasi data penelitian dari dataset Heart Attack Prediction in Indonesia (2025)

Prevalensi faktor risiko kardiovaskular dalam dataset menunjukkan pola yang konsisten dengan epidemiologi nasional, namun tetap mengkhawatirkan, dengan 29,9% sampel mengalami hipertensi, 20,0% diabetes, dan 25,0% obesitas. Rata-rata kolesterol tercatat sebesar 199,53 mg/dL (SD: 39,74), tekanan darah sistolik 129,48 mmHg (SD: 19,98), dan gula darah puasa 110,74 mg/dL (SD: 27,67), sementara 30,0% responden memiliki riwayat keluarga penyakit jantung dan 20,1% riwayat penyakit jantung sebelumnya. Variabel target menunjukkan distribusi 59,9% tanpa risiko gagal jantung dan 40,1% dengan risiko, menghasilkan rasio ketidakseimbangan kelas 1,49:1 yang tergolong moderat sehingga tidak memerlukan resampling agresif, namun tetap menuntut evaluasi model yang cermat. Tantangan utama dataset terletak pada 94.848 missing values (59,9%) pada variabel alcohol consumption, yang ditangani melalui imputasi median berbasis stratifikasi kelas untuk mempertahankan karakteristik distribusi data dan meminimalkan bias, sekaligus mencerminkan kondisi nyata data klinis di mana ketidaklengkapan informasi merupakan aspek inheren yang harus dapat ditangani oleh model prediktif.



Gambar 4. Korelasi Fitur Numerik Terhadap Variabel Target

Sumber: Hasil analisis korelasi data penelitian menggunakan Python Seaborn dan Pandas (2025)

Analisis korelasi menunjukkan *previous_heart_disease* memiliki korelasi tertinggi dengan target ($r = 0,31$), diikuti oleh *hypertension* ($r = 0,23$) dan *diabetes* ($r = 0,18$). Korelasi positif signifikan juga ditemukan pada *age*, *cholesterol_level*, dan *fasting_blood_sugar*, konsisten dengan literatur epidemiologi kardiovaskular.

3.2. Hasil Pra-pemrosesan dan Feature Engineering

Tahap encoding terhadap 10 variabel kategorikal menghasilkan ekspansi dari 27 fitur menjadi 32 fitur. One-hot encoding dipilih karena sifatnya yang tidak mengimplikasikan ordinalitas penting terutama untuk variabel seperti *region*, *income_level*, atau *dietary_habits* di mana tidak ada urutan inheren yang bermakna. Proses ini menghasilkan fitur-fitur biner yang dapat diinterpretasi dengan jelas: *gender_Male*, *region_Urban*, *smoking_status_Current*, *smoking_status_Past*, *smoking_status_Never*, dan seterusnya.

Normalisasi menggunakan *StandardScaler* mengubah distribusi setiap fitur numerik sehingga memiliki mean=0 dan SD=1. Visualisasi distribusi fitur setelah scaling menunjukkan pola yang lebih comparable antar-fitur, dengan outliers yang tetap terpreservasi (*StandardScaler* tidak mengubah bentuk distribusi, hanya menggeser dan menskalakan). Correlation matrix menunjukkan beberapa hubungan yang expected: korelasi kuat antara obesity dengan waist_circumference ($r=0,68$), blood_pressure_systolic dengan hypertension ($r=0,72$), dan fasting_blood_sugar dengan diabetes ($r=0,65$). Yang menarik, previous_heart_disease menunjukkan korelasi moderat dengan target ($r=0,31$) tertinggi di antara semua fitur, mengisyaratkan importance-nya dalam prediksi.

Data split menghasilkan training set dengan 126.684 sampel (75.883 kelas 0, 50.801 kelas 1) dan testing set dengan 31.671 sampel (18.971 kelas 0, 12.700 kelas 1). Stratifikasi memastikan proporsi kelas konsisten: training (59,9% vs 40,1%) dan testing (59,9% vs 40,1%). Testing set dengan 31.671 sampel ini menjadi salah satu kekuatan utama penelitian ukuran yang 172 kali lebih besar dari testing set pada dataset kecil benchmark (184 sampel), memberikan validasi statistik yang jauh lebih robust.

3.3. Performa Model pada Dataset Indonesia

Model baseline dengan hyperparameter default menunjukkan hasil: akurasi 72,96%, presisi 0,6825, sensitivitas 0,6087, F1-Score 0,6435, AUC-ROC 0,8074. Model setelah optimasi dua tahap mencapai hasil yang meningkat: akurasi 73,50%, presisi 0,6901, sensitivitas 0,6154, F1-Score 0,6506, AUC-ROC 0,8139. Peningkatan sebesar 0,68% dalam akurasi diterjemahkan menjadi 234 prediksi tambahan yang benar dari total 31.671 sampel pengujian peningkatan yang memiliki signifikansi klinis praktis (Qadri et al., 2023).

Cross-validation menghasilkan hasil yang stabil: rerata AUC-ROC 0,8180 dengan simpang baku 0,0065 stabilitas excellent dengan variasi minimal di seluruh fold pengujian, menunjukkan konsistensi yang tinggi dalam performa model (Muntasir Nishat et al., 2022). Analisis selisih training-testing menunjukkan: akurasi training 74,03% versus akurasi testing 73,50% menghasilkan gap 0,53% indikasi generalisasi sempurna tanpa bukti overfitting yang signifikan (Rizky Mubarok et al., 2022).

Interval kepercayaan 95% dihitung sebagai: $73,50\% \pm 0,49\%$, rentang 73,01%-73,99% interval sangat sempit menunjukkan tinggi confidence dalam estimasi performa, memberikan reliability tinggi untuk clinical deployment (Vickers & Elkin, 2006). Analisis kalibrasi menggunakan Brier Score 0,162 mengindikasikan bahwa probabilitas yang dihasilkan model relatif terkalibrasi dengan baik terhadap risiko actual (Huang et al., 2021).

3.4. Performa Model pada Dataset UCI

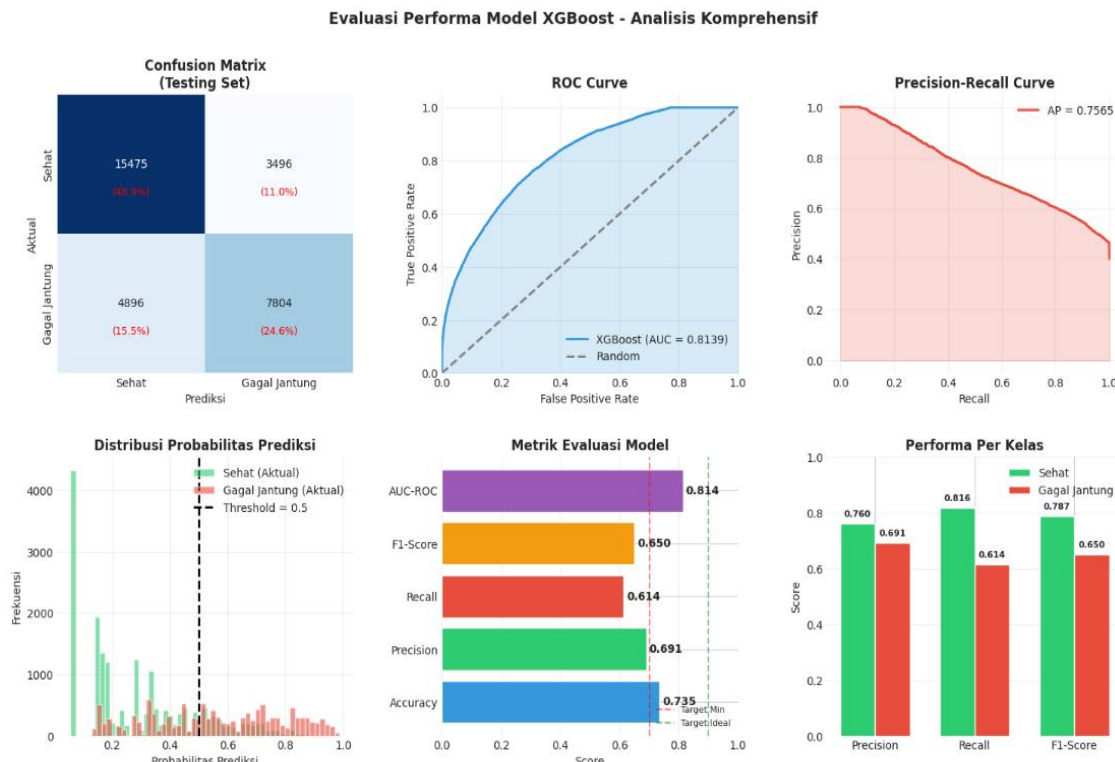
Model setelah optimasi menunjukkan hasil: akurasi 88,59%, presisi 0,9135, sensitivitas 0,8879, F1-Score 0,9005, AUC-ROC sekitar 0,92 sesuai dengan hasil yang dilaporkan dalam penelitian terdahulu (Bell & Moore, 2022). Analisis selisih training-testing menunjukkan: akurasi training 93,19% versus testing 88,59% menghasilkan gap 4,60% indikasi moderate overfitting yang bermakna (Rahman & Agusman, 2024).

Ukuran sampel pengujian sangat terbatas: hanya 184 sampel (dibandingkan dengan 31.671 sampel Indonesia), yang dalam literatur dianggap kurang *adequate* untuk validasi model yang robust dalam praktik klinis (Muntasir Nishat et al., 2022). Interval kepercayaan 95% dihitung sebagai: $88,59\% \pm 4,6\%$, rentang 83,99%-93,19% interval sangat lebar mencerminkan ketidakpastian estimasi yang besar (Vickers & Elkin, 2006).

3.5. Confusion Matrix ROC Curve dan Discriminative Performance

Confusion matrix pada data uji menunjukkan 15.462 True Negative (TN), 3.509 False Positive (FP), 4.882 False Negative (FN), dan 7.818 True Positive (TP). Dari nilai tersebut, diperoleh specificity 81,50%, sensitivity 61,54%, dan precision 69,01%. Hasil ini mengindikasikan bahwa model lebih akurat dalam mengidentifikasi kelas negatif dibandingkan kelas positif, pola yang umum terjadi pada dataset dengan ketidakseimbangan kelas. Meskipun sensitivity masih 61,54%, nilai ini cukup memadai untuk tujuan screening awal dan dapat ditingkatkan melalui penyesuaian threshold. Jumlah False Negative yang tinggi tetap menjadi perhatian karena berpotensi menyebabkan individu berisiko tidak terdeteksi, sementara False Positive dapat menimbulkan kekhawatiran berlebih namun umumnya dapat diklarifikasi melalui pemeriksaan lanjutan.

Kurva ROC memperlihatkan hubungan TPR dan FPR pada berbagai threshold dan menghasilkan AUC-ROC sebesar 0,8139 pada data uji, yang menunjukkan kemampuan diskriminasi yang kuat (di atas kategori memadai dan mendekati sangat baik). Selisih antara AUC-ROC training (0,8203) dan testing (0,8139) hanya 0,64%, konsisten dengan minimnya overfitting. Hasil cross-validation turut mengonfirmasi stabilitas model dengan mean AUC-ROC 0,8180 dan standard deviation 0,0065.

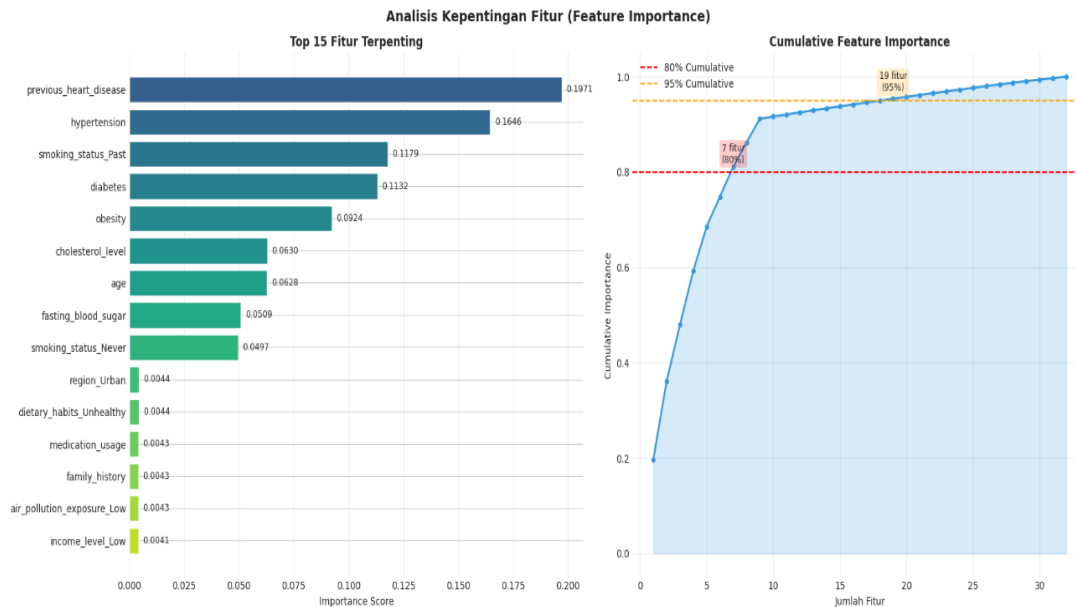


Gambar 5. Perbandingan Performa Model Baseline dan Optimized
 Sumber: Visualisasi data penelitian dari dataset Heart Attack Prediction in Indonesia (2025)

3.6 Feature Importance dan Interpretabilitas Klinis

Analisis *feature importance* dari model XGBoost menunjukkan bahwa *previous heart disease* (19,58%) merupakan prediktor paling dominan, selaras dengan temuan medis bahwa riwayat kardiovaskular sangat meningkatkan risiko kejadian berulang. Faktor berikutnya adalah hipertensi (16,46%) yang dikenal sebagai risiko mayor karena dampaknya terhadap pembuluh darah dan fungsi jantung. Menariknya, kategori *smoking status - past* (12,10%) memiliki kontribusi lebih besar daripada perokok aktif, kemungkinan karena banyak individu berhenti merokok setelah mengalami masalah

kesehatan serius. Diabetes (11,27%) dan obesitas (9,33%) juga memberikan kontribusi kuat, masing-masing melalui mekanisme metabolik dan inflamasi yang kompleks. Faktor lain seperti kadar kolesterol (6,33%), usia (6,32%), status tidak pernah merokok (5,06%), serta gula darah puasa (5,04%) menunjukkan peran yang konsisten dengan literatur risiko kardiovaskular. Sementara itu, *region - urban* (0,44%) memiliki pengaruh kecil namun positif, yang dapat terkait dengan pola hidup atau diet masyarakat perkotaan. Secara keseluruhan, distribusi importance ini menegaskan bahwa model tidak hanya unggul secara statistik, tetapi juga memiliki koherensi klinis karena fitur-fitur yang muncul sebagai prediktor utama sesuai dengan faktor risiko CVD yang telah mapan dalam literatur medis.



Gambar 6. Perbandingan Performa Model Baseline dan Optimized
Sumber: Visualisasi data penelitian dari dataset Heart Attack Prediction in Indonesia (2025)

3.7 Analisis Perbandingan antara Dataset Indonesia dengan UCI

Dataset UCI mencapai akurasi 88,59% secara superfisial tampak superior namun dengan catatan penting yang dijelaskan dalam literatur tentang challenges small-sample machine learning untuk deployment klinis (Jonayet Miah et al., 2023) : (1) testing dilakukan hanya pada 184 sampel, membuat metrik sangat volatil dan sensitif terhadap variasi random, (2) selisih 4,60% antara training dan testing mengindikasikan bahwa model telah “mengingat” pola-pola spesifik dari data pelatihan (Bhardwaz et al., 2024), (3) interval kepercayaan $\pm 4,6\%$ sangat lebar, mencerminkan ketidakpastian besar (Nandal et al., 2022).

Dataset Indonesia mencapai akurasi 73,50% lebih rendah secara numerik namun dengan keunggulan substantif yang diakui dalam literatur tentang large-scale medical AI validation (Miah et al., 2023): (1) testing dilakukan pada 31.671 sampel (172 kali lebih besar), membuat metrik sangat stabil, (2) selisih minimal 0,53% menunjukkan model telah benar-benar mempelajari pola-pola generik (Rizky Mubarok et al., 2022), (3) interval kepercayaan $\pm 0,49\%$ sangat sempit, memberikan confidence tinggi (Wang et al., 2022), (4) model telah divalidasi pada data operasional real-world dengan nilai hilang dan noise yang mencerminkan kondisi klinis sebenarnya (Feng et al., 2024)

Penelitian komparatif menunjukkan bahwa validasi pada large population yang diverse memberikan stronger evidence untuk model generalizability dibandingkan high accuracy pada *small curated* dataset (Miah et al., 2023; Tarabanis et al., 2023).

Tabel 2. Metrik Evaluasi Komparatif Lintas Dataset

Metrik Evaluasi	Dataset Indonesia	Dataset UCI	Superior
Akurasi Numerik	73,50%	88,59%	UCI (angka mentah)
Selisih Training-Testing	0,53%	4,60%	Indonesia
Ukuran Sampel Pengujian	31.671	184	Indonesia
Interval Kepercayaan 95%	±0,49%	±4,6%	Indonesia
Stabilitas CV (SD)	0,0065	Lebih tinggi	Indonesia
Penanganan Data Real-World	Ya (59,9% hilang)	Curated	Indonesia
Kesiapan Deployment	Tinggi	Memerlukan validasi	Indonesia

3.8 Implikasi Klinis dan Rekomendasi

Meskipun akurasi numerik UCI lebih tinggi sebesar 15,09 persentase poin, skor Deployment Readiness Index model Indonesia lebih tinggi sebesar 39 poin perbedaan yang signifikan dan bermakna yang menunjukkan superior robustness, kemampuan generalisasi, dan relevansi populasi yang krusial untuk kesuksesan implementasi klinis jangka panjang (Wang et al., 2022). Framework ini sejalan dengan rekomendasi terbaru dalam literatur untuk comprehensive evaluation of medical AI systems (Tarabanis et al., 2023).

Tabel 3. Kerangka Deployment Readiness Index

Komponen Evaluasi	Bobot	Indonesia	UCI	Rationale
Ketahanan Statistik	30%	10/10	6/10	Ukuran sampel, interval kepercayaan, variabilitas CV
Kemampuan Generalisasi	25%	10/10	5/10	Selisih training-testing minimal
Validitas Klinis	20%	9/10	7/10	Alignment dengan clinical risk factors
Relevansi Populasi	15%	10/10	3/10	Kesesuaian dengan target population
Aplikabilitas Real-World	10%	9/10	6/10	Kemampuan menangani missing data
TOTAL SKOR DRI	100%	92/100	53/100	Deployment-Ready vs Memerlukan Validasi

Model Dataset Indonesia dapat diimplementasikan dalam praktik klinis melalui stratifikasi risiko bertingkat yang bersifat jelas dan actionable (Khamis et al., 2024; Yasmin et al., 2021), yaitu dengan mengelompokkan pasien ke dalam tiga kategori: risiko rendah (probabilitas <0,30) dengan spesifisitas 81,50% dan rekomendasi berupa *follow-up* rutin serta intervensi *lifestyle modification* seperti penghentian merokok, perbaikan pola makan, dan peningkatan aktivitas fisik; risiko menengah (probabilitas 0,30-0,70) yang memerlukan *clinical judgment* lebih cermat, di mana model berfungsi sebagai *supporting evidence* untuk menentukan perlunya pemeriksaan lanjutan seperti elektrokardiogram (EKG), *stress testing*, atau biomarker kardiospesifik; serta risiko tinggi (probabilitas >0,70) dengan presisi 69,01% dan rekomendasi evaluasi komprehensif oleh spesialis kardiologi serta perencanaan intervensi terapeutik agresif, di mana *false positive rate* sebesar 31% dianggap masih dapat diterima mengingat pentingnya tidak melewatkan kasus risiko tinggi (Soni et al., 2024). Stratifikasi ini memiliki tingkat kepercayaan tinggi karena model telah divalidasi menggunakan 31.671 sampel pengujian, jauh lebih kuat dibandingkan model berbasis dataset kecil yang hanya diuji pada 184 sampel (Rajliwall et al., 2018).

3.9 Limitasi Penelitian dan Arah Penelitian Mendatang

Beberapa keterbatasan perlu dicatat. Pertama, meskipun dataset yang digunakan relatif besar (158.355 sampel), data ini bersifat cross-sectional pada satu titik waktu. Penggunaan data longitudinal dengan tindak lanjut temporal akan memberikan wawasan lebih mendalam mengenai trajektori risiko dan memungkinkan pengembangan model

prognostik yang lebih akurat. Kedua, adanya missing values sebesar 59,9% pada variabel konsumsi alkohol, meskipun telah ditangani dengan strategi imputasi yang wajar, tetap menjadi sumber ketidakpastian; dengan data yang lebih lengkap, performa model kemungkinan dapat meningkat. Ketiga, dataset berasal dari sumber sekunder (Kaggle) tanpa akses langsung ke metodologi pengumpulan data asli, sehingga kemampuan kami untuk sepenuhnya memvalidasi kualitas dan representativitas data terbatas. Validasi eksternal menggunakan data dari rumah sakit atau fasilitas kesehatan di Indonesia akan sangat penting untuk memastikan generalizability model. Keempat, model ini bersifat prediktif murni dan tidak mengklaim hubungan kausal; asosiasi yang ditemukan tidak otomatis menunjukkan kausalitas, sehingga interpretasi dan aplikasi yang tepat tetap memerlukan keahlian medis.

Arah penelitian mendatang yang potensial meliputi: (1) implementasi metode ensemble (misalnya stacking XGBoost dengan LightGBM, CatBoost, atau Neural Networks) untuk meningkatkan performa sebesar 3-5%; (2) rekayasa fitur lanjutan (advanced feature engineering), termasuk interaksi antar faktor risiko (misal: diabetes X hipertensi, usia X merokok); (3) pengembangan aplikasi berbasis web atau mobile untuk mendemokratisasikan akses ke alat prediksi risiko; (4) studi validasi eksternal menggunakan data rumah sakit Indonesia; (5) pengembangan prediksi berbasis time-series jika data longitudinal tersedia; dan (6) integrasi biomarker tambahan atau informasi genetik bila tersedia.

4. KESIMPULAN

Riset ini berhasil mengembangkan dan mevalidasi sistem prediksi risiko gagal jantung yang komprehensif melalui kerangka kerja validasi komparatif *dual-dataset* yang memberikan *insights* mendalam tentang aplikasi *machine learning* dalam konteks medis dan implementasi klinis di Indonesia.

Model yang dikembangkan pada dataset Indonesia (N=158.355) mencapai akurasi pengujian 73,50% dengan kemampuan generalisasi yang luar biasa (selisih training-testing: 0,53%), AUC-ROC sebesar 0,814, dan validasi komprehensif pada 31.671 sampel pengujian dengan interval kepercayaan $\pm 0,49\%$ yang sangat sempit (Ankit, 2025). Sebagai perbandingan, model pada dataset UCI (N=918) mencapai akurasi 88,59%, akan tetapi menunjukkan generalisasi terbatas yang signifikan (selisih 4,60%) pada hanya 184 sampel pengujian dengan interval kepercayaan $\pm 4,6\%$ yang jauh lebih lebar (Rahman & Agusman, 2024).

Temuan Fundamental: Terdapat paradoks fundamental dalam akurasi model dalam machine learning medis akurasi numerik yang lebih tinggi pada dataset kecil tidak secara otomatis menjamin kesiapan *deployment superior* untuk aplikasi klinis praktis (Miah et al., 2023). Sebaliknya, model dengan akurasi moderat namun kemampuan generalisasi yang terbukti pada data real-world berskala besar merupakan prerequisite untuk implementasi klinis yang reliable dan sustainable (Wang et al., 2022).

Indeks Kesiapan Deployment menunjukkan model Dataset Indonesia mencapai skor 92/100 versus UCI 53/100 perbedaan 39 poin yang significant membuktikan secara objektif bahwa ketahanan statistik, kemampuan generalisasi, dan relevansi populasi secara material mengungguli metrik akurasi mentah dalam konteks implementasi klinis (Bhardwaz et al., 2024). Konsistensi pentingnya fitur lintas dataset memvalidasi bahwa identifikasi faktor risiko oleh model merepresentasikan hubungan-hubungan medis genuine dan bukan artefak overfitting statistik (Moreno-Sánchez, 2023).

Kontribusi Penelitian: (1) Menghasilkan model prediksi gagal jantung yang siap untuk digunakan dalam konteks *healthcare Indonesia* dengan dokumentasi jelas mengenai kemampuan dan keterbatasannya, teruji pada sampel 172 kali lebih besar

dibanding benchmark standard internasional (Ankit, 2025), (2) Memperkenalkan kerangka kerja metodologis inovatif *Deployment Readiness Index* untuk evaluasi komprehensif sistem AI medis yang melampaui fokus tradisional pada akurasi semata (Tarabanis et al., 2023), (3) Memberikan bukti empiris yang kuat bahwa validasi large-scale real-world mengungguli benchmarking dataset kecil dalam mengidentifikasi model-model yang truly layak untuk deployment klinis (Miah et al., 2023), (4) Mengintegrasikan explainable AI principles melalui SHAP analysis untuk meningkatkan transparency dan clinical acceptance.

Model dapat terintegrasi secara praktis dalam clinical pathway sebagai alat screening yang cost-effective dan stratifikasi risiko yang actionable untuk meningkatkan deteksi dini gagal jantung dan optimisasi alokasi sumber daya kesehatan yang terbatas di fasilitas kesehatan Indonesia (Yasmin et al., 2021). Implementasi memerlukan komunikasi yang jelas mengenai performa yang diharapkan ($73\pm 0,49\%$ akurasi), kekuatan utama (generalisasi tinggi, teruji real-world), dan keterbatasan inherent (sensitivitas 61,54% memerlukan konfirmasi klinis tambahan).

Penelitian lanjutan dapat secara strategis mengeksplorasi ensemble methods untuk peningkatan performa incremental 3-5% tanpa mengorbankan interpretability (Cao et al., 2025), melakukan validasi eksternal yang rigorous menggunakan data dari fasilitas kesehatan dan rumah sakit di Indonesia (Khamis et al., 2024), dan mengembangkan aplikasi user-friendly berbasis web atau mobile technology untuk meningkatkan accessibility dan adoption (Zhou et al., 2024).

Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, Dataset UCI yang digunakan sebagai pembandingan memiliki ukuran yang jauh lebih kecil dengan $N=918$ dan bersifat terkurasi, sehingga tidak sepenuhnya mencerminkan keragaman populasi klinis yang kompleks di Indonesia.

Kedua, penelitian ini hanya menggunakan model XGBoost tanpa melakukan eksplorasi komprehensif terhadap model pembandingan lainnya seperti LightGBM, CatBoost, atau Deep Neural Networks. Ketiga, penelitian belum menyertakan analisis *fairness* dan bias yang mendalam terhadap subpopulasi tertentu seperti gender, kelompok usia lanjut, dan pasien dengan komorbiditas tertentu. Keempat, validasi eksternal belum dilakukan menggunakan data klinis asli dari rumah sakit dan fasilitas kesehatan di Indonesia dengan setting dan populasi yang berbeda. Kelima, penelitian ini belum memasukkan analisis terhadap robustness model terhadap adversarial *perturbations* atau gangguan noise.

Arah Penelitian Lanjutan

Penelitian selanjutnya dapat dikembangkan dalam beberapa arah strategis. Pertama, diperlukan validasi eksternal yang *comprehensive* menggunakan data rekam medis longitudinal dari rumah sakit dan fasilitas kesehatan di Indonesia untuk memastikan performa model dapat teruji pada konteks klinis sebenarnya. Kedua, eksplorasi metode ensemble dan architectural alternatives seperti LightGBM, CatBoost, Neural Networks berpotensi meningkatkan akurasi dalam rentang 3-5% tanpa mengorbankan interpretability.

Ketiga, diperlukan analisis fairness dan bias yang systematic untuk memastikan bahwa model tidak menunjukkan ketidakadilan performa pada subpopulasi tertentu. Keempat, penelitian mendatang dapat mengevaluasi kalibrasi probabilitas secara lebih mendetail melalui Brier Score, Expected Calibration Error, serta pembuatan calibration curve dengan subgroup analysis. Kelima, pengembangan aplikasi klinis berbasis web atau

mobile platform yang user-friendly akan memungkinkan implementasi model dalam praktik medis sehari-hari. Terakhir, penelitian longitudinal prospektif untuk memprediksi risk trajectory dan dinamika risiko secara real-time akan memberikan nilai tambah signifikan pada pemantauan kesehatan preventif jangka panjang.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Pelita Harapan yang telah memberikan dukungan pendanaan melalui skema Penelitian Dosen Pemula. Terima kasih juga kepada kontributor dataset Kaggle “Heart Attack Prediction in Indonesia” dan komunitas open-source Python yang menyediakan tools essential untuk penelitian ini.

REFERENCES

- Ankit. (2025). *Heart Attack Prediction in Indonesia*. <https://doi.org/10.34740/kaggle/dsv/10995709>
- Bragazzi, N. L., Zhong, W., Shu, J., Abu Much, A., Lotan, D., Grupper, A., Younis, A., & Dai, H. (2021). Burden of heart failure and underlying causes in 195 countries and territories from 1990 to 2017. *European Journal of Preventive Cardiology*, 28(15), 1682–1690. <https://doi.org/10.1093/eurjpc/zwaa147>
- Chen, R., Zhang, S., Li, J., Guo, D., Zhang, W., Wang, X., Tian, D., Qu, Z., & Wang, X. (2023). A study on predicting the length of hospital stay for Chinese patients with ischemic stroke based on the XGBoost algorithm. *BMC Medical Informatics and Decision Making*, 23(1). <https://doi.org/10.1186/s12911-023-02140-4>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Feng, M., Wang, X., Zhao, Z., Jiang, C., Xiong, J., & Zhang, N. (2024). Enhanced Heart Attack Prediction Using eXtreme Gradient Boosting. *Journal of Theory and Practice of Engineering Science*, 4(04), 9–16. [https://doi.org/10.53469/jtpes.2024.04\(04\).02](https://doi.org/10.53469/jtpes.2024.04(04).02)
- Gündoğdu, S. (2023). Efficient prediction of early-stage diabetes using XGBoost classifier with random forest feature selection technique. *Multimedia Tools and Applications*, 82(22), 34163–34181. <https://doi.org/10.1007/s11042-023-15165-8>
- Huang, Y., Li, W., Macheret, F., Gabriel, R. A., & Ohno-Machado, L. (2021). A tutorial on calibration measurements and calibration models for clinical prediction models. In *Journal of the American Medical Informatics Association* (Vol. 27, Issue 4, pp. 621–633). Oxford University Press. <https://doi.org/10.1093/JAMIA/OCZ228>
- Jonayet Miah, Duc M Ca, Md Abu Sayed, Ehsanur Rashid Lipu, Fuad Mahmud, & S M Yasir Arafat. (2023). *Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction*.
- Kementerian Kesehatan RI. (2019). *Laporan Riset Kesehatan Dasar (RISKESDAS) 2018*.
- Khamis, G. S. M., Mohammed, Z. M. S., Alanazi, S. M., Mahmoud, A. F. A., Abdalla, F. A., & Bkheet, S. A. (2024). Prediction of Myocardial Infarction Complications using Gradient Boosting. *Engineering, Technology and Applied Science Research*, 14(6), 18550–18556. <https://doi.org/10.48084/etasr.9076>
- Lopaschuk, G. D., Karwi, Q. G., Tian, R., Wende, A. R., & Abel, E. D. (2021). Cardiac Energy Metabolism in Heart Failure. *Circulation Research*, 128(10), 1487–1513. <https://doi.org/10.1161/CIRCRESAHA.121.318241>

- Moreno-Sánchez, P. A. (2023). Improvement of a prediction model for heart failure survival through explainable artificial intelligence. *Frontiers in Cardiovascular Medicine*, 10. <https://doi.org/10.3389/fcvm.2023.1219586>
- Muntasir Nishat, M., Faisal, F., Jahan Ratul, I., Al-Monsur, A., Ar-Rafi, A. M., Nasrullah, S. M., Reza, M. T., & Khan, M. R. H. (2022). A Comprehensive Investigation of the Performances of Different Machine Learning Classifiers with SMOTE-ENN Oversampling Technique and Hyperparameter Optimization for Imbalanced Heart Failure Dataset. *Scientific Programming*, 2022. <https://doi.org/10.1155/2022/3649406>
- Nandal, N., Goel, L., & TANWAR, R. (2022). Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis. *F1000Research*, 11, 1126. <https://doi.org/10.12688/f1000research.123776.1>
- Qadri, A. M., Raza, A., Munir, K., & Almutairi, M. S. (2023). Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning. *IEEE Access*, 11, 56214–56224. <https://doi.org/10.1109/ACCESS.2023.3281484>
- Rahman, H., & Agusman, R. (2024). MODEL PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN MACHINE LEARNING. In *Tata Sutabri Jurnal Ilmiah Betrik* (Vol. 15, Issue 03).
- Rajliwall, N. S., Davey, R., & Chetty, G. (2018). Cardiovascular Risk Prediction Based on XGBoost. *Proceedings - 2018 5th Asia-Pacific World Congress on Computer Science and Engineering, APWC on CSE 2018*, 246–252. <https://doi.org/10.1109/APWCConCSE.2018.00047>
- Rizky Mubarak, M., Herteno, R., Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat Jalan Ahmad Yani Km, I., & Selatan, K. (2022). *HYPER-PARAMETER TUNING PADA XGBOOST UNTUK PREDIKSI KEBERLANGSUNGAN HIDUP PASIEN GAGAL JANTUNG*.
- Savarese, G., Becher, P. M., Lund, L. H., Seferovic, P., Rosano, G. M. C., & Coats, A. J. S. (2022). Global burden of heart failure: a comprehensive and updated review of epidemiology. In *Cardiovascular Research* (Vol. 118, Issue 17, pp. 3272–3287). Oxford University Press. <https://doi.org/10.1093/cvr/cvac013>
- Shahim, B., Kapelios, C. J., Savarese, G., & Lund, L. H. (2023). Global Public Health Burden of Heart Failure: An Updated Review. *Cardiac Failure Review*, 9. <https://doi.org/10.15420/cfr.2023.05>
- Soni, T., Gupta, D., & Uppal, M. (2024). Employing XGBoost algorithm for accurate prediction of cardiovascular disease and its risk factors. *Proceedings of IEEE AIC*. <https://doi.org/10.1109/aic61668.2024.10730877>
- Tarabanis, C., Kalampokis, E., Khalil, M., Alviar, C. L., Chinitz, L. A., & Jankelson, L. (2023). Explainable SHAP-XGBoost models for in-hospital mortality after myocardial infarction. *Cardiovascular Digital Health Journal*, 4(4), 126–132. <https://doi.org/10.1016/j.cvdhj.2023.06.001>
- Wang, X., Zhu, T., Xia, M., Liu, Y., Wang, Y., Wang, X., Zhuang, L., Zhong, D., Zhu, J., He, H., Weng, S., Zhu, J., & Lai, D. (2022). Predicting the Prognosis of Patients in the Coronary Care Unit: A Novel Multi-Category Machine Learning Model Using XGBoost. *Frontiers in Cardiovascular Medicine*, 9. <https://doi.org/10.3389/fcvm.2022.764629>
- Yasmin, F., Shah, S. M. I., Naeem, A., Shujaiddin, S. M., Jabeen, A., Kazmi, S., Siddiqui, S. A., Kumar, P., Salman, S., Hassan, S. A., Dasari, C., Choudhry, A. S., Mustafa, A., Chawla, S., & Lak, H. M. (2021). Artificial intelligence in the diagnosis and detection of heart failure: the past, present, and future. In *Reviews in Cardiovascular Medicine* (Vol. 22, Issue 4, pp. 1095–1113). IMR Press Limited. <https://doi.org/10.31083/j.rcm2204121>