JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis Komparatif Model Random Forest dan XGBoost untuk Klasifikasi Penyakit Jantung Berbasis Data Klinis

Angga Guardi Zunus Saputra*, Fikri Budiman

Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1*]111202214680@mhs.dinus.ac.id, [2]fikri.budiman@dsn.dinus.ac.id
Email Penulis Korespondensi: 111202214680@mhs.dinus.ac.id
Submitted **19-12-2025;** Accepted **15-02-2026;** Published **28-02-2026**

Abstrak

Penyakit jantung masih menjadi tantangan serius bagi sektor kesehatan global seiring meningkatnya prevalensi kasus, sehingga diperlukan sistem diagnosis dini yang akurat dan andal. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Random Forest (RF) dan XGBoost dalam klasifikasi penyakit jantung, serta mengidentifikasi model yang paling sesuai untuk mendukung pengambilan keputusan klinis berbasis data. Studi ini mengatasi research gap dalam kajian ensemble learning dengan melakukan evaluasi komparatif yang komprehensif menggunakan UCI Heart Disease Dataset. Metodologi penelitian meliputi tahapan pra-pemrosesan data, feature encoding, normalisasi, penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Oversampling Technique (SMOTE), serta optimasi hyperparameter berbasis RandomizedSearchCV. Evaluasi kinerja dilakukan menggunakan metrik accuracy, precision, recall, F1-score, ROC-AUC, dan Matthews Correlation Coefficient (MCC), yang diperkuat dengan analisis Feature Importance. Hasil penelitian menunjukkan bahwa kedua model ensemble memiliki performa yang tinggi dengan nilai F1-score konsisten di atas 0,88. XGBoost menunjukkan performa agregat terbaik dengan F1-score tertinggi sebesar 0,8995 dan Precision sebesar 0,8785, sehingga lebih efektif dalam meminimalkan kesalahan False Positive. Sebaliknya, Random Forest unggul dalam sensitivitas dengan Recall tertinggi sebesar 0,9510 dan ROC-AUC sebesar 0,9582, serta stabilitas cross-validation yang lebih baik. Temuan ini menegaskan bahwa pemilihan algoritma klasifikasi penyakit jantung perlu disesuaikan dengan tujuan klinis, sehingga hasil penelitian ini diharapkan dapat memberikan kontribusi praktis dalam pengembangan sistem pendukung keputusan medis berbasis machine learning.

Kata Kunci: Random Forest; XGBoost; Klasifikasi Penyakit Jantung; Feature Importance; SMOTE

Abstract

Heart disease remains a major global health challenge due to its increasing prevalence, highlighting the need for accurate and reliable early diagnostic systems. This study aims to analyze and compare the performance of Random Forest (RF) and XGBoost algorithms for heart disease classification, and to identify the most suitable model for data-driven clinical decision support. Addressing a research gap in ensemble learning studies, this research conducts a comprehensive comparative evaluation using the UCI Heart Disease Dataset. The proposed methodology includes data preprocessing, feature encoding, normalization, class imbalance handling using the Synthetic Minority Oversampling Technique (SMOTE), and hyperparameter optimization based on RandomizedSearchCV. Model performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and Matthews Correlation Coefficient (MCC), supported by Feature Importance analysis. The results demonstrate that both ensemble models achieve strong predictive performance, with consistently high F1-scores above 0.88. XGBoost exhibits superior overall performance, achieving the highest F1-score of 0.8995 and Precision of 0.8785, making it more effective in minimizing False Positive predictions. In contrast, Random Forest shows superior sensitivity, with the highest Recall of 0.9510 and ROC-AUC of 0.9582, along with better cross-validation stability. These findings indicate that the selection of heart disease classification algorithms should be aligned with specific clinical objectives, and the results of this study are expected to contribute to the development of effective machine learning–based clinical decision support systems.

Keywords: Random Forest; XGBoost; Heart Disease Classification; Feature Importance; SMOTE

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab kematian terbesar di dunia dan masih menjadi tantangan serius bagi sektor kesehatan global [1], [2]. Laporan terbaru *World Health Organization* (WHO) tahun 2024 menunjukkan bahwa lebih dari 18 juta jiwa meninggal setiap tahunnya akibat penyakit kardiovaskular, yang mewakili sekitar 32% dari seluruh kematian global [3]. Kondisi tersebut juga tercermin di Indonesia, di mana hasil Riset Kesehatan Dasar (*Riskesdas*) 2018 menunjukkan prevalensi penyakit jantung sebesar 1,5%, sementara Survei Kesehatan Indonesia (*SKI*) 2023 mencatat prevalensi penyakit jantung sebesar 0,85% berdasarkan diagnosis dokter [4]. Peningkatan ini tidak hanya terjadi pada kelompok lanjut usia, tetapi juga pada usia produktif akibat perubahan gaya hidup, konsumsi makanan tinggi lemak, dan tingginya tingkat stres [5]. Kompleksitas faktor risiko—termasuk hipertensi, kolesterol, obesitas, diabetes, kebiasaan merokok, dan *gaya hidup sedentari*—menjadikan proses deteksi dini penyakit jantung semakin menantang bagi tenaga medis [6].

Identifikasi dini penyakit jantung sangat penting untuk menekan risiko komplikasi dan kematian. Namun, keterbatasan kapasitas diagnostik konvensional membuat kebutuhan akan sistem yang mampu melakukan analisis cepat, akurat, dan berbasis data semakin mendesak [7]. Dalam konteks tersebut, perkembangan *Artificial Intelligence* (AI) dan *Machine Learning* (ML) menawarkan solusi potensial untuk mendukung diagnosis berbasis data klinis pasien, terutama melalui pemodelan berbasis data tabular seperti *UCI Heart Disease Dataset* [8]. Algoritma ML mampu mengidentifikasi pola kompleks dalam variabel klinis seperti usia, tekanan darah, kolesterol, denyut jantung, dan hasil pemeriksaan elektrokardiografi sehingga membantu meningkatkan akurasi prediksi risiko penyakit jantung [9].

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berbagai penelitian sebelumnya telah mengevaluasi performa algoritma ML dalam prediksi penyakit jantung. Random Forest (*RF*), misalnya, dikenal efektif dalam menangani data multivariabel serta mampu memberikan performa yang stabil karena mekanisme *bagging* yang mengurangi risiko *overfitting* [10]. Pada *dataset* UCI, *Random Forest* menghasilkan akurasi 0,80, namun performanya menurun pada data tidak seimbang [11]. Untuk mengatasi ketidakseimbangan data tersebut, penelitian lain menunjukkan bahwa integrasi *Synthetic Minority Oversampling Technique* (SMOTE) berhasil meningkatkan akurasi model meskipun demikian, metode ini masih menghadapi kendala pada kestabilan *cross-validation* [12].

Pada sisi lain, algoritma XGBoost (*Extreme Gradient Boosting*) semakin populer karena kemampuannya mengelola data tabular secara efisien melalui optimasi berbasis *gradient boosting* [13], [14]. Studi terdahulu menunjukkan bahwa *XGBoost* mampu mencapai *F1-score* 0,87 pada *UCI Heart Dataset*, namun performanya terbukti menurun ketika variabel kategori tidak di-*encode* secara konsisten [15]. Selain itu, penelitian lain mencatat bahwa *XGBoost* memiliki sensitivitas yang lebih baik dibanding Random Forest, tetapi model ini tetap menunjukkan kecenderungan *overfitting* pada data dengan distribusi tidak seimbang [16]. Secara umum, sebagian besar penelitian dalam lima tahun terakhir menempatkan performa model ML untuk klasifikasi penyakit jantung pada kisaran *F1-score* 0,85–0.90, yang mengindikasikan perlunya pendekatan metodologis yang mampu menghasilkan performa lebih tinggi dan lebih stabil[17], [18], [19].

Meskipun sejumlah studi telah membandingkan performa algoritma ML pada dataset penyakit jantung, terdapat beberapa keterbatasan yang masih belum banyak dibahas. Pertama, sebagian penelitian tidak mengintegrasikan proses *hyperparameter tuning* secara komprehensif, sehingga performa model belum optimal [20]. Kedua, penggunaan SMOTE sebagai metode penyeimbang data jarang divalidasi melalui visualisasi distribusi data sebelum dan sesudah *oversampling*, sehingga kurang memberikan gambaran yang transparan mengenai kualitas data latih [21]. Selain itu, penelitian sebelumnya masih berfokus pada satu model tertentu tanpa melakukan evaluasi komparatif yang mendalam terhadap dua model *ensemble* paling dominan, yaitu Random Forest dan XGBoost [22]. Kondisi tersebut menunjukkan adanya *research gap* berupa kebutuhan akan studi komparatif yang terstandar, transparan, dan evaluatif.

Penelitian ini berkontribusi dalam bidang informatika kesehatan melalui pengembangan kerangka kerja klasifikasi penyakit jantung yang lebih akurat dan transparan. Kontribusi utama penelitian terletak pada penerapan metodologi terstandar yang mengintegrasikan visualisasi distribusi data sebelum dan sesudah penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Oversampling Technique (SMOTE), serta penggunaan hyperparameter tuning berbasis RandomizedSearchCV untuk memastikan komparasi yang adil antara algoritma Random Forest dan XGBoost. Berdasarkan celah penelitian (research gap) pada studi komparatif ensemble learning sebelumnya, penelitian ini bertujuan untuk melakukan analisis komparatif yang komprehensif terhadap kedua algoritma menggunakan UCI Heart Disease Dataset. Metodologi yang diterapkan mencakup tahapan data preprocessing, feature encoding, normalisasi, penyeimbangan kelas, serta evaluasi kinerja model menggunakan metrik akurasi, precision, recall, F1-score, ROC-AUC, dan Matthews Correlation Coefficient (MCC), yang dilengkapi dengan validasi silang untuk menilai stabilitas model. Selain itu, analisis feature importance digunakan untuk meningkatkan interpretabilitas model. Dengan pendekatan ini, penelitian diharapkan mampu menghasilkan model klasifikasi penyakit jantung yang lebih andal, stabil, dan transparan sebagai dasar pengembangan sistem pendukung keputusan klinis berbasis data.

2. METODOLOGI PENELITIAN

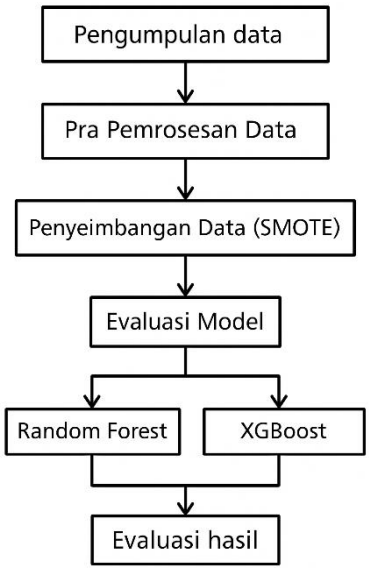
2.1 Tahap Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif komparatif untuk mengevaluasi kinerja dua algoritma *ensemble learning* berbasis *decision tree*, yaitu Random Forest (*RF*) dan XGBoost, dalam kasus klasifikasi penyakit jantung menggunakan *UCI Heart Disease Dataset*. Pemilihan *RF* didasarkan pada kemampuannya mengatasi hubungan non-linier dan menjaga stabilitas prediksi melalui mekanisme *bagging* [10], sementara *XGBoost* dipilih berkat optimasi *gradient boosting* yang efisien serta kinerja unggul yang telah terbukti pada data tabular medis [13], [14].

Pemilihan kedua algoritma tersebut juga mempertimbangkan karakteristik data medis yang bersifat heterogen, *non-linear*, serta mengandung interaksi kompleks antar *fitur*. Kedua *model ensemble* ini memanfaatkan struktur pohon keputusan untuk menangkap pola hubungan yang kompleks pada data tabular. *Random Forest* menawarkan stabilitas prediksi melalui agregasi banyak pohon keputusan, sedangkan *XGBoost* mengoptimalkan proses pembelajaran secara bertahap melalui mekanisme *boosting* yang terkontrol. Pendekatan ini memungkinkan analisis komparatif yang sistematis antara dua strategi *ensemble* utama, yaitu *bagging* dan *boosting*, dalam konteks klasifikasi penyakit jantung.

Selain itu, pemilihan kedua model *ensemble* ini juga didorong oleh adanya *research gap* dari studi terdahulu yang mengindikasikan perlunya implementasi *hyperparameter tuning* yang komprehensif [20] dan analisis perbandingan yang lebih mendalam serta terstandar antara *RF* dan *XGBoost* [22]. Oleh karena itu, penelitian ini mengimplementasikan alur kerja terstruktur yang holistik, dimulai dari pemahaman data, *pra-pemrosesan*, penyeimbangan kelas menggunakan SMOTE, pembangunan model dengan *hyperparameter tuning* berbasis *RandomizedSearchCV*, hingga tahapan evaluasi performa menggunakan metrik prediktif terukur, sebagaimana alur metodologi ini ditunjukkan secara rinci pada Gambar 1.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 1. Alur Penelitian

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari *UCI Heart Disease Dataset*, sebuah *dataset* standar yang telah diterima luas dalam penelitian diagnosis penyakit jantung berbasis *Machine Learning* [8]. *Dataset* ini mencakup 920 sampel dengan 16 *feature* yang merepresentasikan karakteristik klinis pasien. *Feature* yang tersedia mencakup kombinasi data numerik, seperti usia, tekanan darah istirahat (*trestbps*), kadar kolesterol (*chol*), denyut jantung maksimum (*thalch*), dan nilai *oldpeak*, serta data kategorikal seperti jenis kelamin dan tipe nyeri dada. *Dataset* ini juga memuat variabel target (*num*) yang bersifat biner, mengindikasikan ada atau tidaknya penyakit jantung pada pasien. Karena terdapat kombinasi atribut numerik dan kategorikal, *dataset* ini memerlukan tahapan *pra-pemrosesan* (meliputi *imputasi* nilai hilang, *encoding*, dan normalisasi) agar dapat diproses secara optimal oleh model *Machine Learning*. Pemilihan *dataset* ini didasarkan pada kelengkapan dan keragaman atribut klinisnya, serta relevansinya yang tinggi dengan berbagai penelitian terdahulu dalam bidang prediksi penyakit jantung [8].

2.3 Pra Pemrosesan Data

Tahap *pra-pemrosesan* data merupakan langkah krusial untuk memastikan kualitas dan konsistensi data sebelum proses pelatihan model, mengingat bahwa performa algoritma *Machine Learning* sangat dipengaruhi oleh kebersihan data yang digunakan [8], [9]. Proses ini dimulai dengan Identifikasi Tipe *Feature*, di mana 16 atribut diklasifikasikan menjadi *feature* numerik (*age*, *trestbps*, *chol*, *thalch*, *oldpeak*, dan *ca*) dan *feature* kategorikal (*sex*, *dataset*, *cp*, *fbs*, *restecg*, *exang*, *slope*, dan *thal*). Beberapa *feature* numerik dengan kategori terbatas dikonversi menjadi kategorikal untuk memastikan model membaca atribut sesuai sifat aslinya [9], [15]. Selanjutnya, dilakukan Penanganan Nilai Hilang (*missing values*) yang teridentifikasi pada beberapa atribut kunci; untuk menjaga integritas data dan mempertahankan distribusi aslinya, digunakan *SimpleImputer* dengan pendekatan median untuk *feature* numerik dan modus (*most frequent*) untuk *feature* kategorikal, strategi yang *robust* terhadap *outlier* dan penting dalam pemodelan data medis [11], [13]. Setelah itu, *feature* numerik dinormalisasi menggunakan *StandardScaler*; standarisasi ini diperlukan agar model berbasis *gradient boosting* seperti XGBoost dapat bekerja lebih stabil, meskipun model berbasis *tree* kurang sensitif terhadap skala [13], [14]. Terakhir, dilakukan *Encoding feature* kategorikal menggunakan metode *One-Hot Encoding* (OHE) dengan skema *drop-first* untuk menghindari risiko *multikolinearitas* dan memungkinkan model menangkap variasi kategori secara tepat [9], [16]. Seluruh proses *pra-pemrosesan* ini diintegrasikan ke dalam satu alur kerja (*pipeline*) untuk memastikan penerapan yang otomatis dan konsisten.

2.4 Pembagian Data

Setelah seluruh tahapan pra-pemrosesan data dilakukan, *dataset* dibagi menjadi data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20 menggunakan metode *stratified train-test split* untuk menjaga proporsi kelas pada variabel target tetap konsisten. Pembagian ini bertujuan menyediakan data latih yang memadai untuk proses pelatihan dan optimasi *model*, sekaligus mempertahankan data uji yang independen guna mengevaluasi kemampuan generalisasi *model*. Seluruh proses penyeimbangan data menggunakan *Synthetic Minority Oversampling Technique* (*SMOTE*), pelatihan *model*, serta optimasi *hyperparameter* hanya diterapkan pada data latih, sedangkan data uji digunakan secara eksklusif untuk evaluasi akhir, sehingga risiko *data leakage* dapat dihindari.

2.5 Penyeimbangan Data (SMOTE)

Dataset yang digunakan pada penelitian ini memiliki ketidakseimbangan kelas, di mana terdapat 509 sampel berlabel 1 dan 411 sampel berlabel 0. Kondisi ini dapat menyebabkan model cenderung lebih “berpihak” pada kelas mayoritas

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

sehingga menurunkan kemampuan deteksi terhadap kelas minoritas, khususnya pada kasus penyakit jantung yang bersifat kritis. Untuk mengatasi masalah tersebut, penelitian ini menggunakan Synthetic Minority Oversampling Technique (SMOTE), yaitu metode oversampling berbasis interpolasi yang menghasilkan sampel sintetis baru pada kelas minoritas dengan memanfaatkan kedekatan antar-sampel pada ruang fitur [21]. Pada penelitian ini, SMOTE diterapkan hanya pada data latih setelah seluruh proses transformasi fitur (imputasi, standarisasi, dan encoding) selesai, sehingga mencegah terjadinya *data leakage* dan menjaga validitas evaluasi model. Pendekatan ini menghasilkan distribusi data latih yang lebih seimbang serta meningkatkan sensitivitas model dalam mengenali kasus positif penyakit jantung tanpa mengurangi kemampuan generalisasi.

2.6 Pembangunan Model

2.6.1 Random Forest

Random Forest digunakan sebagai model utama karena mampu menangani hubungan non-linear dan variabel multiskala melalui pendekatan *bagging*. Model membangun banyak *decision tree* yang dilatih pada subset data yang di-*bootstrap*, lalu hasil prediksi digabungkan menggunakan *majority voting*. Optimasi dilakukan dengan RandomizedSearchCV menggunakan ruang parameter n_estimators, max_depth, min_samples_split, min_samples_leaf, dan max_features, dengan pemilihan berbasis F1-score pada Stratified K-Fold.

$$\hat{y} = mode\{h1(x), h2(x), \ldots, hB(x)\} \quad (1)$$

Prediksi akhir $\hat{y}$ ditentukan dari kelas yang paling sering muncul (mode) berdasarkan prediksi seluruh pohon $h_1 \ldots h_B$. Pendekatan ini menurunkan varians dan mengurangi risiko overfitting.

2.6.2 XGBoost

XGBoost digunakan sebagai model pembanding karena efisiensinya dalam *gradient boosting* serta kemampuannya mengoreksi kesalahan prediksi secara bertahap. Setiap pohon baru dibangun untuk meminimalkan fungsi loss yang telah diregulasi. Optimasi dilakukan menggunakan RandomizedSearchCV dengan parameter seperti learning_rate, max_depth, min_child_weight, subsample, colsample_bytree, gamma, reg_lambda, reg_alpha, dan n_estimators.

$$Obj = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega\left(f_k\right) \quad (2)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (3)$$

Fungsi objektif XGBoost menggabungkan *loss function* dengan *regularization term* Ωuntuk mengontrol kompleksitas pohon (T) dan bobot daun (w_j). Regularisasi inilah yang membuat XGBoost lebih stabil dan tahan overfitting dibanding model boosting klasik.

2.7 Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja komparatif algoritma Random Forest dan XGBoost dalam klasifikasi penyakit jantung. Penelitian ini menggunakan enam metrik evaluasi utama: *Accuracy*, *Precision*, *Recall*, *F1-score*, *ROC-AUC*, dan *Matthews Correlation Coefficient* (MCC). Selain itu, untuk menilai stabilitas dan kemampuan generalisasi model pada berbagai subset data, diterapkan *10-Fold Stratified Cross-Validation*, yang hasilnya diukur melalui perolehan nilai *CV F1 Mean* dan *CV F1 STD*. Evaluasi akhir juga didukung oleh analisis *Confusion Matrix* dan *ROC Curve*. Kombinasi evaluasi ini bertujuan memberikan gambaran menyeluruh mengenai tingkat akurasi, konsistensi, dan kemampuan generalisasi kedua model.
Enam metrik evaluasi didefinisikan sebagai berikut.

a. Accuracy

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (4)$$

Accuracy mengukur sejauh mana model mampu memberikan prediksi yang benar terhadap seluruh data uji. Nilai ini memperhitungkan proporsi data yang berhasil diklasifikasikan dengan tepat, baik pasien yang benar-benar sakit maupun yang sehat. Meskipun umum digunakan, metrik ini dapat menjadi kurang representatif jika *dataset* tidak seimbang karena *FP* dan *FN* tidak terlihat secara proporsional.

b. Precision

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (5)$$

Precision menunjukkan tingkat ketepatan model dalam menghasilkan prediksi positif. Nilai ini mengukur seberapa banyak pasien yang diprediksi memiliki penyakit benar-benar terbukti positif. *Precision* penting terutama ketika kesalahan *FP* (*False Positive*) memiliki konsekuensi besar, misalnya ketika pasien sehat tidak boleh salah didiagnosis sebagai penderita.

c. Recall (Sensitivity)

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \quad (6)$$

Recall mengukur kemampuan model dalam mendeteksi semua kasus positif. Nilai ini menunjukkan seberapa banyak pasien yang benar-benar memiliki penyakit berhasil dikenali oleh model. *Recall* sangat penting dalam konteks medis, karena *FN* (*False Negative*) berarti pasien sakit tidak terdeteksi, yang dapat berdampak serius terhadap penanganan dan diagnosis.

d. F1-Score

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \quad (7)$$

F1-score memberikan ukuran performa yang seimbang antara *Precision* dan *Recall*. Metrik ini relevan ketika terdapat ketidakseimbangan kelas atau ketika penting untuk mempertimbangkan kesalahan *FP* dan *FN* secara bersamaan. Nilai *F1* yang tinggi menunjukkan bahwa model tidak hanya akurat dalam memprediksi positif tetapi juga mampu menemukan sebagian besar kasus positif.

e. ROC-AUC

$$TPR = \frac{TP}{\text{TP+FN}}, \text{FPR} = \frac{FP}{\text{FP+TN}} \quad (8)$$

ROC-AUC menilai kemampuan model dalam memisahkan kelas positif dan negatif pada berbagai ambang batas keputusan. *AUC* (*Area Under Curve*) yang tinggi menunjukkan bahwa model konsisten memberikan skor probabilitas yang lebih tinggi kepada pasien yang benar-benar memiliki penyakit dibandingkan dengan pasien sehat. Metrik ini memberikan gambaran performa model secara lebih menyeluruh daripada sekadar satu nilai ambang.

f. Matthews Correlation Coefficient (MCC)

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (9)$$

MCC memberikan evaluasi komprehensif terhadap performa model dengan mempertimbangkan seluruh komponen pada *confusion matrix*. Nilai *MCC* berada pada rentang -1 hingga 1, di mana 1 menunjukkan prediksi sempurna dan 0 menunjukkan performa acak. *MCC* sangat bermanfaat pada *dataset* tidak seimbang karena memberikan penilaian yang lebih adil terhadap kesalahan *FP* dan *FN* dibanding metrik konvensional seperti *Accuracy*.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah *Heart Disease UCI*, sebuah gabungan data klinis penyakit jantung yang umum digunakan dalam studi klasifikasi (*Machine Learning*). *Dataset* ini berisi 920 sampel data dengan 16 *feature* utama (termasuk *age*, *sex*, *cp*, *trestbps*, *chol*, dan *thalch*) yang merepresentasikan kondisi pasien, ditambah variabel target (*num*). *Dataset* ini diakses melalui: https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data. Berikut perincian *feature* dalam Tabel 1 di bawah ini.

Tabel 1. Sampel Dataset Penyakit Diabetes

id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
1	63	Male	Cleveland	typical angina	145	233	TRUE	lv hypertrophy	150	FALSE	2.3	downsloping	0	fixed defect	0
2	67	Male	Cleveland	asymptomatic	160	286	FALSE	lv hypertrophy	108	TRUE	1.5	flat	3	normal	2
3	67	Male	Cleveland	asymptomatic	120	229	FALSE	lv hypertrophy	129	TRUE	2.6	flat	2	reversable defect	1
4	37	Male	Cleveland	non-angina	130	250	FALSE	normal	187	FALSE	3.5	downsloping	0	normal	0
5	41	Female	Cleveland	atypical angina	130	204	FALSE	lv hypertrophy	172	FALSE	1.4	upsloping	0	normal	0

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.2 Pra Pemrosesan

Tahap *pra-pemrosesan* data merupakan langkah krusial untuk menyiapkan data mentah, memastikan data bersih, konsisten, dan optimal untuk pelatihan model klasifikasi serta meminimalkan bias. Proses ini mencakup:

3.2.1 Penanganan Nilai Hilang (Imputasi)

Dalam tahap pra-pemrosesan ini, dilakukan imputasi terhadap nilai hilang (*missing values*) yang terdeteksi pada beberapa *feature* di *dataset*. Proses ini krusial untuk menjaga kualitas data dan mencegah bias selama pelatihan model. Pada fitur numerik, diterapkan *median imputation* karena metode ini *robust* terhadap *outlier*. Sementara pada fitur kategorikal, digunakan *most frequent imputation* (*modus*). Hasil dari proses ini, di mana nilai-nilai kosong telah berhasil diisi, dapat dilihat pada Tabel 2.

Tabel 2. Hasil Penanganan Nilai Hilang

Id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	Oldpeak	slope	ca	thal	num
88.0	53.0	Female	Cleveland	non-anginal	128.0	216.0	FALSE	lv hypertrophy	115.0	FALSE	0.0	upsloping	0.0	normal	0.0
167.0	52.0	Male	Cleveland	non-anginal	138.0	223.0	FALSE	normal	169.0	FALSE	0.0	upsloping	0.0	normal	0.0
193.0	43.0	Male	Cleveland	asymptomatic	132.0	247.0	TRUE	lv hypertrophy	143.0	TRUE	0.1	flat	0.0	reversable defect	1.0
267.0	52.0	Male	Cleveland	asymptomatic	128.0	204	TRUE	normal	156.0	TRUE	1.0	flat	0.0	normal	1.0
288.0	58.0	Male	Cleveland	atypical angina	125.0	220.0	FALSE	normal	144.0	FALSE	0.4	flat	0.0	reversable defect	0.0

Berdasarkan Tabel 2, terlihat bahwa proses imputasi telah berhasil mengatasi nilai kosong yang semula terdeteksi. Nilai NaN pada *feature* ca (numerik) telah diisi dengan nilai 0.0 (hasil dari *median imputation*), dan nilai kosong pada *feature* thal (kategorikal) telah diisi dengan kategori normal (hasil dari *most frequent imputation*). Dengan terisinya semua nilai hilang, data menjadi lebih bersih dan konsisten, siap untuk tahapan pemrosesan berikutnya.

3.2.2 Normalisai Fitur Numerik

Pada tahap ini normalisasi terhadap seluruh *feature* numerik menggunakan metode *StandardScaler*. Metode ini menstandarkan setiap nilai *feature* agar memiliki rata-rata mendekati 0 dan standar deviasi mendekati 1. Proses normalisasi bertujuan untuk memastikan bahwa skala antar-*feature* berada dalam rentang yang sebanding, sehingga algoritma pembelajaran tidak memberikan bobot berlebih pada *feature* yang memiliki rentang nilai besar. Tabel 3 menampilkan contoh nilai *feature* setelah normalisasi diterapkan.

Tabel 3. Hasil Setelah Normalisai

id_scaled	age_scaled	trestbps_scaled	chol_scaled	thalch_scaled	oldpeak_scaled
-1.705421	1.005269	0.72193	0.290428	0.484757	1.363575
-1.701689	1.432635	1.56027	0.779084	-1.183186	0.608285
-1.697958	1.432635	-0.675305	0.253548	-0.349215	1.646809
-1.694226	-1.772612	-0.116411	0.447167	1.954135	2.496509
-1.690495	-1.345245	-0.116411	0.02305	1.358441	0.513874

Berdasarkan Tabel 3, dapat dilihat bahwa nilai seluruh *feature* numerik, seperti age, chol, dan thalch, telah berhasil ditransformasi dan berada dalam rentang skala yang sebanding (rata-rata mendekati 0 dan variansi 1). Transformasi ini memastikan bahwa jarak antar-titik data (misalnya, pada *feature* chol yang memiliki rentang ratusan dan oldpeak yang memiliki rentang satuan) memiliki bobot yang setara selama proses pelatihan model, sehingga menghindari dominasi *feature* dengan skala nilai yang besar.

3.2.3 One-Hot Encoding Fitur Kategorikal

Setela normalisasi *feature* numerik, *One-Hot Encoding* (OHE) dilakukan pada seluruh *feature* kategorikal seperti sex, dataset, cp, restecg, slope, thal, dan ca. Proses *OHE* ini bertujuan mengubah setiap kategori menjadi representasi numerik

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

biner (0 atau 1), memungkinkan algoritma pembelajaran mesin memproses data kategorikal tanpa salah menafsirkan nilai sebagai ordinal. Tabel 4 menampilkan contoh hasil *One-Hot Encoding* pada *feature* kategorikal, yang mengilustrasikan proses konversi kategori menjadi representasi numerik biner.

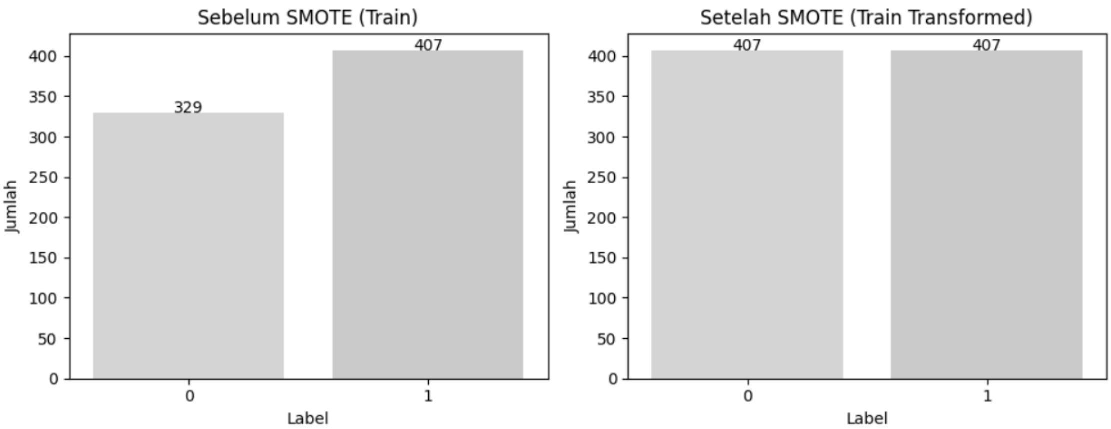
Tabel 4. Contoh Data Setelah One-Hot Encoding pada Feature Kategorikal

sex_Male	dataset_Hungary	dataset_Switzerland	dataset_VA Long Beach	cp_atypical angina	cp_non-anginal	cp_typical angina	fbs_True	restecg_normal	restecg_st-t abnormality	exang_True	slope_flat	slope_upsloping	thal_normal	thal_reversable defect	ca_1.0	ca_2.0	ca_3.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Dari hasil *One-Hot Encoding* pada Tabel 4 di atas, diketahui *feature* kategorikal berhasil dikonversi menjadi bentuk numerik biner, menghasilkan total 18 *feature* untuk pelatihan. Sebagai contoh, *feature* dataset yang memiliki empat kategori (Cleveland, Hungary, Switzerland, VA Long Beach) dikonversi menjadi kolom biner terpisah (misalnya, dataset_Hungary, dataset_Switzerland). Nilai 1.0 menunjukkan kategori yang dimiliki pasien, sedangkan nilai 0.0 menandakan tidak termasuk kategori tersebut. Proses ini krusial agar model dapat mengolah data kategorikal secara numerik tanpa kehilangan informasi penting dari tiap atribut.

3.2.4 Balancing Data Menggunakan SMOTE

Pada tahap ini, dilakukan *balancing* data menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk menyeimbangkan distribusi kelas pada data latih (*train data*). Proses ini bertujuan mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan dalam mengenali pola pada kelas minoritas (Pasien Sehat, Kelas 0). Perbandingan hasil penyeimbangan kelas dapat dilihat dalam Gambar 2 berikut.



Gambar 2. Distribusi Data Latih Sebelum dan Sesudah SMOTE

Gambar 2 memperlihatkan bahwa distribusi kelas awal pada data latih tidak seimbang, di mana jumlah sampel pada Kelas 0 (329 sampel) jauh lebih sedikit dibandingkan Kelas 1 (407 sampel). Setelah penerapan SMOTE, jumlah sampel pada kedua kelas menjadi seimbang, masing-masing 407 sampel. Penyeimbangan ini penting untuk mencegah bias model terhadap kelas mayoritas serta meningkatkan kemampuan algoritma dalam mengenali pola pada kelas minoritas, sehingga model dapat melakukan klasifikasi secara lebih objektif dan akurat terhadap kasus penyakit jantung.

3.3 Evaluasi Hasil

Evaluasi model dilakukan untuk menilai kinerja komparatif algoritma Random Forest (RF) dan XGBoost dalam klasifikasi penyakit jantung, dengan fokus pada enam metrik evaluasi utama: *Accuracy*, *Precision*, *Recall*, *F1-score*, *ROC-AUC*, dan *Matthews Correlation Coefficient* (MCC). Stabilitas dan kemampuan generalisasi model dinilai lebih lanjut melalui penerapan 10-Fold Stratified Cross-Validation, yang hasilnya diukur melalui nilai CV_F1_Mean dan CV_F1_STD. Kombinasi evaluasi ini, bersama dengan analisis *Confusion Matrix* dan *ROC Curve*, bertujuan memberikan gambaran menyeluruh mengenai tingkat akurasi dan konsistensi kedua model. Setelah proses validasi dan pengujian, metrik performa komparatif dari RF dan XGBoost ditampilkan pada Tabel 5 di bawah ini.

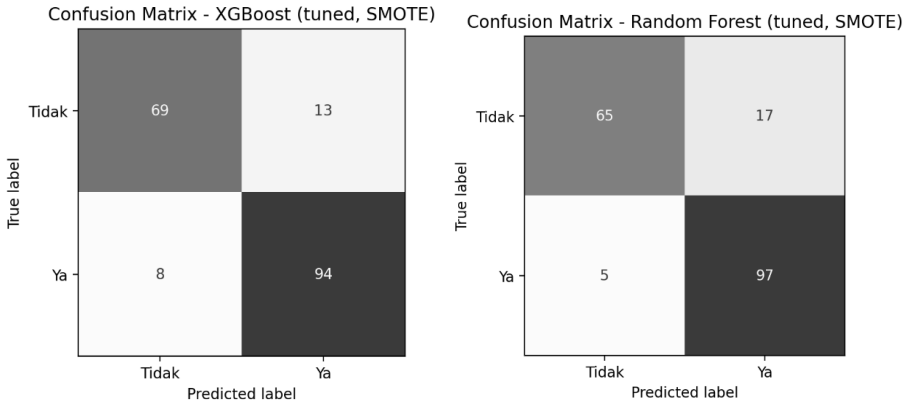
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Tabel 5. Hasil Evaluasi Metrik Kinerja Model dan Cross-Validation

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	MCC	CV_F1_Mean	CV_F1_STD
Random Forest	0,8804	0,8509	0,9510	0,8981	0,9582	0,7613	0,8917	0,0312
XGBoost	0,8859	0,8785	0,9216	0,8995	0,9519	0,7688	0,8899	0,0377

Berdasarkan hasil evaluasi komparatif yang disajikan dalam Tabel 5, observasi menunjukkan bahwa kedua model, Random Forest dan XGBoost, mencapai tingkat kinerja yang superior, dicerminkan oleh nilai *Accuracy* dan *F1-score* yang seimbang dan melebihi ambang batas 0.88. Secara spesifik, XGBoost menonjol sebagai model dengan performa agregat terbaik. Hal ini didukung oleh pencapaian *Akurasi* tertinggi sebesar 0.8859 dan *F1-score* 0.8995, serta unggul pada metrik *Precision* (0.8785) dan *Matthews Correlation Coefficient* (MCC) sebesar 0.7688. Namun, Random Forest menunjukkan kapabilitas yang lebih baik pada metrik yang berorientasi pada sensitivitas, yaitu *Recall* tertinggi (0.9510) dan *ROC-AUC* tertinggi (0.9582). Selain itu, pengujian stabilitas model melalui *Cross-Validation* (CV) menunjukkan bahwa kedua algoritma memiliki performa yang andal (CV_F1_Mean di atas 0.88). Menariknya, Random Forest memperlihatkan konsistensi yang sedikit lebih baik dalam hasil prediksinya, ditandai oleh nilai standar deviasi CV yang lebih kecil (CV_F1_STD 0.0312), yang menegaskan keandalan model saat diterapkan pada berbagai *subset* data.

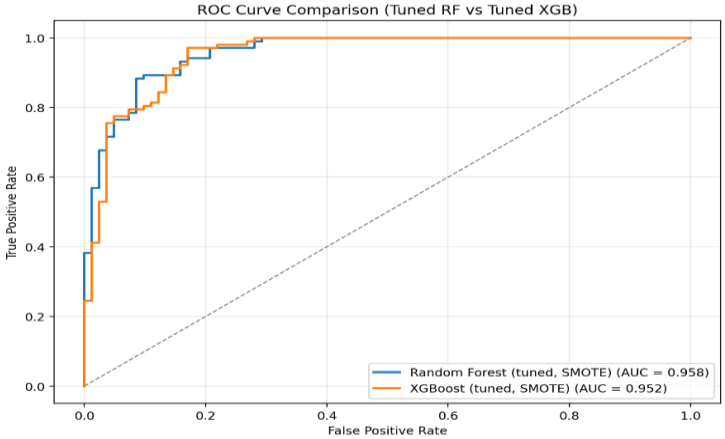
Perbedaan antara *Precision* yang lebih tinggi pada XGBoost dan *Recall* yang lebih tinggi pada Random Forest dijelaskan melalui analisis *Confusion Matrix*, yang digunakan untuk mengamati rincian kesalahan klasifikasi (Gambar 3).



Gambar 3. Confusion Matrix Model Random Forest dan XGBoost

Berdasarkan hasil visualisasi pada Gambar 3, terlihat jelas bahwa Random Forest menunjukkan kapabilitas superior dalam mengidentifikasi kelas positif, dibuktikan dengan jumlah *True Positive* yang mencapai 97 dan hanya menghasilkan 5 *False Negative*. Meskipun demikian, kemampuan deteksi yang tinggi ini berkonsekuensi pada jumlah *False Positive* yang lebih besar (17), yang menjelaskan mengapa nilai *Precision* model tersebut sedikit lebih rendah. Sebaliknya, XGBoost memperlihatkan kinerja yang lebih baik dalam meminimalkan prediksi positif yang keliru (Kesalahan Tipe I), dengan hanya 13 *False Positive* dan *True Negative* mencapai 69. Distribusi ini secara langsung menegaskan tingginya nilai *Precision* pada XGBoost. Secara keseluruhan, perbedaan distribusi kesalahan ini mencerminkan *trade-off* yang wajar antara kedua algoritma, Random Forest mengedepankan sensitivitas, sementara XGBoost mengutamakan ketepatan hasil positif.

Analisis kurva *Receiver Operating Characteristic* (ROC) digunakan untuk menilai kemampuan diskriminatif model dalam membedakan kelas positif dan negatif pada berbagai ambang batas probabilitas (Gambar 4).



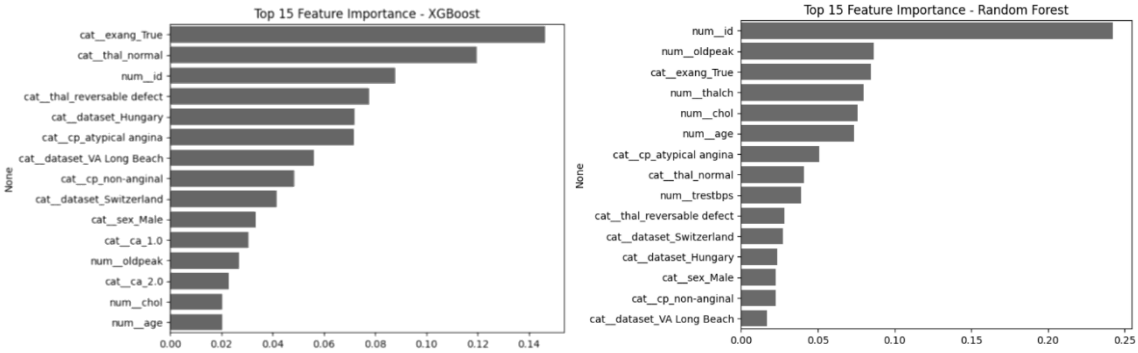
Gambar 4. Kurva ROC Model Random Forest dan XGBoost

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Berdasarkan hasil pengujian pada Gambar 4, kedua model menghasilkan nilai *ROC-AUC* yang sangat tinggi, keduanya melebihi 0.95. Model Random Forest mencatat nilai tertinggi sebesar 0.958, menunjukkan kapabilitas diskriminatif yang sedikit lebih unggul. Visualisasi kurva *ROC* memperlihatkan bahwa kedua model berada sangat dekat dengan sudut kiri atas grafik, menandakan *True Positive Rate* yang tinggi serta *False Positive Rate* yang rendah. Nilai AUC yang tinggi pada seluruh model menunjukkan konsistensi performa klasifikasi. Secara keseluruhan, XGBoost unggul dalam akurasi dan ketepatan (*Precision*), sementara Random Forest unggul dalam sensitivitas (*Recall* dan *ROC-AUC*) dan konsistensi (CV_F1_STD).

3.4 Analisis Feature Importance

Analisis Feature Importance dilakukan untuk mengidentifikasi dan membandingkan variabel prediktor mana yang memberikan kontribusi terbesar terhadap kinerja model klasifikasi, baik pada Random Forest maupun XGBoost. Perbandingan ini penting untuk memahami perbedaan cara kerja internal (*internal mechanism*) kedua model dalam memprioritaskan informasi. Hasil perbandingan tingkat kepentingan fitur pada kedua model ditampilkan pada Gambar 5.



Gambar 5. Feature Importance Model Random Forest dan XGBoost

Analisis visual pada Gambar 5 mengungkapkan adanya perbedaan prioritas fitur yang lebih baik antara kedua model, meskipun keduanya merupakan *ensemble methods*. Model Random Forest menunjukkan ketergantungan yang sangat kuat pada variabel num__id, yang mendominasi daftar fitur penting, diikuti oleh variabel numerik klinis seperti num__oldpeak dan variabel kategorikal biner cat__exang_True (angina diinduksi oleh latihan). Sementara itu, XGBoost menetapkan cat__exang_True sebagai prediktor terpenting, disusul oleh fitur klinis lainnya seperti cat__thal_normal dan cat__thal_reversable defect, serta num__id di posisi yang lebih rendah. Perbedaan ini menunjukkan bahwa Random Forest lebih mengandalkan fitur yang secara struktural memiliki variabilitas tinggi, sementara XGBoost mendistribusikan bobot kepentingan secara lebih merata ke fitur-fitur klinis utama dan hasil tes, mencerminkan kemampuannya yang adaptif untuk menangkap hubungan yang kompleks di antara variabel kategorikal. Meskipun demikian, kesamaan prioritas terhadap fitur-fitur klinis kritis seperti cat__exang_True, num__oldpeak, dan cat__thal pada kedua model menguatkan relevansi variabel-variabel tersebut dalam diagnosis penyakit jantung.

3.5 Pembahasan

Pembahasan ini bertujuan untuk menginterpretasikan kinerja komparatif algoritma *Random Forest* (*RF*) dan *XGBoost* dalam klasifikasi penyakit jantung serta mengaitkan hasil evaluasi *model* dengan analisis *Feature Importance*. Secara umum, kedua *model ensemble* berbasis *tree* menunjukkan kemampuan prediksi yang superior dalam menangani data medis yang kompleks. Hal ini tercermin dari nilai *Accuracy* dan *F1-score* yang konsisten berada di atas 0,88 setelah data melalui tahapan pra-pemrosesan, normalisasi, dan penyeimbangan kelas menggunakan *SMOTE*. Hasil ini menegaskan bahwa pendekatan *ensemble* bekerja secara optimal pada data yang dipersiapkan secara sistematis, di mana mekanisme penggabungan beberapa pohon keputusan mampu menekan *bias* dan *variance* secara efektif.

Perbedaan utama antara kedua *model* terletak pada *trade-off* antara *Precision* dan *Recall*, yang merupakan aspek krusial dalam konteks pengambilan keputusan klinis. *XGBoost* menunjukkan kinerja *Precision* yang lebih tinggi (0,8785) serta *F1-score* agregat yang lebih unggul (0,8995), mengindikasikan bahwa prediksi kelas positif yang dihasilkan memiliki tingkat ketepatan yang tinggi. Hal ini juga didukung oleh hasil *Confusion Matrix* yang menunjukkan jumlah *False Positive* yang lebih rendah, yaitu 13 kasus. Dalam praktik klinis, keunggulan *Precision* ini berperan penting dalam mengurangi risiko diagnosis positif palsu yang dapat berujung pada tindakan medis lanjutan yang tidak diperlukan.

Sebaliknya, *Random Forest* menunjukkan keunggulan signifikan pada metrik *Recall* atau sensitivitas (0,9510), yang tercermin dari jumlah *False Negative* yang sangat minimal, yakni hanya 5 kasus. Nilai *Recall* yang tinggi menjadi prioritas utama dalam skenario skrining penyakit jantung karena mampu menurunkan risiko kegagalan deteksi pada pasien yang benar-benar menderita penyakit. Selain itu, *RF* memperlihatkan konsistensi performa yang lebih baik pada berbagai *subset* data, ditunjukkan oleh nilai *CV_F1_STD* yang lebih rendah (0,0312), serta kemampuan diskriminatif yang lebih kuat dengan nilai *ROC-AUC* sebesar 0,9582.

Analisis *Receiver Operating Characteristic* (*ROC*) memberikan dukungan tambahan terhadap kemampuan diskriminatif kedua *model*. Baik *Random Forest* maupun *XGBoost* mencapai nilai *ROC-AUC* di atas 0,95, yang

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 1, Februari 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i1.9427
Hal 254-264
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

mengindikasikan performa klasifikasi yang sangat baik dalam memisahkan kelas positif dan negatif. Secara keseluruhan, meskipun *XGBoost* unggul dalam akurasi agregat dan ketepatan prediksi positif, *Random Forest* menjadi pilihan yang lebih sesuai apabila tujuan utama sistem pendukung keputusan klinis adalah memprioritaskan sensitivitas tinggi guna meminimalkan kemungkinan kasus penyakit yang tidak terdeteksi.

Analisis *Feature Importance* memberikan wawasan mengenai transparansi dan mekanisme internal kedua *model*. Kedua algoritma secara konsisten memprioritaskan fitur-fitur klinis utama seperti *cat__exang_True* (angina yang dipicu aktivitas), *num__oldpeak*, dan variabel *thal* sebagai prediktor dominan. Namun, terdapat perbedaan pola penekanan antar *model*. *Random Forest* cenderung memberikan bobot yang relatif tinggi pada *num__id*, sebuah fitur yang tidak memiliki relevansi klinis, yang mengindikasikan bahwa *model* mungkin menangkap variasi struktural dalam data. Sebaliknya, *XGBoost* menunjukkan distribusi bobot yang lebih proporsional dan relevan secara klinis dengan menempatkan *cat__exang_True* sebagai prediktor terkuat. Karakteristik *gradient boosting* yang melakukan koreksi kesalahan secara bertahap ini menghasilkan profil kepentingan fitur yang lebih adaptif dan lebih mudah diinterpretasikan dalam konteks medis.

4. KESIMPULAN

Secara umum, penelitian ini berhasil menerapkan dan membandingkan kinerja dua *model ensemble tree-based*, yaitu *Random Forest* (*RF*) dan *XGBoost*, dalam klasifikasi penyakit jantung. Hasil evaluasi menunjukkan bahwa kedua algoritma memiliki performa yang tinggi, dengan nilai *Accuracy* dan *F1-score* yang stabil di atas 0.88. Secara kuantitatif, *XGBoost* mencapai kinerja terbaik dari sisi akurasi agregat dan ketepatan prediksi positif, dengan nilai *Accuracy* sebesar 0.8859 dan *F1-score* tertinggi sebesar 0.8995, sehingga lebih efektif dalam meminimalkan kesalahan *False Positive* pada konteks klinis. Di sisi lain, *Random Forest* menunjukkan keunggulan dalam sensitivitas deteksi, ditunjukkan oleh nilai *Recall* tertinggi sebesar 0.9510, *ROC-AUC* sebesar 0.9582, serta konsistensi performa yang lebih baik dengan *CV_F1_STD* sebesar 0.0312, sehingga lebih sesuai untuk aplikasi klinis yang memprioritaskan deteksi dini dan meminimalkan kasus penyakit yang tidak teridentifikasi (*False Negative*). Analisis *Feature Importance* mengonfirmasi bahwa fitur-fitur klinis utama seperti *cat__exang_True*, *num__oldpeak*, dan *thal* merupakan prediktor dominan pada kedua *model*. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada ukuran dan keragaman *dataset* serta ruang eksplorasi *hyperparameter tuning*. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan *dataset* yang lebih besar dan beragam, menerapkan teknik optimasi *hyperparameter* yang lebih *advanced* seperti *Bayesian Optimization*, serta memanfaatkan metode interpretabilitas lanjutan seperti *SHAP* (*SHapley Additive exPlanations*) untuk meningkatkan transparansi dan relevansi klinis *model*.

REFERENCES

[1] "WHO EMRO - World Heart Day - 29 September 2024: campaigning for cardiovascular health." Accessed: Dec. 15, 2025. [Online]. Available: https://www.emro.who.int/media/news/world-heart-day-29-september-2024-campaigning-for-cardiovascular-health.html
[2] "Katalog Data - Layanan Permintaan Data | Kementerian Kesehatan RI." Accessed: Nov. 27, 2025. [Online]. Available: https://layanandata.kemkes.go.id/katalog-data/ski/ketersediaan-data/ski-2023
[3] "Cardiovascular diseases (CVDs)." Accessed: Nov. 27, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[4] "LAPORAN SKI 2023 DALAM ANGKA_REVISI I_OK.pdf," Google Docs. Accessed: Dec. 15, 2025. [Online]. Available: https://drive.google.com/file/d/1rjNDG_f8xG6-Y9wmhJUnXhJ-vUFevVJC/view?usp=sharing&usp=embed_facebook
[5] D. Triyono, R. Liani, A. W. Utami, S. Tristiyanti, and A. Supriatna, "PENYAKIT JANTUNG KORONER DI INDONESIA: PERAN FAKTOR RISIKO DAN UPAYA PENCEGAHAN," *HUMANIS J. Ilmu-Ilmu Sos. Dan Hum.*, vol. 17, no. 1, pp. 86–94, Jan. 2025, doi: 10.52166/humanis.v17i01.8798.
[6] C. Muazizah and H. Novida, "Faktor Risiko Kematian Pada Pasien Diabetes Melitus dan Penyakit Jantung: Systematic Review," *J. Ilmu Kedokt. Dan Kesehat. Indones.*, vol. 4, no. 2, pp. 01–13, Jul. 2024, doi: 10.55606/jikki.v4i2.3908.
[7] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," *J. Big Data*, vol. 10, no. 1, p. 144, Sep. 2023, doi: 10.1186/s40537-023-00817-1.
[8] H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Sci. Rep.*, vol. 14, no. 1, p. 23277, Oct. 2024, doi: 10.1038/s41598-024-74656-2.
[9] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Ensemble learning with explainable AI for improved heart disease prediction based on multiple datasets," *Sci. Rep.*, vol. 15, no. 1, p. 13912, Apr. 2025, doi: 10.1038/s41598-025-97547-6.
[10] M. W. Nugroho, "Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi," *J. JTIK J. Teknol. Inf. Dan Komun.*, vol. 9, no. 4, pp. 1562–1571, Oct. 2025, doi: 10.35870/jtik.v9i4.4236.
[11] A.-A.-R. Asif *et al.*, "Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease," vol. 29, no. 2, 2021.
[12] G. W. Sirait and T. Widodo, "Implementasi Logika Fuzzy Mamdani pada Aplikasi Sistem Pakar Diagnosis Penyakit Ternak Babi Berbasis Web," *J. Ris. Komput. JURIKOM*, vol. 12, no. 6, pp. 974–985, Dec. 2025, doi: 10.30865/jurikom.v12i6.9302.
[13] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized XGBoost based diagnostic system for effective prediction of heart disease," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4514–4523, Jul. 2022, doi: 10.1016/j.jksuci.2020.10.013.
[14] R. H. Laftah and K. H. K. Al-Saedi, "Explainable Ensemble Learning Models for Early Detection of Heart Disease".

[15] A. Rohman and S. Mujiyono, "Komparasi Algoritma Machine Learning dan Ensemble Methods dalam Prediksi Penyakit Jantung dengan Dataset yang Bervariasi," vol. 4, no. 2, 2022.
[16] Y. Amelia, "PERBANDINGAN METODE MACHINE LEARNING UNTUK MENDETEKSI PENYAKIT JANTUNG," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 220–225, Jul. 2023, doi: 10.36080/idealis.v6i2.3043.
[17] D. Setiawan, A. Muhammad, and S. H. F. Dewi, "Penerapan Algoritma Klasifikasi untuk Deteksi Dini Penyakit Jantung Koroner Berdasarkan Gejala Klinis".
[18] M. K. Biddinika, A. Masitha, H. Herman, and V. A. N. Fatimah, "Machine Learning Techniques for Heart Disease Prediction Using a Multi-Algorithm Approach," *JUITA J. Inform.*, vol. 12, no. 2, p. 149, Nov. 2024, doi: 10.30595/juita.v12i2.24153.
[19] N. Nasution, M. A. Hasan, and F. Bakri Nasution, "Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset," *IT J. Res. Dev.*, vol. 9, no. 2, pp. 140–150, Apr. 2025, doi: 10.25299/itjrd.2025.17941.
[20] D. Wijayanto, R. Marco, A. Sidauruk, and M. Sulistiyono, "The Effect of SMOTE and Optuna Hyperparameter Optimization on TabNet Performance for Heart Disease Classification," *J. Sisfokom Sist. Inf. Dan Komput.*, vol. 14, no. 2, pp. 156–164, May 2025, doi: 10.32736/sisfokom.v14i2.2348.
[21] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang," *MATRIK J. Manaj. Tek. Inform. Dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, Jul. 2022, doi: 10.30812/matrik.v21i3.1726.
[22] J. P. Anggraini, Chaya Gladys Zhafirah A, and A. Desiani, "Perbandingan Algoritma Random Forest dan Extreme Gradient Boosting (XGBoost) dalam Klasifikasi Penyakit Gagal Jantung," *Komputika J. Sist. Komput.*, vol. 14, no. 2, pp. 149–157, Nov. 2025, doi: 10.34010/komputika.v14i2.16618.