 ISSN:2302-8149
e-ISSN:2540-9719

Optimasi *Hyperparameter Ensemble Learning* untuk Prediksi Penyakit Jantung berdasarkan Data Pasien

Hyperparameter Optimization of Ensemble Learning for Heart Disease Prediction using Patient Data

[1]Nikko Listio Wicaksono*, [2]Kusrini
[1,2]Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta
[1,2]Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta, Indonesia
*e-mail: *nikkolistiow@students.amikom.ac.id**, *kusrini@amikom.ac.id*

(***received*:** 26 December 2025, ***revised*:** 10 February 2026, ***accepted*:** 12 February 2026)

Abstrak

Penelitian ini mengevaluasi dampak optimasi hyperparameter terhadap performa empat algoritma machine learning—Extra Trees, XGBoost, Random Forest, dan AdaBoost—dalam prediksi penyakit jantung. Hasil menunjukkan bahwa tuning hyperparameter secara signifikan meningkatkan performa model pada tiga dari empat algoritma, dengan variasi antar algoritma. Extra Trees menunjukkan peningkatan paling konsisten, mencapai Area Under Curve (AUC) tertinggi sebesar 0.9107 dan recall sebesar 80.93%, yang sangat krusial dalam konteks medis untuk mendeteksi kasus penyakit secara akurat. XGBoost mengalami lonjakan akurasi terbesar dari 78.11% menjadi 81.49%, sementara Random Forest meningkat dalam recall dan F1-Score. Sebaliknya, AdaBoost mengalami penurunan ringan, menunjukkan bahwa model tersebut sudah optimal sebelum tuning. Secara keseluruhan, Extra Trees dengan hyperparameter tuning terbukti sebagai algoritma terbaik untuk aplikasi prediksi penyakit jantung, menawarkan keandalan tinggi dalam mengidentifikasi pasien berisiko.

Kata kunci: *adaboost, extra trees*, optimasi hyperparameter, algoritma *machine learning*, *random forest, xgboost.*

Abstract

This study evaluates the impact of hyperparameter optimization on the performance of four machine learning algorithms—Extra Trees, XGBoost, Random Forest, and AdaBoost—in heart disease prediction. The results show that hyperparameter tuning significantly improves model performance for three out of the four algorithms, with varying effects across models. Extra Trees demonstrates the most consistent improvement, achieving the highest Area Under the Curve (AUC) of 0.9107 and a recall of 80.93%, which is particularly crucial in medical contexts for accurately identifying disease cases. XGBoost exhibits the largest increase in accuracy, rising from 78.11% to 81.49%, while Random Forest shows improvements in both recall and F1-score. In contrast, AdaBoost experiences a slight decline in performance, suggesting that the model was already near optimal prior to tuning. Overall, Extra Trees with hyperparameter optimization emerges as the best-performing algorithm for heart disease prediction, offering high reliability in identifying at-risk patients.

Keywords: *adaboost, extra trees, hyperparameter optimization, machine learning algorithms, random forest, xgboost.*

1. Pendahuluan

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia. Menurut laporan World Health Organization, penyakit kardiovaskular menyumbang sekitar 32% dari seluruh kematian global setiap tahun [1]. Kondisi ini juga menjadi salah satu beban kesehatan paling signifikan di Indonesia, dengan prevalensi yang meningkat seiring perubahan gaya hidup, pola makan tidak sehat, dan tingginya tingkat stres masyarakat [2].

Situasi ini menunjukkan pentingnya upaya mitigasi melalui deteksi dini dan pencegahan. Diagnosis dini penyakit jantung berperan penting dalam menekan risiko komplikasi serius serta meningkatkan efektivitas pengobatan. Namun, metode diagnosis konvensional masih memiliki

 ISSN:2302-8149
e-ISSN:2540-9719

keterbatasan karena bergantung pada analisis manual dokter yang dapat dipengaruhi oleh bias subjektif dan pengalaman klinis [3].

Perkembangan teknologi komputasi membuka peluang untuk meningkatkan akurasi dan konsistensi proses diagnosis melalui pendekatan berbasis data. Metode machine learning telah banyak diterapkan dalam sistem pendukung keputusan (decision support system) untuk mengidentifikasi pola dan faktor risiko penyakit secara lebih objektif serta efisien. Namun, sebagian besar penelitian terdahulu masih menghadapi tantangan, khususnya terkait kualitas data medis yang tidak seimbang (imbalanced) serta ketergantungan performa model terhadap pemilihan hyperparameter yang sering kali belum dianalisis secara mendalam dan komprehensif.

Berangkat dari permasalahan tersebut, penelitian ini berfokus pada pengembangan dan evaluasi model machine learning berbasis ensemble learning dengan optimasi hyperparameter untuk meningkatkan performa prediksi risiko penyakit jantung. Berbeda dengan penelitian sebelumnya yang umumnya hanya mengevaluasi satu algoritma atau menitikberatkan pada metrik akurasi, penelitian ini membandingkan beberapa algoritma ensemble dengan karakteristik yang berbeda serta menekankan metrik evaluasi yang krusial dalam konteks medis, seperti AUC dan recall. Selain itu, penelitian ini juga mempertimbangkan teknik penanganan data tidak seimbang guna memaksimalkan kemampuan model dalam mengidentifikasi pasien dengan risiko tinggi secara lebih andal.

2. Tinjauan Literatur

Machine learning (ML) telah menjadi pendekatan yang menjanjikan dalam mendeteksi dan memprediksi penyakit karena kemampuannya mengenali pola kompleks dalam data medis. Salah satu pendekatan ML yang terbukti efektif adalah ensemble learning, yaitu teknik yang menggabungkan beberapa model prediksi untuk menghasilkan performa yang lebih stabil dan akurat. Berbagai algoritma ensemble, seperti Random Forest, AdaBoost, dan Gradient Boosting, telah menunjukkan hasil yang kuat dalam klasifikasi risiko penyakit jantung [4].

Meskipun demikian, kinerja algoritma–algoritma tersebut sangat dipengaruhi oleh pengaturan hyperparameter yang optimal. Oleh karena itu, optimasi hyperparameter menjadi bagian penting dalam meningkatkan akurasi prediksi model ensemble [5]. Penelitian oleh Nurkholifah juga menegaskan bahwa algoritma machine learning dapat membantu menganalisis data pasien secara lebih efektif dan mengidentifikasi pola yang tidak mudah ditemukan melalui analisis manual [6].

Tabel 1 Tabel perbandingan penelitian terdahulu

Tahun	Algoritma	Optimasi	Dataset	Hasil
2024	Ensemble Learning (RF, AdaBoost, Bagging)	Optimasi hyperparameter	NASA JM1	Model ensemble teroptimasi mencapai akurasi 0,82, ROC 0,79, dan F-measure 0,79.
2019	Ensemble Boosting	Seleksi model & iterasi boosting	ILPD	Boosting menghasilkan akurasi 71,53% dan AUC 0,722 pada dataset ILPD.
2025	RF, XGBoost, LightGBM, Extra Trees	Random Search	Data klorofil	LightGBM unggul dengan NRMSE 0,63 dan waktu komputasi 2,81 ms.
2020	Ensemble Tree Models	Feature extraction & selection	CICAndMal2017	Extra Trees mencapai akurasi 87,75% (deteksi), 79,97% (kategori), dan 66,71% (keluarga malware).
2020	LR, RF, Extra Trees, Voting	Feature selection & imbalance handling	Data sensor sapi perah	Model terbaik mencapai AUC 0,72–0,79 (mastitis) dan 0,66–0,71 (lameness).
2020	Extra Trees, XGBoost, SVM	Model selection & preprocessing	BigML Telecom	Extra Trees mencapai AUC 0,8439, F1-score 0,5666, dan MCC 0,7098 pada prediksi churn.

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan penelitian terdahulu pada tabel 1, algoritma ensemble learning seperti Random Forest, Extra Trees, XGBoost, dan AdaBoost telah banyak diterapkan pada berbagai domain, termasuk prediksi perangkat lunak, kesehatan, lingkungan, keamanan siber, dan analisis bisnis. Hasil penelitian menunjukkan bahwa metode ensemble mampu meningkatkan performa prediksi dibandingkan algoritma tunggal, terutama ketika dikombinasikan dengan optimasi hyperparameter [5], [7], [8].

Beberapa penelitian menunjukkan bahwa optimasi hyperparameter dapat meningkatkan akurasi dan stabilitas model secara signifikan. Misalnya, penelitian [5] menunjukkan bahwa optimasi hyperparameter pada model ensemble dapat meningkatkan performa prediksi bug perangkat lunak, sementara penelitian [7] membuktikan bahwa optimasi parameter pada model tree-based ensemble mampu menghasilkan performa terbaik pada prediksi konsentrasi klorofil. Selain itu, penelitian [8] menunjukkan bahwa pemilihan model terbaik melalui perbandingan algoritma supervised learning dapat meningkatkan kualitas prediksi churn pelanggan.

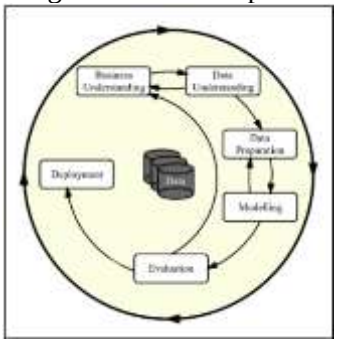
Dalam bidang kesehatan, metode ensemble juga telah diterapkan untuk prediksi penyakit, seperti pada penelitian [9] dan [10]. Penelitian [9] menunjukkan bahwa algoritma boosting, khususnya LogitBoost, memberikan performa terbaik pada prediksi penyakit hati, sedangkan penelitian [10] menunjukkan bahwa model ensemble dapat memberikan performa yang kompetitif dalam klasifikasi kondisi kesehatan berdasarkan data sensor. Namun, sebagian besar penelitian tersebut masih berfokus pada satu algoritma tertentu atau satu strategi optimasi, sehingga belum memberikan pemahaman komprehensif mengenai perbandingan efektivitas optimasi hyperparameter pada berbagai algoritma ensemble.

Selain itu, beberapa penelitian belum secara mendalam mempertimbangkan permasalahan data tidak seimbang (imbalanced data), yang merupakan karakteristik umum pada dataset medis. Padahal, ketidakseimbangan data dapat memengaruhi kinerja model, terutama pada metrik yang krusial dalam konteks medis seperti recall dan AUC [9], [10]. Di sisi lain, sebagian penelitian masih menitikberatkan evaluasi pada metrik akurasi, tanpa mempertimbangkan metrik lain yang lebih relevan dalam konteks prediksi risiko penyakit.

Berdasarkan celah penelitian tersebut, penelitian ini bertujuan untuk membandingkan performa beberapa algoritma ensemble learning, yaitu Random Forest, Extra Trees, XGBoost, dan AdaBoost, dengan menerapkan optimasi hyperparameter serta teknik penanganan data tidak seimbang. Penelitian ini juga menggunakan metrik evaluasi yang lebih komprehensif, seperti AUC, recall, precision, dan F1-Score, guna memperoleh model prediksi penyakit jantung yang lebih akurat, stabil, dan relevan secara klinis. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam memperjelas hubungan antara karakteristik algoritma ensemble, strategi optimasi hyperparameter, dan performa model dalam domain kesehatan.

3. Metode Penelitian

Penelitian ini menggunakan kerangka kerja CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai metodologi utama dalam proses analisis data. CRISP-DM dipilih karena mampu memberikan pendekatan yang sistematis, terstruktur, dan telah terbukti efektif dalam menyelesaikan permasalahan data mining pada berbagai domain, termasuk prediksi di bidang kesehatan [11]. Metodologi ini terdiri dari enam tahap yang saling berkaitan, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Setiap tahap memiliki peran penting dalam menghasilkan model prediksi yang akurat dan dapat diandalkan.



Gambar 1 Alur proses CRISP-DM [11]

 ISSN:2302-8149
e-ISSN:2540-9719

Lingkaran terluar pada Gambar 1 menunjukkan bahwa data mining bersifat berulang. Setelah solusi diterapkan, proses tidak berhenti karena temuan dan pengalaman sebelumnya sering menghasilkan pertanyaan bisnis baru yang lebih spesifik. Siklus berikutnya pun menjadi lebih efektif berkat pembelajaran dari tahap sebelumnya [11].

3.1. Business Understanding

Tahap awal ini bertujuan memahami konteks dan tujuan dari penelitian, yakni membangun model prediksi penyakit jantung berbasis data pasien. Penyakit jantung dipilih karena merupakan salah satu penyebab kematian utama secara global, dan deteksi dini dapat meningkatkan tingkat keselamatan pasien. Dalam tahap ini, ditentukan tujuan utama yaitu membangun model ensemble learning yang mampu mengklasifikasikan pasien berisiko secara akurat, sehingga dapat digunakan sebagai alat bantu keputusan medis .

3.2. Data Understanding

Setelah tujuan ditetapkan, tahap berikutnya adalah mengumpulkan dan memahami data yang akan digunakan. Dataset yang digunakan merupakan data publik berisi atribut medis pasien, seperti tekanan darah, kolesterol, kadar gula, usia, jenis kelamin, dan lainnya. Pada tahap ini dilakukan eksplorasi awal untuk mendeteksi distribusi data, anomali, nilai kosong (missing values), serta hubungan antar fitur menggunakan teknik statistik dan visualisasi data .

3.3. Data Preparation

Tahap ini meliputi persiapan data agar dapat digunakan dalam proses pemodelan. Data dibersihkan dari outlier, nilai kosong diatasi dengan teknik imputasi, dan dilakukan normalisasi agar semua fitur memiliki skala yang seragam. Dataset kemudian dibagi menjadi training set dan testing set agar model dapat diuji dan divalidasi dengan baik.

3.4. Modeling

Pada tahap ini dilakukan pembangunan model prediksi menggunakan algoritma ensemble learning yaitu Random Forest, XGBoost, AdaBoost, dan Extra Trees. Pemilihan algoritma tersebut didasarkan pada perbedaan karakteristik pendekatan ensemble, sehingga memungkinkan perbandingan performa model dalam memprediksi risiko penyakit jantung.

Untuk meningkatkan kinerja model, dilakukan optimasi hyperparameter menggunakan RandomizedSearchCV dengan skema 10-fold cross validation. Parameter yang dioptimasi meliputi n_estimators, max_depth, learning_rate, dan min_samples_split dengan rentang nilai yang disajikan pada Tabel 2. Model kemudian dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, dan AUC, dengan penekanan pada recall dan AUC karena penting dalam konteks deteksi risiko penyakit jantung.

Tabel 2 Rentang hyperparameter yang dioptimasi

Algoritma	Hyperparameter	Rentang Nilai
Random Forest	n_estimators	100,200
	max_depth	5,10,None
	min_samples_split	2,5
XGBoost	n_estimators	100,200
	max_depth	3,6
	learning_rate	0.01, 0.1, 0.2
AdaBoost	n_estimators	50, 75, 100
	learning_rate	0.1, 0.5, 1.0
Extra Trees	n_estimators	100,200
	max_depth	5,10,None
	min_samples_split	2,5

3.4.1 Random Forest

Random Forest merupakan metode ensemble learning yang mengintegrasikan sejumlah pohon keputusan untuk memperoleh hasil prediksi yang lebih akurat dan konsisten. Setiap pohon keputusan dilatih menggunakan subset data dan fitur yang berbeda sehingga mampu

 ISSN:2302-8149
e-ISSN:2540-9719

mengurangi risiko overfitting [12]. Prediksi akhir ditentukan berdasarkan hasil voting mayoritas dari seluruh pohon keputusan, sebagaimana dirumuskan pada Persamaan (1).

$$\hat{y} = mode\ (\{h_1(x), h_2(x), \dots, h_T(x)\}) \quad (1)$$

$\hat{y}$ = hasil prediksi akhir
$h_t(x)$ = prediksi dari pohon ke-terhadap input *x*
T = jumlah pohon dalam *Random Forest*

3.4.2 XGBoost

XGBoost (Extreme Gradient Boosting) merupakan algoritma ensemble berbasis gradient boosting yang banyak digunakan dalam tugas klasifikasi karena kemampuannya membangun kumpulan pohon keputusan secara bertahap. Setiap pohon baru dibentuk untuk memperbaiki kesalahan dari model sebelumnya. Proses pelatihan model dilakukan dengan meminimalkan fungsi loss pada klasifikasi biner, yaitu binary log loss, sebagaimana ditunjukkan pada Persamaan (2). Fungsi ini mengukur perbedaan antara label aktual dan probabilitas prediksi yang dihasilkan oleh model [13].

$$L(y, \hat{y}) = \sum_{i=1}^{n} [\ (\hat{y}i) + (1 - yi) \log(1 - \hat{y}i)] \quad (2)$$

y = label asli (*ground truth*), bernilai 0 atau 1
$\hat{y}$ = probabilitas prediksi (nilai antara 0 dan 1
$\hat{y}$ = jumlah total data

3.4.3 AdaBoost

AdaBoost (Adaptive Boosting) merupakan metode ensemble learning yang mengombinasikan sejumlah klasifikator lemah, umumnya berupa decision stump, untuk menghasilkan model prediksi yang lebih akurat. Pada setiap iterasi, bobot data latih akan diperbarui sehingga data yang salah diklasifikasikan memperoleh perhatian lebih besar pada proses pelatihan berikutnya [14]. Prediksi akhir ditentukan berdasarkan kombinasi berbobot dari seluruh weak classifier, sebagaimana dirumuskan pada Persamaan (3).

$$\hat{y}(x) = sign\ (\sum_{t=1}^{T} \alpha_t\ h_t(x)) \quad (3)$$

$\hat{y}(x)$ = hasil prediksi akhir
$h_t(x)$ = *weak classifier* (misal stump) ke-t
α_t = bobot dari *weak classifier* ke-t, tergantung pada akurasinya
T = jumlah total iterasi
sign = fungsi tanda (positif = +1, negatif = -1)

3.4.4 Extra Trees

Extra Trees (Extremely Randomized Trees) merupakan salah satu varian dari algoritma Random Forest yang digunakan untuk menyelesaikan permasalahan klasifikasi maupun regresi [12]. Tidak seperti Random Forest yang mencari titik split terbaik berdasarkan kriteria seperti information gain atau Gini impurity, Extra Trees melakukan pemilihan fitur dan nilai split secara acak, sehingga proses pembentukan pohon menjadi lebih acak dan cepat (4).

$$\hat{y} = mode\ (\{h_1(x), h_2(x),\ \dots\ h_T(x)\}) \quad (4)$$

$h_1(x)$ = prediksi dari pohon ke-*t* untuk input *x*
T = jumlah total pohon dalam ensemble
mode = nilai kelas yang paling sering muncul

3.5. Evaluation

Model yang telah dibangun dievaluasi berdasarkan metrik evaluasi yang relevan. Proses evaluasi ini bertujuan memastikan bahwa model tidak hanya akurat, tetapi juga memiliki generalisasi yang baik terhadap data baru. Analisis dilakukan terhadap kesalahan prediksi serta kekuatan dan kelemahan masing-masing model.

 ISSN:2302-8149
e-ISSN:2540-9719

3.6. Deployment

Tahap akhir adalah implementasi model ke dalam sistem yang dapat digunakan secara praktis, misalnya dalam bentuk antarmuka pengguna sederhana atau integrasi ke dalam sistem klinis. Meskipun dalam penelitian ini deployment hanya sebatas simulasi, namun tahapan ini penting untuk menggambarkan bagaimana model dapat digunakan dalam praktik nyata untuk membantu deteksi dini penyakit jantung.

4. Hasil dan Pembahasan

Pada bagian ini, hasil eksperimen terkait optimasi hyperparameter pada berbagai algoritma ensemble learning untuk prediksi penyakit jantung berdasarkan data pasien akan dibahas secara mendetail. Model yang diuji meliputi Random Forest, XGBoost, AdaBoost, dan Extra Trees, dengan evaluasi menggunakan metrik akurasi, AUC, recall, precision, dan F1-score. Optimasi hyperparameter dilakukan untuk meningkatkan kemampuan masing-masing model dalam membedakan pasien yang memiliki risiko penyakit jantung dan yang tidak. Analisis hasil dilakukan dengan membandingkan kinerja model sebelum dan sesudah optimasi, serta meninjau distribusi klasifikasi menggunakan confusion matrix dan kurva AUC-ROC. Selain itu, pembahasan juga mencakup analisis pengaruh masing-masing fitur dalam proses prediksi, interpretasi hasil menggunakan teknik feature importance, serta strategi yang dapat diterapkan untuk lebih meningkatkan kinerja model pada data serupa di masa mendatang.

4.1. Dataset

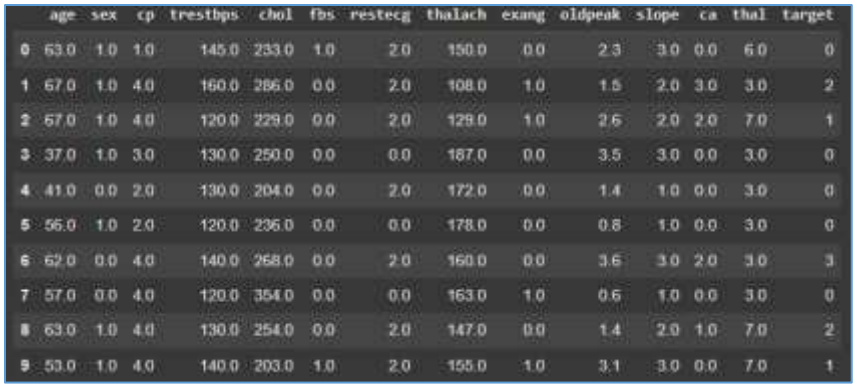
Dataset yang digunakan dalam penelitian ini adalah Heart Disease Dataset yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi data rekam medis pasien yang digunakan untuk mengidentifikasi kemungkinan adanya penyakit jantung berdasarkan sejumlah parameter klinis dan demografis. Dataset terdiri dari 13 fitur prediktor (features) yang mewakili karakteristik medis pasien, serta satu variabel target (target variable) yang menunjukkan ada atau tidaknya penyakit jantung. Berikut penjelasan singkat masing-masing fitur pada Tabel 3.

Tabel 3 Gambaran umum dataset

Fitur	Keterangan
Age	usia pasien dalam tahun.
Sex	jenis kelamin pasien
CP (Chest Pain Type)	tipe nyeri dada yang dialami pasien, dengan kategori 0–3 (misalnya typical angina, atypical angina, non-anginal pain, asymptomatic)
Trestbps	tekanan darah istirahat pasien (mm Hg
Chol	kadar kolesterol serum dalam mg/d
FBS (Fasting Blood Sugar)	kadar gula darah puasa (1 jika > 120 mg/dl, 0 jika ≤ 120 mg/dl)
Restecg	hasil pemeriksaan elektrokardiografi istirahat (0 = normal, 1 = kelainan gelombang ST-T, 2 = hipertrofi ventrikel kiri)
Thalach	detak jantung maksimum yang dicapai pasien selama uji latihan
Exang	detak adanya angina akibat latihan
Oldpeak	penurunan segmen ST pada elektrokardiogram setelah latihan (diukur dalam mm)
Slope	kemiringan segmen ST pada puncak latihan
CA	jumlah pembuluh darah besar (0–3) yang terlihat melalui fluoroskopi
Thal	status thalassemia pasien

Dataset yang digunakan dalam penelitian ini adalah Heart Disease Dataset yang diperoleh dari UCI Machine Learning Repository ditampilkan pada Gambar 2.

 ISSN:2302-8149
e-ISSN:2540-9719



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0	0
1	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2
2	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
3	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
4	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0
5	56.0	1.0	2.0	120.0	236.0	0.0	0.0	178.0	0.0	0.8	1.0	0.0	3.0	0
6	62.0	0.0	4.0	140.0	268.0	0.0	2.0	160.0	0.0	3.6	3.0	2.0	3.0	3
7	57.0	0.0	4.0	120.0	354.0	0.0	0.0	163.0	1.0	0.6	1.0	0.0	3.0	0
8	63.0	1.0	4.0	130.0	254.0	0.0	2.0	147.0	0.0	1.4	2.0	1.0	7.0	2
9	53.0	1.0	4.0	140.0	203.0	1.0	2.0	155.0	1.0	3.1	3.0	0.0	7.0	1

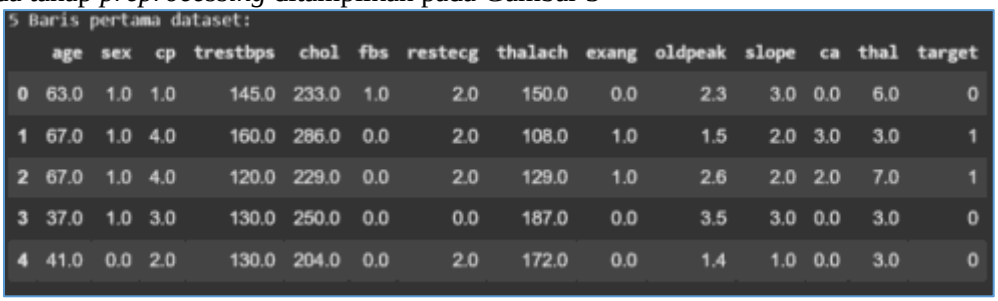
Gambar 2 Tabel dataset UCI

4.2. Preprocessing Dataset

Preprocessing dataset dilakukan untuk membersihkan dan mempersiapkan data agar siap digunakan dalam pemodelan klasifikasi biner (target: 0=sehat, 1=sakit). Dataset asli berukuran 303 baris × 14 kolom, dengan fitur sebagian besar numerik. Berikut tahapan utamanya secara berurutan:

a. Pemuatan Data: Dataset dimuat dari URL UCI menggunakan Pandas (pd.read_csv), dengan kolom diberi nama eksplisit (age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal, target).
b. Penanganan Nilai Hilang: Nilai "?" (missing values, terutama di kolom 'ca' dan 'thal') diganti dengan NaN menggunakan df.replace("?", np.nan). Pendekatan ini mengidentifikasi missing values secara standar tanpa imputasi, karena proporsinya rendah (<2%).
c. Konversi Tipe Data: Semua kolom dikonversi ke numerik menggunakan pd.to_numeric(errors='coerce'). Kolom kategorikal (seperti 'cp', 'thal') yang sudah di-encode sebagai angka tetap numerik; nilai tidak valid menjadi NaN.
d. Biner Label Target: Kolom 'target' (nilai asli 0–4) diubah menjadi biner: 0 (sehat) jika ≤0, dan 1 (sakit) jika >0, menggunakan fungsi lambda.
e. Pemisahan Fitur dan Label: Fitur (X) diambil dari 13 kolom selain 'target', sementara label (y) adalah kolom 'target'.

Hasil pada tahap *preprocessing* ditampilkan pada Gambar 3



5 Baris pertama dataset:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63.0	1.0	1.0	145.0	233.0	1.0	2.0	150.0	0.0	2.3	3.0	0.0	6.0	0
1	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	1
2	67.0	1.0	4.0	120.0	229.0	0.0	2.0	129.0	1.0	2.6	2.0	2.0	7.0	1
3	37.0	1.0	3.0	130.0	250.0	0.0	0.0	187.0	0.0	3.5	3.0	0.0	3.0	0
4	41.0	0.0	2.0	130.0	204.0	0.0	2.0	172.0	0.0	1.4	1.0	0.0	3.0	0

Gambar 3 Tabel hasil preprocessing

4.3. Pemodelan

Dalam penelitian ini, pemodelan diterapkan untuk menentukan model ensemble optimal yang paling akurat dalam memprediksi penyakit jantung berdasarkan data klinis pasien. Berbagai metode evaluasi model yang diterapkan pada studi ini meliputi:

a. Akurasi merupakan metrik evaluasi yang digunakan untuk mengukur proporsi jumlah prediksi yang benar terhadap seluruh data. Metrik ini efektif untuk menilai performa umum model pada dataset yang seimbang, namun kurang representatif pada data yang tidak seimbang karena cenderung bias terhadap kelas mayoritas [15]. Perhitungan akurasi ditunjukkan pada Persamaan (4).

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

TP = (*True Positive*) : jumlah data positif yang diprediksi benar

TN = (*True Negative*) : jumlah data negatif yang diprediksi benar

FP = (*False Positive*) : jumlah data negatif yang diprediksi sebagai positif

 ISSN:2302-8149
e-ISSN:2540-9719

FN = (*False Negative*) : jumlah data positif yang diprediksi sebagai negatif

b. AUC (Area Under ROC Curve / ROC-AUC). AUC mengukur luas di bawah kurva ROC, yang memetakan TPR (TP / (TP + FN)) vs FPR (FP / (FP + TN)) pada berbagai threshold. Metrik ini threshold-independent, ideal untuk membedakan kelas pada data imbalance seperti deteksi fraud, dan sering digunakan untuk tuning model [16].

c. Recall (Sensitivity atau True Positive Rate) merupakan metrik evaluasi yang mengukur kemampuan model dalam mengidentifikasi seluruh data positif yang benar. Metrik ini berfokus pada pengurangan kesalahan false negative, sehingga sangat penting pada kasus di mana kegagalan mendeteksi kelas positif harus diminimalkan [17]. Perhitungan recall ditunjukkan pada Persamaan (5).

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

d. recision (Positive Predictive Value) merupakan metrik evaluasi yang mengukur ketepatan model saat memprediksi kelas positif, yaitu perbandingan antara jumlah prediksi positif yang benar dengan seluruh prediksi positif yang dihasilkan model [17]. Perhitungan precision ditunjukkan pada Persamaan (6).

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

e. F1-Score merupakan metrik komposit yang menggabungkan precision dan recall menggunakan harmonic mean untuk memberikan ukuran kinerja yang seimbang, terutama ketika kedua metrik tersebut sama pentingnya [18]. Perhitungan F1-Score ditunjukkan pada Persamaan (7).

$$F1\ Score = \frac{2 \times Presisi \times Recall}{Presisi+Recall} \quad (7)$$

Tabel 4 Hasil evaluasi algoritma

Algoritma	Akurasi	AUC	Recall	Presisi	F1-Score
Random Forest	0.8184	0.9031	0.7588	0.8388	0.7920
XGBoost	0.7811	0.8654	0.7516	0.7845	0.7594
AdaBoost	0.8317	0.8895	0.8033	0.8333	0.8129
Extra Trees	0.8111	0.8997	0.7797	0.8124	0.7906

Berdasarkan tabel 4 hasil evaluasi keempat algoritma, AdaBoost menempati posisi terbaik secara keseluruhan dengan mencetak nilai tertinggi pada Akurasi (0.8317), Recall (0.8033), dan F1-Score (0.8129), yang menunjukkan kemampuannya dalam menghasilkan prediksi yang akurat sekaligus mendeteksi lebih banyak instance positif dengan keseimbangan yang optimal. Sementara itu, Random Forest unggul dalam aspek kepercayaan prediksi positif dengan nilai Presisi tertinggi (0.8388) serta memiliki daya beda model terbaik yang ditunjukkan oleh nilai AUC tertinggi (0.9031). XGBoost dan Extra Trees menunjukkan performa yang solid namun konsisten berada di peringkat tengah. Dengan demikian, pemilihan model akhir dapat diarahkan pada AdaBoost jika fokusnya adalah keseimbangan dan deteksi positif, atau Random Forest jika yang diutamakan adalah akurasi prediksi positif dan kemampuan diskriminasi model.

4.4. Optimasi Hyperparameter

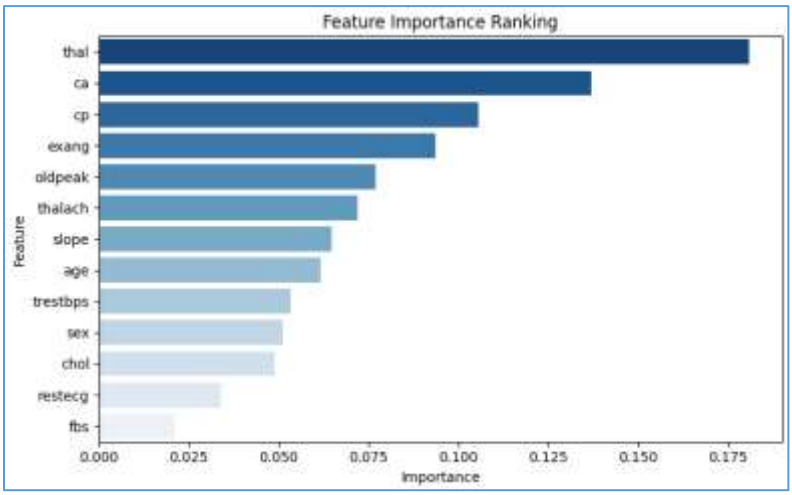
Tabel 5 Hasil optimasi Hyperparameter tunning

Algoritma	Akurasi	AUC	Recall	Presisi	F1-Score
Random Forest	0.8184	0.9031	0.7588	0.8388	0.7920
Random Forest + Hyperparameter tunning	0.8184	0.9029	0.7808	0.8278	0.7981
XGBoost	0.7811	0.8654	0.7516	0.7845	0.7594
XGBoost + Hyperparameter tunning	0.8149	0.8833	0.7225	0.8560	0.7807
AdaBoost	0.8317	0.8895	0.8033	0.8333	0.8129
Ada Boost + Hyperparameter tunning	0.8282	0.8962	0.7879	0.8282	0.8067
Extra Trees	0.8111	0.8997	0.7797	0.8124	0.7906
Extra Trees + Hyperparameter tunning	0.8283	0.9107	0.8093	0.8268	0.8122

 ISSN:2302-8149
e-ISSN:2540-9719

Berikut adalah hasil evaluasi beberapa algoritma klasifikasi beserta performanya berdasarkan metrik Akurasi, AUC, Recall, Presisi, dan F1-Score pada Tabel 5. Algoritma Random Forest menunjukkan akurasi sebesar 0.8184 dengan AUC 0.9031, Recall 0.7588, Presisi 0.8388, dan F1-Score 0.7920. Ketika Random Forest dikombinasikan dengan tuning hyperparameter, performanya sedikit berubah dengan akurasi tetap 0.8184, AUC 0.9029, Recall meningkat menjadi 0.7808, presisi sedikit menurun menjadi 0.8278, dan F1-Score naik menjadi 0.7981. Algoritma XGBoost memiliki akurasi 0.7811, AUC 0.8654, Recall 0.7516, presisi 0.7845, dan F1-Score 0.7594. Dengan hyperparameter tuning, XGBoost memperlihatkan akurasi 0.8149, AUC 0.8833, recall menurun menjadi 0.7225, presisi meningkat menjadi 0.8560, dan F1-Score menjadi 0.7807. Pada dataset yang sama, AdaBoost menunjukkan hasil akurasi 0.8317, AUC 0.8895, Recall 0.8033, presisi 0.8333, dan F1-Score 0.8129. Setelah tuning hyperparameter, ada perubahan menjadi akurasi 0.8282, AUC 0.8962, recall turun menjadi 0.7879, presisi turun menjadi 0.8282, dan F1-Score menjadi 0.8067. Algoritma Extra Trees memiliki akurasi 0.8111, AUC 0.8997, Recall 0.7797, presisi 0.8124, dan F1-Score 0.7906. Dengan tuning hyperparameter, akurasinya menjadi 0.8283, AUC meningkat ke 0.9107, recall meningkat ke 0.8093, presisi menjadi 0.8268, dan F1-Score menjadi 0.8122. Secara keseluruhan, tuning hyperparameter cenderung meningkatkan AUC dan F1-Score pada beberapa algoritma, meskipun ada fluktuasi pada nilai recall dan presisi.

4.5. Feature Importance



Gambar 4 *Ranking feature importance*

Gambar 4 menunjukkan hasil feature importance dari model Extra Trees, yang merupakan algoritma dengan performa terbaik berdasarkan nilai AUC dan recall. Analisis ini digunakan untuk mengidentifikasi fitur yang paling berpengaruh dalam prediksi risiko penyakit jantung. Hasil menunjukkan bahwa fitur thal, ca, cp, dan exang memiliki kontribusi paling besar terhadap prediksi model. Dominannya fitur-fitur tersebut mengindikasikan bahwa kondisi fisiologis dan fungsional jantung berperan penting dalam menentukan risiko penyakit jantung, sejalan dengan temuan dalam literatur medis.

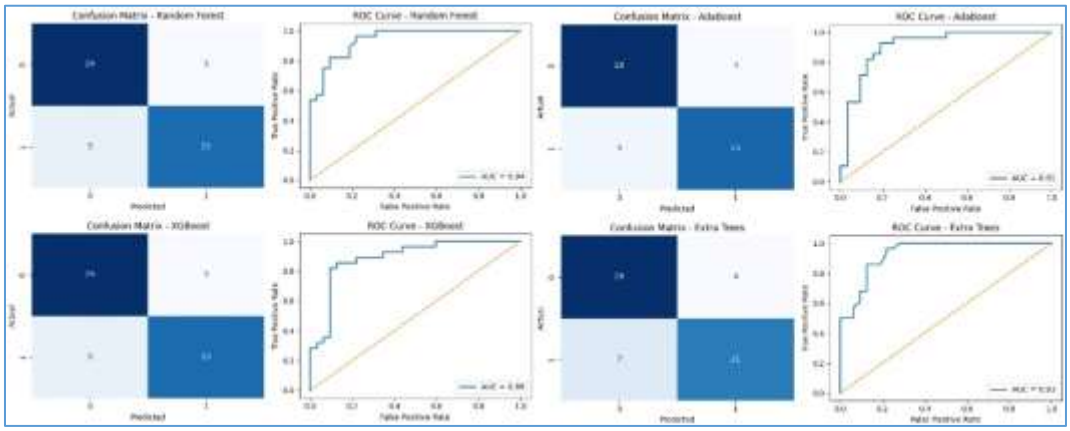
Fitur oldpeak dan thalach juga memberikan kontribusi signifikan, yang berkaitan dengan indikasi iskemia jantung dan kapasitas kardiovaskular pasien. Sebaliknya, fitur age, chol, trestbps, serta fbs menunjukkan pengaruh yang relatif lebih rendah dalam model.

4.6. Analisis Hasil Pemodelan dan Pembahasan

Berdasarkan tabel 4 hasil evaluasi beberapa algoritma, dapat dilihat bahwa ada variasi pada metrik akurasi, AUC, recall, presisi, dan F1-Score untuk setiap metode. Algoritma AdaBoost menunjukkan akurasi tertinggi sebesar 0.8317 dan recall terbaik sebesar 0.8033, sementara nilai presisi dan F1-Score juga cukup kompetitif dengan masing-masing sebesar 0.8333 dan 0.8129. Ketika hyperparameter tuning diterapkan, performa algoritma cenderung bervariasi; misalnya, Random Forest dengan hyperparameter tuning memiliki recall tertinggi 0.7808, tetapi presisinya sedikit menurun menjadi 0.8278 dibandingkan tanpa tuning. Extra Trees dengan tuning menunjukkan peningkatan di hampir semua metrik, khususnya pada AUC sebesar 0.9107 dan recall 0.8093, yang

 ISSN:2302-8149
e-ISSN:2540-9719

berdampak pada F1-Score yang mencapai 0.8122. Sementara itu, XGBoost tanpa tuning memiliki akurasi dan metrik lain yang lebih rendah dibandingkan dengan metode tunning seperti Extra Trees dan AdaBoost. Secara keseluruhan, meskipun hyperparameter tuning tidak selalu menghasilkan peningkatan signifikan di semua metrik, beberapa algoritma seperti Extra Trees dan AdaBoost mendapatkan hasil yang lebih baik setelah tuning dibandingkan tanpa tuning.



Gambar 5 ***Confusion matrix dan ROC Curve***

Berdasarkan Gambar 5 hasil evaluasi menggunakan Confusion Matrix dan ROC Curve, seluruh model ensemble learning menunjukkan performa yang baik dalam memprediksi risiko penyakit jantung. Model Random Forest memberikan kinerja terbaik dengan nilai AUC sebesar 0,94, yang menunjukkan kemampuan pemisahan kelas yang sangat baik, diikuti oleh Extra Trees dengan AUC 0,93 dan AdaBoost dengan AUC 0,91, sementara XGBoost memperoleh AUC 0,89 dan tetap berada pada kategori performa yang baik. Analisis Confusion Matrix menunjukkan bahwa sebagian besar model mampu menghasilkan jumlah prediksi benar yang tinggi dengan tingkat kesalahan klasifikasi yang relatif rendah. Secara keseluruhan, hasil ini menegaskan bahwa pendekatan ensemble learning efektif untuk prediksi penyakit jantung, dengan Random Forest sebagai model paling optimal pada dataset yang digunakan dalam penelitian ini.

5. Kesimpulan

Penelitian ini menunjukkan bahwa optimasi hyperparameter berpengaruh terhadap peningkatan performa model prediksi penyakit jantung, meskipun tingkat peningkatannya bervariasi pada setiap algoritma. Hasil eksperimen memperlihatkan bahwa tiga dari empat algoritma mengalami perbaikan performa setelah proses tuning, dengan Extra Trees menunjukkan kinerja paling stabil dan unggul berdasarkan nilai AUC sebesar 91,07% dan recall 80,93%. Temuan ini menegaskan bahwa pemilihan strategi optimasi yang tepat dapat meningkatkan kemampuan model dalam mendeteksi pasien berisiko, yang merupakan aspek penting dalam sistem pendukung keputusan di bidang kesehatan. Secara analitis, efektivitas optimasi hyperparameter bersifat algoritma-spesifik karena karakteristik masing-masing metode ensemble memengaruhi sensitivitasnya terhadap proses tuning, sehingga memberikan kontribusi pada pemahaman hubungan antara strategi optimasi dan performa model *machine learning* dalam konteks prediksi medis. Meskipun demikian, penelitian ini masih memiliki keterbatasan karena pengujian dilakukan menggunakan dataset publik dengan jumlah fitur dan ruang parameter yang terbatas, sehingga performa model belum sepenuhnya merepresentasikan kondisi klinis nyata; oleh karena itu, penelitian selanjutnya disarankan untuk melakukan validasi menggunakan data pasien riil, memperluas metode optimasi seperti *Genetic Algorithm* atau *Bayesian Optimization*, serta menambahkan teknik *feature engineering* guna meningkatkan kemampuan generalisasi dan akurasi model.

Referensi

[1] "WHO." Accessed: Feb. 10, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] A. Sekarbumi *et al.*, "*Literature Review*: Peran Manajemen Stres dan Pola Hidup Sehat dalam

 ISSN:2302-8149
e-ISSN:2540-9719

mencegah *Diabetes Mellitus* Tipe 2 pada Remaja," *J. Penelit. Inov.*, Vol. 5, No. 2, pp. 2219–2228, 2025, DOI: 10.54082/jupin.1456.

[3] R. Ajartha, J. Sudirman No, L. Pakam, and K. Deli Serdang, "Penanganan Awal Gagal Jantung Akut bagi Tenaga Kesehatan di Instalasi Gawat Darurat," *J. Pengabdi. Kpd. Masy.*, Vol. 4, No. 1, p. 2024, 2024.

[4] N. Chandrasekhar and S. Peddakrishna, "*Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization*," *Processes*, Vol. 11, No. 4, 2023, DOI: 10.3390/pr11041210.

[5] D. Al-fraihat, Y. Sharrab, H. Alshishani, and A. Algarni, "*Hyperparameter Optimization for Software Bug Prediction using Ensemble Learning*," *IEEE Access*, Vol. 12, No. January, pp. 51869–51878, 2024, DOI: 10.1109/ACCESS.2024.3380024.

[6] M. Nurkholifah, Jasmarizal, Y. Umar, and Rahmaddeni, "Analisa Performa Algoritma *Machine Learning* dalam Prediksi Penyakit Liver," *J. Indones. Manaj. Inform. dan Komun.*, Vol. 4, No. 3, pp. 989–998, 2023.

[7] J. Tjen and G. Hoendarto, "*Assessing the Performance of Different Variations of Ensembled Tree Models in Chlorophyll Concentration Prediction*," Vol. 09, No. 0, pp. 6401–6409, 2024.

[8] S. R. Labhsetwar, "*Predictive Analysis of Customer Churn in Telecom Industry using Supervised Learning*," pp. 2054–2060, 2020, DOI: 10.21917/ijsc.2020.0291.

[9] N. Nahar, F. Ara, K. Andersson, and V. Barua, "*A Comparative Analysis of the Ensemble Method for Liver Disease Prediction*," pp. 23–24, 2019.

[10] C. Models, "*Using Sensor Data to Detect Lameness and Mastitis Treatment Events in Dairy Cows : A Comparison of Classification Models*," pp. 1–18, 2020.

[11] R. Wirth and J. Hipp, "*CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39*," *Proc. Fourth Int. Conf. Pract. Appl. Knowl. Discov. Data Min.*, No. 24959, pp. 29–39, 2000.

[12] S. M. Ganie, P. K. D. Pramanik, M. B. Malik, A. Nayyar, and K. S. Kwak, "*An Improved Ensemble Learning Approach for Heart Disease Prediction using Boosting Algorithms*," *Comput. Syst. SCI. Eng.*, Vol. 46, No. 3, pp. 3993–4006, 2023, DOI: 10.32604/csse.2023.035244.

[13] A. H. Yusufi, A. Kharisma, A. D. Adinata, D. F. Ramzy, and M. M. Santoni, "Prediksi Resiko Kematian pada Penderita Penyakit Kadiovaskular menggunakan Metode *Ensemble Learning*," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 531–542, 2022.

[14] C. Chen *et al.*, "*DNN-DTIs: Improved Drug-Target Interactions Prediction using XGBoost Feature Selection and Deep Neural Network*," *Comput. Biol. Med.*, Vol. 136, No. March, p. 104676, 2021, DOI: 10.1016/j.compbiomed.2021.104676.

[15] D. Chicco and G. Jurman, "*The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 Score and Accuracy in Binary Classification Evaluation*," *BMC Genomics*, Vol. 21, No. 1, pp. 1–14, 2020, DOI: 10.1186/s12864-019-6413-7.

[16] K. Hajian-Tilaki, "*Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation*," *Casp. J. Intern. Med.*, Vol. 4, No. 2, pp. 627–635, 2013.

[17] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "*The Impact of Class Imbalance in Classification Performance Metrics based on the Binary Confusion Matrix*," *Pattern Recognit.*, Vol. 91, pp. 216–231, 2019, DOI: 10.1016/j.patcog.2019.02.023.

[18] A. Tharwat, "*Classification Assessment Methods*," *Appl. Comput. Informatics*, Vol. 17, No. 1, pp. 168–192, 2018, DOI: 10.1016/j.aci.2018.08.003.