

**IMPLEMENTASI XGBOOST UNTUK KLASIFIKASI
PENYAKIT JANTUNG DENGAN DATASET CLEVELAND
CLINIC HEART DISEASE**

TUGAS AKHIR

Diajukan oleh:

HARIST ISLAMI RIDHA

NIM. 210705038

Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi



**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI AR-RANIRY
BANDA ACEH
2025 M/1447 H**

LEMBAR PERSETUJUAN

IMPLEMENTASI XGBOOST UNTUK KLASIFIKASI PENYAKIT JANTUNG DENGAN DATASET CLEVELAND CLINIC HEART DISEASE

TUGAS AKHIR

Diajukan kepada Fakultas Sains dan Teknologi
Universitas Islam Negeri Ar-Raniry Banda Aceh
Sebagai Salah Satu Beban Studi Memperoleh Gelar Sarjana (S1)
Dalam Prodi Teknologi Informasi

Oleh.

HARIST ISLAMI RIDHA

NIM. 210705038

**Mahasiswa Fakultas Sains dan Teknologi
Program Studi Teknologi Informasi**

Disetujui Untuk Munaqasyah Oleh:

Pembimbing I,

Pembimbing II,

Malahayati, M.T

NIP.198301272015032003

Aga Maulana, M.Kom

جامعة الرانيري
Mengetahui,

Ketua Program Studi Teknologi Informasi

Malahayati, M.T

NIP.198301272015032003

LEMBAR PENGESAHAN

IMPLEMENTASI XGBOOST UNTUK KLASIFIKASI PENYAKIT JANTUNG DENGAN DATASET CLEVELAND CLINIC HEART DISEASE

TUGAS AKHIR

Telah Diuji Oleh Panitia Ujian Munaqasyah Tugas Akhir
Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh dan Dinyatakan Lulus
Serta Diterima Sebagai Salah Satu Beban Studi Program Sarjana (S-1)
Dalam Program Studi Teknologi Informasi

Pada Hari/Tanggal: Rabu, 6 Agustus 2025

12 Safar 1447 H

Di Darussalam, Banda Aceh

Panitia Ujian Munaqasyah Tugas Akhir:

Ketua,

Malahayati, M.T.
NIP. 198301272015032003

Sekretaris,

Aga Maulana, M.Kom.

Penguji I

Khairan AR, M.Kom.
NIP. 198607042014031001

Penguji II,

Mulkan Fadhli, M.T.
NIP.198811282020121006

UIN
جامعة الرانيري
AR - RANIRY

Mengetahui:

Dekan Fakultas Sains dan Teknologi

UIN Ar-Raniry Banda Aceh,



Prof. Dr. Ir. Muhammad Dirhamsyah, M.T., I.P.U.
NIP.196210021988111001

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Harist Islami Ridha

NIM : 210705038

Program Studi : Teknologi Informasi

Fakultas : Sains dan Teknologi

Implementasi XGBoost Untuk Klasifikasi Penyakit Jantung

Judul : Dengan Dataset Cleveland Clinic Heart Disease

Dengan ini menyatakan bahwa dalam penulisan tugas akhir ini, saya:

1. Tidak menggunakan ide orang lain tanpa mampu mengembangkan dan mempertanggungjawabkan;
2. Tidak melakukan plagiasi terhadap naskah karya orang lain;
3. Tidak menggunakan karya orang lain tanpa menyebutkan sumber asli atau tanpa izin pemilik karya;
4. Tidak memanipulasi dan memalsukan data;
5. Mengerjakan sendiri karya ini dan mampu bertanggungjawab atas karya ini.

Bila dikemudian hari ada tuntutan dari pihak lain atas karya saya, dan telah melalui pembuktian yang dapat dipertanggungjawabkan dan ternyata memang ditemukan bukti bahwa saya telah melanggar pernyataan ini, maka saya siap dikenai sanksi berdasarkan aturan yang berlaku di Fakultas Sains dan Teknologi UIN Ar-Raniry Banda Aceh. Demikian pernyataan ini saya buat dengan sesungguhnya dan tanpa paksaan dari pihak manapun.

Banda Aceh, 6 Agustus 2025

Yang Menyatakan


Harist Islami Ridha

KATA PENGANTAR

Bismillahirrahmanirrahim,

Segala puji bagi Allah SWT atas limpahan rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan Tugas Akhir yang berjudul **“Implementasi XGBoost untuk Klasifikasi Penyakit Jantung dengan Dataset Cleveland Clinic Heart Disease”**. Shalawat serta salam semoga senantiasa tercurah kepada junjungan kita, Nabi Muhammad SAW, beserta keluarga, para sahabat, dan seluruh pengikutnya yang telah membawa umat manusia ke jalan kebenaran.

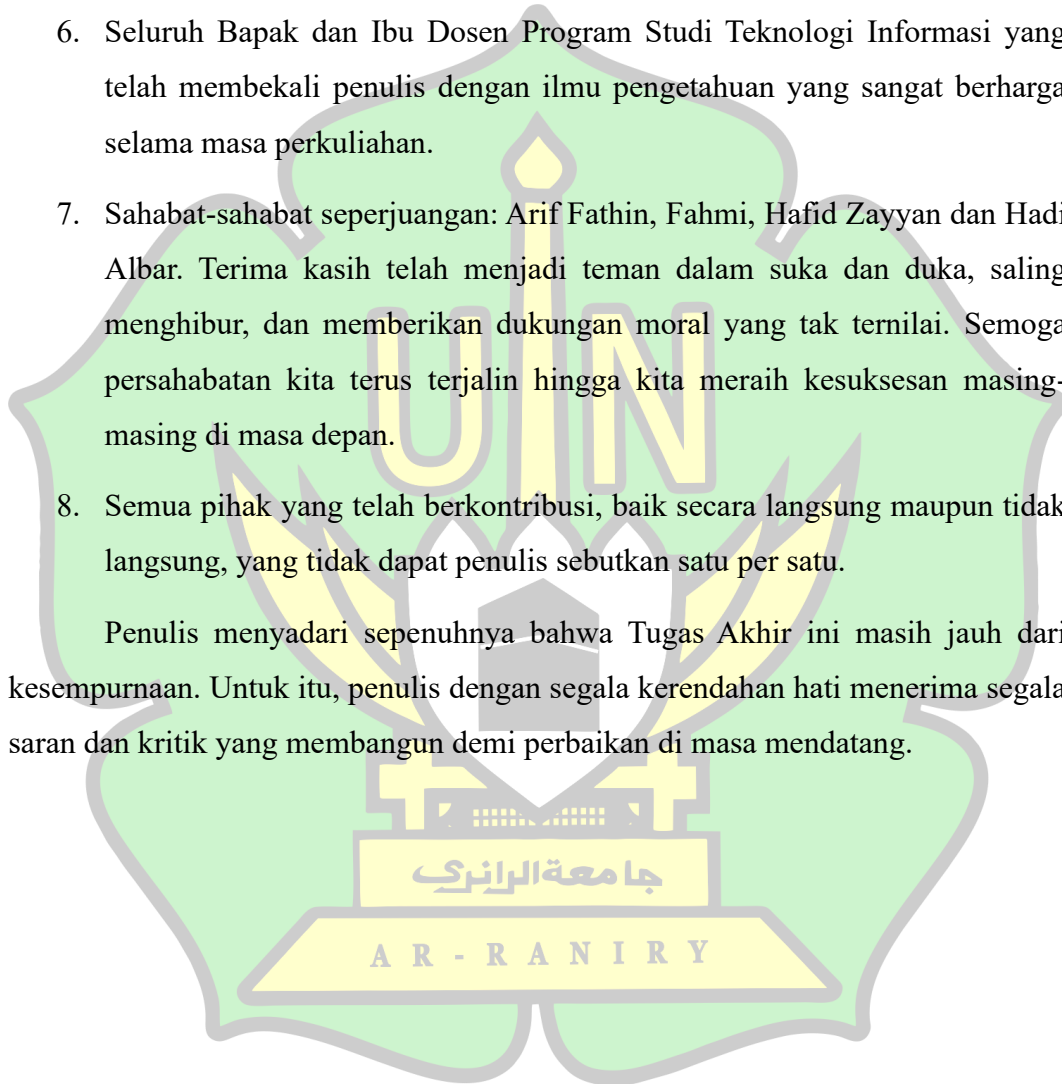
Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana (S1) pada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Ar-Raniry.

Penyelesaian Tugas Akhir ini tidak akan terwujud tanpa bimbingan, dukungan, serta doa tulus dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati dan rasa syukur yang mendalam, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Cinta pertama dan sumber inspirasi saya, Ibunda Cut Idawati, serta pahlawan dan panutan hidup saya, Ayahanda Sarino. Terima kasih atas kasih sayang yang tak terbatas, doa yang tak pernah putus, serta dukungan moril dan materiel yang menjadi sumber kekuatan terbesar bagi penulis dalam menjalani studi hingga tahap akhir ini. Semoga Allah SWT senantiasa melimpahkan kesehatan dan kebahagiaan untuk Ayah dan Bunda.
2. Ibu Malahayati, M.T., yang dengan penuh dedikasi menjalankan amanah sebagai Ketua Program Studi Teknologi Informasi sekaligus Pembimbing I penulis. Terima kasih yang tak terhingga atas bimbingan akademis, arahan, kritik, saran, serta waktu yang selalu diluangkan di tengah kesibukan Beliau.
3. Bapak Aga Maulana, M.Kom, selaku Pembimbing II, atas bimbingan, masukan, dan diskusi konstruktif yang sangat membantu dalam mengarahkan dan menyempurnakan penelitian ini.

4. Bapak Dr. Zainal Abidin, M.Pd., selaku Penasihat Akademik, yang senantiasa memberikan arahan, nasihat, dan motivasi kepada penulis sejak awal masa perkuliahan.
5. Ibu Cut Ida Rahmadiana, S.Si., yang dengan penuh kesabaran telah banyak membantu dalam proses pengurusan berkas administrasi sidang. Bantuan Ibu sangat berarti bagi kelancaran penyelesaian studi penulis.
6. Seluruh Bapak dan Ibu Dosen Program Studi Teknologi Informasi yang telah membekali penulis dengan ilmu pengetahuan yang sangat berharga selama masa perkuliahan.
7. Sahabat-sahabat seperjuangan: Arif Fathin, Fahmi, Hafid Zayyan dan Hadi Albar. Terima kasih telah menjadi teman dalam suka dan duka, saling menghibur, dan memberikan dukungan moral yang tak ternilai. Semoga persahabatan kita terus terjalin hingga kita meraih kesuksesan masing-masing di masa depan.
8. Semua pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari kesempurnaan. Untuk itu, penulis dengan segala kerendahan hati menerima segala saran dan kritik yang membangun demi perbaikan di masa mendatang.



Akhir kata, semoga Tugas Akhir ini dapat bermanfaat bagi penulis, pembaca, dan pengembangan ilmu pengetahuan. Semoga Allah SWT mencatatnya sebagai amal kebaikan.

Aamiin yaa Rabbal'aalamiin.

Banda Aceh, 25 Juli 2025

Penulis,

Harist Islami Ridha



ABSTRAK

Penyakit jantung merupakan salah satu penyebab utama kematian global, sehingga deteksi dini menjadi sangat krusial untuk penanganan yang efektif. Pemanfaatan teknologi *machine learning* menawarkan potensi besar dalam mengembangkan alat deteksi dini yang cepat dan akurat. Penelitian ini mengimplementasikan algoritma *Extreme Gradient Boosting* (XGBoost), sebuah metode *machine learning* berbasis pohon keputusan, untuk membangun model klasifikasi biner penyakit jantung. Tujuan utama dari penelitian ini adalah untuk mengevaluasi performa model XGBoost dan menganalisis sejauh mana pengaruh optimasi *hyperparameter* terhadap peningkatan akurasi dan keandalan model. Penelitian ini menggunakan *dataset Cleveland Clinic Heart Disease* yang terdiri dari 303 sampel dan 13 fitur klinis. Model dievaluasi dalam dua tahap: sebelum (baseline) dan sesudah optimasi *hyperparameter* menggunakan metode *RandomizedSearchCV*. Hasil penelitian menunjukkan peningkatan performa yang sangat signifikan setelah optimasi. Model baseline mencapai akurasi 87% dan AUC 0.94. Setelah dioptimalkan, model akhir berhasil mencapai akurasi sebesar 95%, presisi 96%, recall 93%, F1-score 95% untuk kelas penyakit, dan *Area Under Curve* (AUC) sebesar 0.98. Kesimpulan dari penelitian ini adalah algoritma XGBoost sangat efektif untuk klasifikasi penyakit jantung, dan proses optimasi *hyperparameter* terbukti menjadi langkah krusial yang secara substansial meningkatkan akurasi, presisi, dan keandalan model prediktif.

Kata Kunci: Penyakit Jantung, XGBoost, *Machine Learning*, Klasifikasi Biner, *Tuning Hyperparameter*, Deteksi Dini.

ABSTRACT

Heart disease is a leading cause of global mortality, making early detection critical for effective treatment. The use of machine learning offers great potential for developing fast and accurate early detection tools. This research implements the Extreme Gradient Boosting (XGBoost) algorithm, a decision-tree-based machine learning method, to develop a binary classification model for heart disease. The primary objectives of this study were to evaluate the performance of the XGBoost model and to analyze the impact of hyperparameter optimization on improving its accuracy and reliability. This study utilized the Cleveland Clinic Heart Disease dataset, which comprises 303 samples and 13 clinical features. The model was evaluated in two stages: before (baseline) and after hyperparameter optimization using the RandomizedSearchCV method. The results demonstrated a significant improvement in performance following optimization. The baseline model achieved an accuracy of 87% and an AUC of 0.94. After optimization, the final model successfully achieved an accuracy of 95%, a precision of 96%, a recall of 93%, an F1-score of 95% for the disease class, and an Area Under Curve (AUC) of 0.98. This study concludes that the XGBoost algorithm is highly effective for heart disease classification, and that the hyperparameter optimization process is a crucial step that substantially enhances the accuracy, precision, and reliability of the predictive model.

Keywords: *Heart Disease, XGBoost, Machine Learning, Binary Classification, Hyperparameter Tuning, Early Detection.*

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR	iii
KATA PENGANTAR	iv
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN	2
1.1 Latar Belakang	2
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II LANDASAN TEORI	6
2.1 Penelitian Terdahulu.....	6
2.2 Penyakit Jantung	9
2.3 Machine learning.....	11
2.4 Dataset Cleveland Heart Disease	12
2.5 Python	13
2.6 Google Colaboratory	14
2.7 Algoritma Klasifikasi: Extreme Gradient Boosting (XGBoost)	14
2.7.1 Arsitektur dan Mekanisme Kerja XGBoost	15

2.7.2	Rumus Utama XGBoost.....	16
2.7.3	Perbedaan Algoritma XGBoost Studi Kasus Jantung dan Kanker	17
2.7.4	Peran Umum XGBoost dalam Bidang Medis	18
2.8	Tuning Hyperparameter	19
2.9	Evaluasi Model.....	20
BAB III METODOLOGI PENELITIAN		22
3.1	Jenis Penelitian.....	22
3.2	Tahapan Penelitian	22
3.3	Pengumpulan Dataset.....	23
3.4	Prapemrosesan Data	24
3.4.1	Pembersihan Data.....	24
3.4.2	Transformasi Data	25
3.5	Split Data.....	26
3.6	Pembangunan Model XGBoost.....	27
3.7	Tuning Hyperparameter	28
3.8	Evaluasi Model.....	30
3.9	Perangkat Keras	31
3.10	Perangkat Lunak.....	32
BAB IV HASIL DAN PEMBAHASAN.....		33
4.1	Hasil Prapemrosesan Data.....	33
4.2	Hasil Pembangunan Model XGBoost	34
4.2.1	Performa Model XGBoost Sebelum Tuning (Baseline).....	34
4.2.2	Proses Tuning Hyperparameter	39
4.2.3	Performa Model XGBoost Setelah <i>Tuning</i>	42
4.3	Analisis dan Pembahasan	47
4.3.1	Analisis Kinerja Model Sebelum dan Sesudah Tuning.....	48

4.3.2	Analisis Kepentingan Fitur (Feature Importance Analysis).....	50
BAB V KESIMPULAN DAN SARAN		52
5.1	Kesimpulan	52
5.2	Saran.....	53
DAFTAR PUSTAKA.....		54



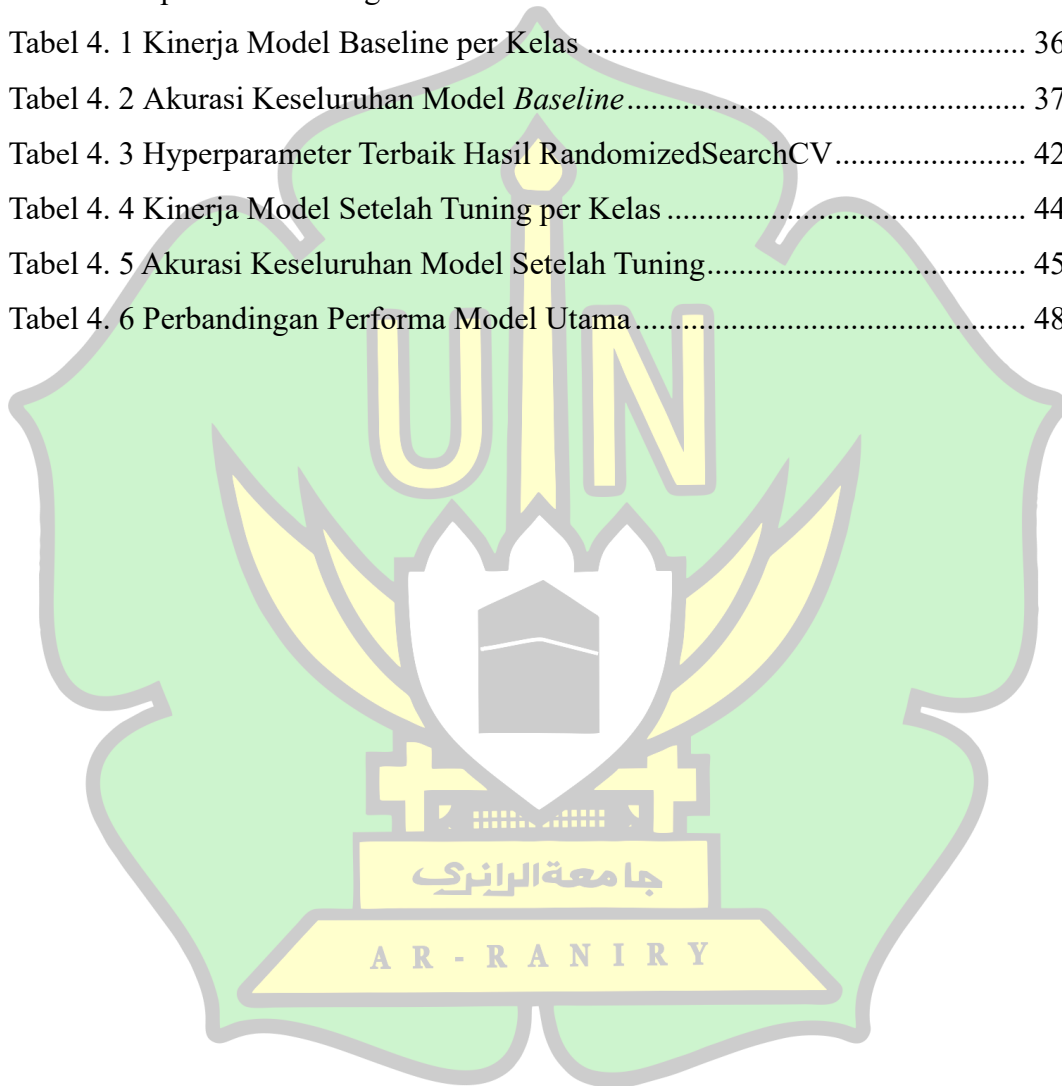
DAFTAR GAMBAR

Gambar 2. 1	Arsitektur umum algoritma XGBoost.....	15
Gambar 3. 1	Proses Tahapan Penelitian.....	23
Gambar 4. 1	Confusion Matrix Sebelum Tuning.....	38
Gambar 4. 2	Confusion Matrix Setelah Tuning.....	46
Gambar 4. 3	Fitur Paling Berpengaruh (Top 10 by Gain).....	50



DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu.....	8
Tabel 3. 1 Ruang Pencarian <i>Hyperparameter</i> XGBoost.....	30
Tabel 3. 2 Spesifikasi Perangkat Keras	31
Tabel 3. 3 Spesifikasi Perangkat Lunak	32
Tabel 4. 1 Kinerja Model Baseline per Kelas	36
Tabel 4. 2 Akurasi Keseluruhan Model <i>Baseline</i>	37
Tabel 4. 3 Hyperparameter Terbaik Hasil <i>RandomizedSearchCV</i>	42
Tabel 4. 4 Kinerja Model Setelah Tuning per Kelas	44
Tabel 4. 5 Akurasi Keseluruhan Model Setelah Tuning.....	45
Tabel 4. 6 Perbandingan Performa Model Utama.....	48



BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit jantung merupakan salah satu penyebab kematian tertinggi secara global. Menurut *World Health Organization* (WHO), penyakit kardiovaskular bertanggung jawab atas sekitar 17,9 juta kematian setiap tahunnya, yang mencakup 31% dari total kematian dunia (Jitendra P Chaudhari dkk., 2025). Penyakit ini mencakup berbagai kondisi yang mempengaruhi jantung, termasuk penyakit jantung koroner, gagal jantung, dan aritmia. Faktor risiko utama penyakit jantung meliputi hipertensi, kadar kolesterol tinggi, obesitas, serta gaya hidup yang tidak sehat seperti merokok dan kurangnya aktivitas fisik (Theerthagiri P, 2021). Oleh karena itu, deteksi dini penyakit jantung menjadi sangat penting untuk mengurangi angka kematian serta membantu tenaga medis dalam mengambil keputusan yang tepat.

Di Indonesia, penyakit jantung menjadi salah satu penyebab utama kematian dengan prevalensi mencapai 2.650.340 kasus. Berdasarkan data terbaru, sekitar 1.5% dari populasi mengalami penyakit jantung koroner, sedangkan gagal jantung memiliki prevalensi 0.3%. Faktor risiko utama termasuk hipertensi, gaya hidup tidak sehat, dan obesitas, yang semakin meningkat dalam beberapa tahun terakhir (Destarani, 2025). Oleh karena itu, penggunaan algoritma *machine learning* seperti XGBoost diharapkan dapat membantu dalam proses deteksi dini penyakit jantung di Indonesia.

Salah satu tantangan utama dalam penanganan penyakit jantung adalah keterlambatan diagnosis. Banyak pasien baru menyadari bahwa mereka mengidap penyakit jantung ketika kondisinya sudah parah dan sulit ditangani. Deteksi dini menjadi sangat penting untuk menurunkan risiko komplikasi dan meningkatkan peluang hidup pasien. Namun, metode konvensional seperti wawancara medis, pemeriksaan fisik, dan beberapa tes laboratorium sering kali membutuhkan waktu dan sumber daya yang tidak sedikit. Oleh karena itu, dibutuhkan pendekatan yang

lebih efisien dan terukur untuk membantu proses diagnosis secara cepat dan akurat (Shahid, 2025).

Salah satu algoritma *machine learning* yang populer dan efektif adalah *Extreme Gradient Boosting* (XGBoost). XGBoost merupakan pengembangan dari teknik *boosting*, yang bekerja dengan menggabungkan sejumlah model pohon keputusan sederhana (*weak learners*) secara bertahap untuk membentuk model prediktif yang kuat. Algoritma ini dirancang dengan fokus pada efisiensi komputasi, mendukung paralelisasi proses pelatihan, dan dilengkapi dengan mekanisme regularisasi untuk mengurangi risiko *overfitting*. Berkat kemampuannya dalam menangani data *non-linear* dan skala besar, XGBoost banyak digunakan dalam berbagai bidang, termasuk deteksi dan prediksi penyakit (Jafari dkk., 2024).

Seiring dengan perkembangan teknologi, metode berbasis kecerdasan buatan dan *machine learning* mulai digunakan dalam diagnosis penyakit jantung. Metode ini memungkinkan analisis data yang lebih cepat dan akurat dibandingkan dengan metode konvensional (Hajiarbabi, 2024). Pemilihan XGBoost dalam penelitian ini didasarkan pada keunggulannya dibandingkan *Random Forest* dan SVM dalam klasifikasi penyakit jantung. XGBoost menggunakan *boosting*, yang memungkinkan model memperbaiki kesalahan prediksi secara iteratif, meningkatkan akurasi, dan mengurangi *overfitting* melalui regularisasi L1 dan L2. Selain itu, XGBoost lebih efisien secara komputasi karena mendukung paralelisasi dan optimasi memori, membuatnya lebih cepat dalam menangani *dataset* besar. Hajiarbabi dalam penelitiannya menunjukkan bahwa XGBoost memiliki akurasi lebih tinggi dibandingkan SVM dalam klasifikasi penyakit jantung berbasis ECG, menjadikannya pilihan unggul dalam analisis data medis.

Dataset yang digunakan adalah *Cleveland Heart Disease Dataset*, yang merupakan salah satu *dataset benchmark* dalam penelitian diagnosis penyakit jantung. *Dataset* ini terdiri dari berbagai *parameter* klinis seperti tekanan darah, kadar kolesterol, detak jantung maksimal, serta hasil pemeriksaan elektrokardiogram. *Dataset* ini sering digunakan dalam penelitian karena ketersediaannya yang terbuka dan struktur datanya yang cukup *representative* (Hajiarbabi, 2024).

Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma XGBoost dalam klasifikasi biner penyakit jantung dan menganalisis sejauh mana proses *tuning hyperparameter* dapat meningkatkan performa model. Evaluasi dilakukan menggunakan metrik performa seperti *accuracy*, *precision*, *recall*, *F1-score*, dan *Area Under Curve* (AUC). Dengan pendekatan ini, diharapkan hasil penelitian dapat memberikan kontribusi dalam pengembangan model prediksi berbasis data klinis yang sederhana, efisien, dan cukup akurat untuk mendukung deteksi dini penyakit jantung.

Selain itu, penelitian ini diharapkan dapat menjadi landasan awal untuk pengembangan sistem pendukung keputusan medis berbasis *machine learning*, khususnya pada kasus-kasus dengan sumber daya terbatas. Dengan algoritma yang relatif ringan dan tidak terlalu kompleks, model yang dikembangkan dalam penelitian ini dapat menjadi dasar awal untuk pengembangan sistem serupa di masa depan, tanpa mengkaji aspek implementasi langsung dalam lingkungan klinis.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, penelitian ini berfokus pada beberapa permasalahan utama:

1. Bagaimana performa XGBoost dalam melakukan klasifikasi biner untuk mendeteksi penyakit jantung?
2. Bagaimana pengaruh optimasi *hyperparameter* terhadap performa model XGBoost dalam klasifikasi penyakit jantung?

1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah:

1. Menganalisis performa algoritma XGBoost dalam klasifikasi biner untuk mendeteksi penyakit jantung.
2. Mengoptimalkan *hyperparameter* XGBoost guna meningkatkan hasil klasifikasi berdasarkan evaluasi akurasi dan metrik terkait.

1.4 Batasan Penelitian

Untuk memastikan penelitian ini tetap fokus dan terarah, beberapa batasan yang diterapkan adalah:

1. Jenis klasifikasi yang dilakukan adalah klasifikasi biner, yaitu membedakan antara pasien yang memiliki penyakit jantung dan yang tidak, tanpa mempertimbangkan tingkat keparahan penyakit.
2. *Dataset* yang digunakan adalah *dataset Cleveland Heart Disease* dari UCI *Machine learning Repository*, yang terdiri dari data klinis pasien dengan 13 fitur utama dan 1 variabel target.
3. Model yang digunakan adalah algoritma XGBoost, tanpa membandingkannya secara mendalam dengan algoritma klasifikasi lainnya.
4. Evaluasi performa model terbatas pada metrik *accuracy*, *precision*, *recall*, *F1-score*, dan AUC, tanpa mengkaji aspek implementasi dalam sistem klinis nyata.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi dunia medis, hasil penelitian ini dapat mendukung proses deteksi dini penyakit jantung dengan menyediakan model prediksi yang akurat, sehingga dapat membantu tenaga medis dalam mengambil keputusan secara lebih cepat dan tepat.
2. Bagi bidang ilmu data dan kecerdasan buatan, penelitian ini dapat menjadi referensi dalam penerapan algoritma XGBoost untuk klasifikasi biner pada *dataset* medis, khususnya dalam konteks optimasi performa model melalui *tuning hyperparameter*.
3. Bagi penulis, penelitian ini menjadi sarana untuk memperdalam pemahaman mengenai algoritma *machine learning*, khususnya dalam proses pelatihan, evaluasi, serta peningkatan performa model klasifikasi penyakit berdasarkan data klinis.