

**OPTIMASI PREDKSI RISIKO PENYAKIT JANTUNG
MENGUNAKAN METODE HYBRID GRADIENT BOOSTIG
BERBASIS KOMBINASI MODEL XGBOOST DAN LIGHTGBM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

JOE FAUZSANNNO RETTOB

22.11.4794

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**OPTIMASI PREDKSI RISIKO PENYAKIT JANTUNG
MENGUNAKAN METODE HYBRID GRADIENT BOOSTIG
BERBASIS KOMBINASI MODEL XGBOOST DAN LIGHTGBM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

JOE FAUZSANNO RETTOB

22.11.4794

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**OPTIMASI PREDKSI RISIKO PENYAKIT JANTUNG
MENGUNAKAN METODE HYBRID GRADIENT BOOSTIG
BERBASIS KOMBINASI MODEL XGBOOST DAN LIGHTGBM**

yang disusun dan diajukan oleh

Joe Fauzsanno Rettob

22.11.4794

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Januari 2026

Dosen Pembimbing,



Dr. Emigawaty S.Kom., M.Kom

NIK. 190302226

HALAMAN PENGESAHAN

SKRIPSI

**OPTIMASI PREDKSI RISIKO PENYAKIT JANTUNG
MENGUNAKAN METODE HYBRID GRADIENT BOOSTIG
BERBASIS KOMBINASI MODEL XGBOOST DAN LIGHTGBM**

yang disusun dan diajukan oleh

Joe Fauzsanno Rettob
22.11.4794

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Januari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
<u>Ike Verawati, S.Kom., M.Kom</u> NIK. 190302237	
<u>Ria Andriani, S.Kom., M.Kom</u> NIK. 190302458	
<u>Dr. Emigawaty, S.Kom., M.Kom</u> NIK. 190302226	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Joe Fauzsanno Rettob
NIM : 22.11.4794

Menyatakan bahwa Skripsi dengan judul berikut:

**Optimasi Predksi Risiko Penyakit Jantung Menggunakan Metode
Hybrid Gradient Boostig berbasisi Kombinasi Model Xgboost Dan
Lightgbm**

Dosen Pembimbing : Dr. Emigawaty S.Kom., M.Kom

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

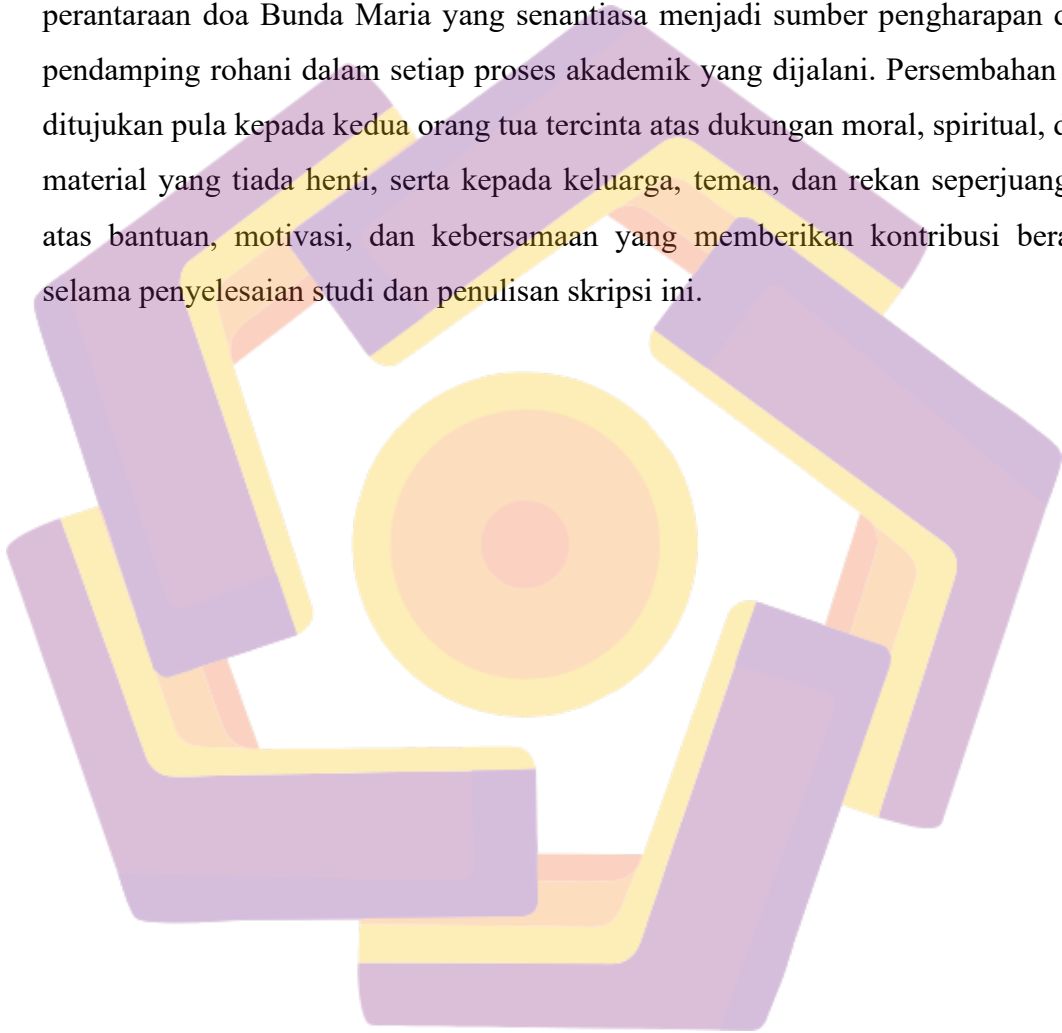
Yogyakarta, 20 Januari 2026

Yang Menyatakan,


Joe Fauzsanno Rettob

HALAMAN PERSEMBAHAN

Dengan segala keredahan hati dan rasa syukur, penulis mempersembahkan karya ilmiah ini kepada Tuhan Yesus Kristus, yang telah menganugerahkan kekuatan, hikmat, dan penyertaan hingga terselesaikannya skripsi ini, serta melalui perantaraan doa Bunda Maria yang senantiasa menjadi sumber pengharapan dan pendamping rohani dalam setiap proses akademik yang dijalani. Persembahan ini ditujukan pula kepada kedua orang tua tercinta atas dukungan moral, spiritual, dan material yang tiada henti, serta kepada keluarga, teman, dan rekan seperjuangan atas bantuan, motivasi, dan kebersamaan yang memberikan kontribusi berarti selama penyelesaian studi dan penulisan skripsi ini.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Optimasi Prediksi Risiko Penyakit Jantung Menggunakan Metode Hybrid Gradient Boosting Berbasis Kombinasi Model XGBoost dan LightGBM”. Penyusunan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi yang penulis tempuh. Dalam penyusunan skripsi ini, penulis menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. M. Suysnto, MM, selaku Rektor Universitas yang telah memberikan kesempatan dan fasilitas selama penulis menempuh pendidikan.
2. Prof. Dr. Kusrini, M.Kom, selaku Dekan Fakultas Ilmu Komputer yang telah memberikan dukungan dan layanan akademik bagi penulis selama masa studi.
3. Eli Pujastuti, M.Kom, yang telah memberikan arahan dan dukungan selama penulis mengikuti proses perkuliahan.
4. Dr. Emigawaty S.Kom., M.Kom , yang telah memberikan bimbingan, masukan, serta motivasi secara berkesinambungan sehingga skripsi ini dapat terselesaikan dengan baik.
5. Dosen Program Studi S1 Informatika, yang telah memberikan ilmu, wawasan, dan pengalaman berharga selama perkuliahan dan tidak dapat penulis sebutkan satu per satu, namun kontribusinya sangat berarti dalam penyelesaian skripsi ini.

Yogyakarta, 21 Januari 2026

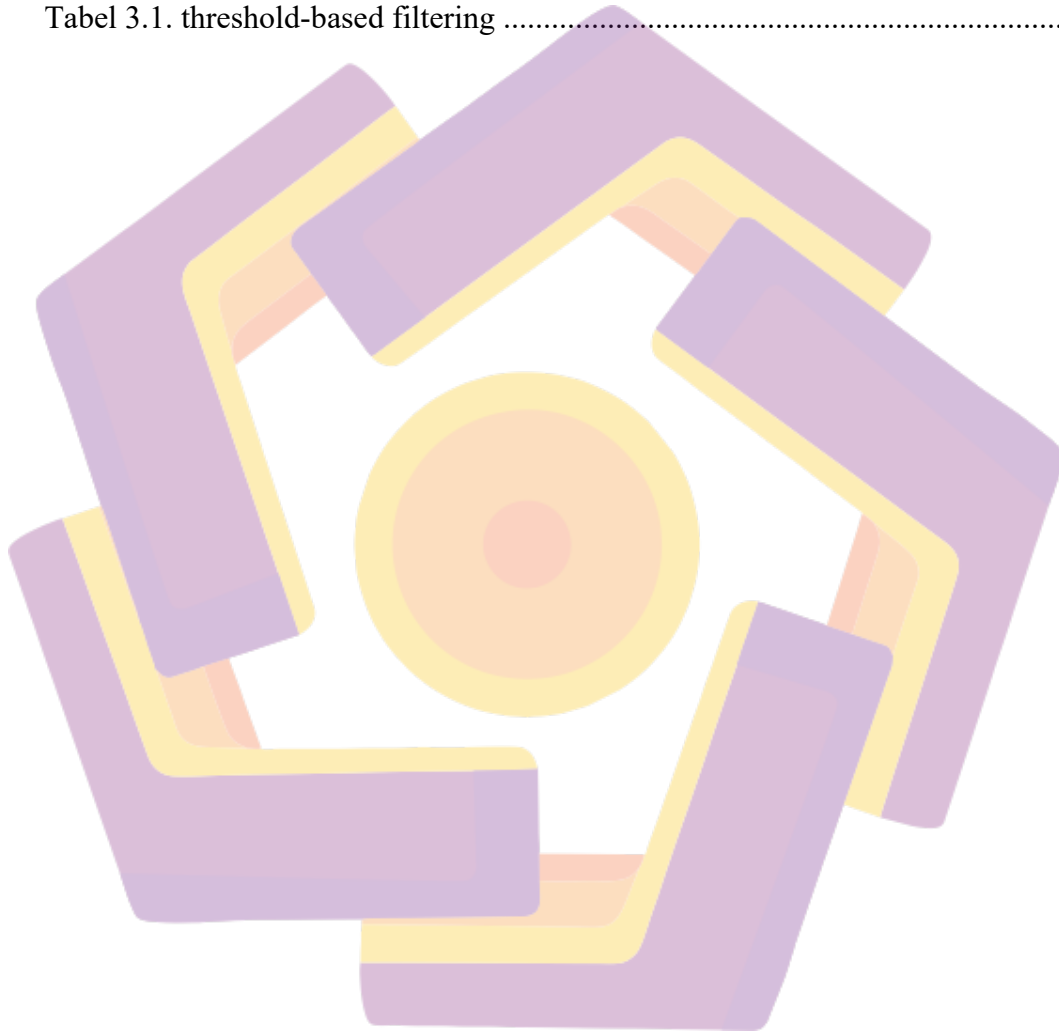
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMBANG DAN SINGKATAN	x
DAFTAR ISTILAH.....	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	18
BAB III METODE PENELITIAN	24
3.1 Objek Penelitian.....	24
3.2 Alur Penelitian	25
3.3 Alat dan Bahan.....	35
BAB IV HASIL DAN PEMBAHASAN	36
BAB V PENUTUP	60
5.1 Kesimpulan	60
5.2 Saran	61
REFERENSI	62

DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian	1
Tabel 3.1. Bahan	2
Tabel 3.2. Alat	3
Tabel 4.1. Deskripsi Data	4
Tabel 3.1. threshold-based filtering	5



DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	1
Gambar 4.1. Akuisisi Data	2
Gambar 4.2. Data Info	3
Gambar 4.3. Heatmap Korelasi antar Fitur	4
Gambar 4.4. Distribusi Kelas Target	5
Gambar 4.5. Feature Engineering dan Normalisasi Data.....	6
Gambar 4.6. Feature Selection	7
Gambar 4.7. Split Data	8
Gambar 4.8. Hasil Evaluasi XGBoost.....	9
Gambar 4.9. Hasil Evaluasi LightGBM.....	10
Gambar 4.10. hasil cross Validation.....	11
Gambar 4.11 hasil Optimisasi Optuna.....	12
Gambar 4.12 Hasil Evaluasi Voting Classifier.....	13
Gambar 4.13. Evaluasi Model Hybrid Gradien Boosting dan confusion Matrix Hybrid Gradient Boosting	9
Gambar 4.14. Intepretasi (SHAP)	10
Gambar 4.15. Hasil Deploy Prediksi Penyakit Jantung	11

DAFTAR LAMBANG DAN SINGKATAN

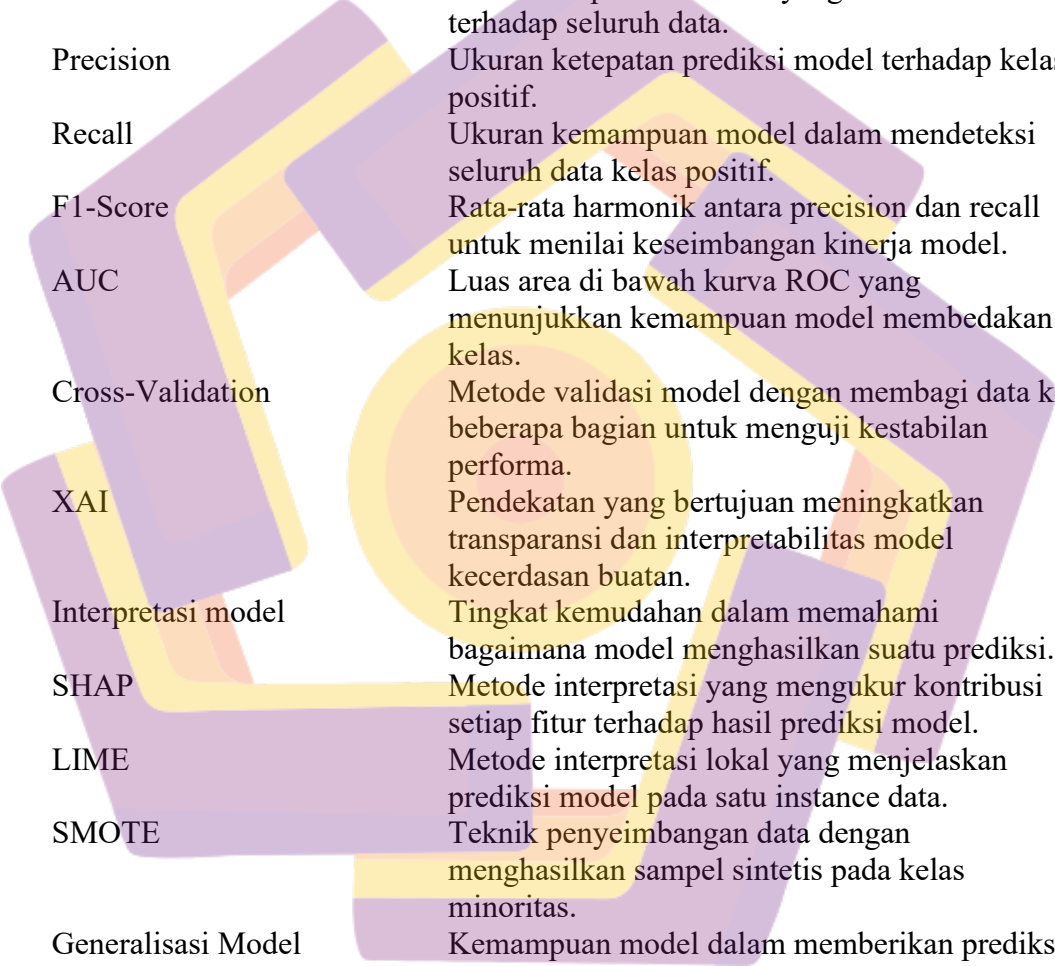
BMI	<i>Body Mass Index</i> , indeks massa tubuh
Kg	Kilogram, satuan berat
m	Meter, satuan tinggi
x	Matriks fitur (variabel independen)
y	y label atau variabel target
RobutScaler	Metode normalisasi berbasis median dan IQR
Median(x)	Nilai Tengah dari data X
IQR	<i>Interquartile Range</i> , selisih antara Q3 dan Q1
Q1	Kuartil pertama
Q3	Kuartil ketiga
X_scaled	Data fitur asli normalisasi
SelectFromModel	Metode seleksi fitur berbasis model
Feature Importance	Ukuran kontribusi fitur terhadap performa model
ΔL	Penurunan nilai fungsi (<i>loss</i>)
f	Fitur dalam dataset
T	Jumlah pohon keputusan dalam model
N	Jumlah total data
X_selected	Data fitur hasil seleksi fitur
y_res	Variabel target hasil proses resampling
Test_size	Proporsi data uji
Stratified Sampling	Teknik pembagian data berbasis proporsi kelas
Random_state	Parameter pengendali reproduktibilitas
$\hat{y}$	Nilai prediksi model
$f_k(x_i)$	Fungsi pohon ke-k pada data ke-i
Obj	Fungsi objektif model
$L(y, \hat{y})$	Fungsi kerugian (<i>loss function</i>)
$\Omega(f)$	Fungsi regularisasi
Log loss	Fungsi kerugian untuk klasifikasi biner
g_i	Gradien orde pertama
h_i	Hessian atau gradien orde kedua
w_j	Bobot optimal pada daun pohon
λ	Parameter regularisasi.
Learning_rate	Parameter laju pembelajaran
num_leaves	Jumlah daun maksimum pada LightGBM
feature_fraction	Proporsi fitur yang digunakan
bagging_fraction	Proporsi data yang digunakan
Cross Validation	Metode evaluasi berbasis validasi silang
K	Jumlah fold pada cross validation
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
TPR	True Positive Rate
FPR	False Positive Rate
TP	True Positif
FP	False Positif

TN	True Pegative
FN	False Negative
Optuna	Framework optimasi hyperparameter
Bayesian Optimization	Metode optimasi berbasis probablistik
n_estimators	Jumlah pohon Keputusan
max_depth	Kedalam maksimum pohon
subsample	Proposisi data untuk setiap iterasi
colsample_bytree	Proposisi fitur pada setiap pohon
VotingClassifier	Metode ensemble berbasis voting
Soft Voting	Metode ensemble berbasis rata-rata probabilitas
P_XGB	Probabilitas prediksi XGBoost
P_LGBM	Probabilitas prediksi LightGBM
P_hybrid	Porbabilitas prediksi model hybrid
WHO	<i>World Health Organization</i> , organisasi kesehatan dunia.
CVD	<i>Cardiovascular Disease</i> atau penyakit kardiovaskular.
ML	<i>Machine Learning</i> , cabang kecerdasan buatan untuk pembelajaran mesin.
XGBoost	<i>Extreme Gradient Boosting</i> , algoritma ensemble berbasis boosting.
LightGBM	<i>Light Gradient Boosting Machine</i> , algoritma gradient boosting berbasis pohon.
RF	<i>Random Forest</i> , algoritma ensemble berbasis bagging.
DT	<i>Decision Tree</i> , algoritma klasifikasi berbasis pohon Keputusan.
MLP	<i>Multilayer Perceptron</i> , jaringan saraf tiruan dengan banyak lapisan.
LR	<i>Logistic Regression</i> , algoritma klasifikasi berbasis model linear.
AUC	<i>Area Under the Curve</i> , metrik evaluasi performa model klasifikasi.
XAI	<i>Explainable Artificial Intelligence</i> , pendekatan untuk meningkatkan interpretabilitas model.
SHAP	<i>SHapley Additive exPlanations</i> , metode interpretasi model berbasis nilai kontribusi fitur.
LIME	<i>Local Interpretable Model-agnostic Explanations</i> , metode interpretasi prediksi model
SMOTE	<i>Synthetic Minority Over-sampling Technique</i> , teknik penyeimbangan data tidak seimbang
Esembel Learning	Metode penggabungan beberapa model untuk meningkatkan kinerja prediksi
Gardiant Boosting	Model ensemble yang menggabungkan lebih dari satu jenis algoritma

Cross-Validation	Teknik ensemble berbasis pembelajaran bertahap untuk meminimalkan kesalahan
F1-Score	Teknik validasi silang untuk mengevaluasi kestabilan model
Recal	Metrik evaluasi yang menggabungkan precision dan recall
Precision	Metrik evaluasi yang mengukur ketepatan prediksi kelas positif
Accuracy	Metrik evaluasi yang mengukur tingkat ketepatan prediksi secara keseluruhan
Age	Usia pasien
Gender	Jenis kelamin pasien
Height	Tinggi badan pasien
weight	Berat badan pasien
Ap_hi	Tekanan darah sistolik
Ap_lo	Tekanan darah diastolik
Cholesterol	Kadar kolesterol
Gluc	Kadar glukosa darah
Smoke	Kebiasaan merokok
Alco	Konsumsi alkohol
active	Aktifitas fisik
cardio	Penyakit jantung (label target)

DAFTAR ISTILAH

Peyakit cardiovascular	Kelompok penyakit yang meliputi gangguan pada jantung dan sistem peredaran darah, termasuk serangan jantung dan stroke.
Deteksi dini	Upaya identifikasi risiko penyakit sebelum muncul gejala klinis yang lebih serius.
Klasifikasi	Proses pengelompokan data ke dalam kelas tertentu berdasarkan karakteristik atau pola yang dipelajari model.
Machine Learning	Metode komputasi yang memungkinkan sistem belajar dari data tanpa diprogram secara eksplisit.
Dataset medis	Kumpulan data kesehatan pasien yang berisi informasi klinis dan fisiologis untuk keperluan analisis dan pemodelan.
Model Tunggal	Model prediksi yang hanya menggunakan satu algoritma pembelajaran mesin.
Ensemble Learning	Pendekatan pembelajaran mesin yang menggabungkan beberapa model untuk meningkatkan akurasi dan stabilitas prediksi.
Gradient Boosting	Teknik ensemble yang membangun model secara bertahap untuk memperbaiki kesalahan prediksi sebelumnya.
Hybrid Esemble	Model ensemble yang mengombinasikan dua atau lebih algoritma berbeda dalam satu sistem prediksi.
Hybrid Gradient Boosting	Model ensemble yang menggabungkan algoritma XGBoost dan LightGBM untuk meningkatkan performa klasifikasi.
Overviting	Kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga performanya menurun pada data baru.
Multikolinearitas	Kondisi ketika antar variabel independen saling berkorelasi tinggi sehingga memengaruhi kestabilan model.
Noise Data	Gangguan atau ketidakkonsistenan pada data yang dapat menurunkan kualitas pembelajaran model.
Imbalance Data	Kondisi ketidakseimbangan jumlah data antar kelas, misalnya antara pasien sehat dan pasien berpenyakit.
Preprocessing Data	Tahapan awal pengolahan data sebelum pemodelan, termasuk pembersihan, normalisasi, dan transformasi fitur.



Normalisasi Data	Proses penskalaan data agar berada pada rentang tertentu untuk meningkatkan kinerja model.
Hyperparameter	Parameter yang ditentukan sebelum proses pelatihan model untuk mengatur perilaku algoritma.
Optimasi Hyperparameter	Proses pencarian kombinasi hyperparameter terbaik untuk meningkatkan performa model.
Evaluasi model	Proses pengukuran kinerja model menggunakan metrik tertentu seperti akurasi dan AUC.
Akurasi	Persentase prediksi benar yang dihasilkan model terhadap seluruh data.
Precision	Ukuran ketepatan prediksi model terhadap kelas positif.
Recall	Ukuran kemampuan model dalam mendeteksi seluruh data kelas positif.
F1-Score	Rata-rata harmonik antara precision dan recall untuk menilai keseimbangan kinerja model.
AUC	Luas area di bawah kurva ROC yang menunjukkan kemampuan model membedakan kelas.
Cross-Validation	Metode validasi model dengan membagi data ke beberapa bagian untuk menguji kestabilan performa.
XAI	Pendekatan yang bertujuan meningkatkan transparansi dan interpretabilitas model kecerdasan buatan.
Interpretasi model	Tingkat kemudahan dalam memahami bagaimana model menghasilkan suatu prediksi.
SHAP	Metode interpretasi yang mengukur kontribusi setiap fitur terhadap hasil prediksi model.
LIME	Metode interpretasi lokal yang menjelaskan prediksi model pada satu instance data.
SMOTE	Teknik penyeimbangan data dengan menghasilkan sampel sintetis pada kelas minoritas.
Generalisasi Model	Kemampuan model dalam memberikan prediksi yang baik pada data baru yang belum pernah dilatih.

INTISARI

Penyakit jantung merupakan salah satu penyebab utama kematian dengan angka prevalensi yang terus meningkat dan berdampak signifikan terhadap kualitas hidup masyarakat serta beban sistem kesehatan. Keterlambatan dalam mendeteksi risiko penyakit jantung menyebabkan penanganan medis sering dilakukan pada tahap lanjut, sehingga diperlukan pendekatan prediktif yang mampu mengidentifikasi risiko secara dini dan akurat berbasis data klinis dan demografis. Penelitian ini bertujuan untuk merancang dan mengimplementasikan model Hybrid Gradient Boosting berbasis kombinasi algoritma XGBoost dan LightGBM guna mengoptimalkan performa prediksi risiko penyakit jantung. Tahapan penelitian meliputi pengumpulan dataset Heart Attack Indonesia, preprocessing data, exploratory data analysis (EDA), feature engineering, pembagian data latih dan data uji, pelatihan model XGBoost dan LightGBM, serta evaluasi kinerja model menggunakan metrik akurasi, precision, recall, dan F1-score. Selain itu, dilakukan cross-validation untuk menguji kemampuan generalisasi model serta analisis feature importance dan SHAP untuk meningkatkan interpretabilitas model. Hasil penelitian menunjukkan bahwa pendekatan Hybrid Gradient Boosting mampu menghasilkan performa prediksi yang optimal, stabil, dan memiliki kemampuan generalisasi yang baik. Variabel usia, tekanan darah, kadar kolesterol, dan indeks massa tubuh (BMI) terbukti berpengaruh signifikan terhadap risiko penyakit jantung. Penelitian ini berkontribusi dalam pengembangan model prediksi yang akurat dan dapat diinterpretasikan, serta dapat dimanfaatkan oleh tenaga medis, institusi kesehatan, dan peneliti sebagai alat bantu pengambilan keputusan dan pengembangan sistem pendukung diagnosis penyakit jantung.

Kata kunci: penyakit jantung, hybrid gradient boosting, XGBoost, LightGBM, prediksi risiko.

ABSTRACT

Heart disease is one of the leading causes of death, with a continuously increasing prevalence that significantly affects public quality of life and imposes a substantial burden on healthcare systems. Delays in early risk detection often result in medical treatment being conducted at advanced stages of the disease. Therefore, a predictive approach is required to accurately and reliably identify the risk of heart disease at an early stage based on clinical and demographic data. This study aims to design and implement a Hybrid Gradient Boosting model based on a combination of XGBoost and LightGBM algorithms to optimize the performance of heart disease risk prediction. The research methodology includes data collection from the Heart Attack Indonesia dataset, data preprocessing, exploratory data analysis (EDA), feature engineering, data partitioning into training and testing sets, model training using XGBoost and LightGBM, and performance evaluation using accuracy, precision, recall, and F1-score metrics. Cross-validation is also applied to assess the generalization capability of the proposed model, while feature importance analysis and SHAP are utilized to enhance model interpretability. The results demonstrate that the Hybrid Gradient Boosting approach achieves optimal and stable predictive performance with good generalization ability. Variables such as age, blood pressure, cholesterol levels, and body mass index (BMI) are shown to have a significant influence on heart disease risk. This study contributes to the development of accurate and interpretable predictive models and can be utilized by healthcare professionals, medical institutions, and researchers as a decision-support tool for heart disease risk assessment.

Keyword: heart disease, hybrid gradient boosting, XGBoost, LightGBM, risk prediction.