

**PENGEMBANGAN MODEL TUNED LOGISTIC
REGRESSION UNTUK DETEKSI DINI SERANGAN
JANTUNG DENGAN PENDEKATAN EXPLAINABLE AI
MENGUNAKAN SHAP**

SKRIPSI

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



disusun oleh

ADITYA DWI SURYO HARDIYANTO

22.11.5086

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PENGEMBANGAN MODEL TUNED LOGISTIC
REGRESSION UNTUK DETEKSI DINI SERANGAN
JANTUNG DENGAN PENDEKATAN EXPLAINABLE AI
MENGUNAKAN SHAP**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ADITYA DWI SURYO HARDIYANTO

22.11.5086

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PENGEMBANGAN MODEL TUNED LOGISTIC REGRESSION UNTUK
DETEKSI DINI SERANGAN JANTUNG DENGAN PENDEKATAN
EXPLAINABLE AI MENGGUNAKAN SHAP**

yang disusun dan diajukan oleh

Aditya Dwi Suryo HardiYanto

22.11.5086

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Januari 2026

Dosen Pembimbing,



Anna Baita, S.Kom., M.Kom.
NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

**PENGEMBANGAN MODEL TUNED LOGISTIC REGRESSION UNTUK
DETEKSI DINI SERANGAN JANTUNG DENGAN PENDEKATAN
EXPLAINABLE AI MENGGUNAKAN SHAP**

yang disusun dan diajukan oleh

Aditya Dwi Surya HardiYanto

22.11.5086

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dr. Hartatik, S.T., M.Cs.
NIK. 190302232



Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192



Anna Baita, S.Kom., M.Kom.
NIK. 190302290

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aditya Dwi Suryo HardiYanto
NIM : 22.11.5086

Menyatakan bahwa Skripsi dengan judul berikut:

Pengembangan Model Tuned Logistic Regression untuk Deteksi Dini Serangan Jantung dengan Pendekatan Explainable AI Menggunakan SHAP Pengembangan Model

Dosen Pembimbing : Anna Baita, 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, 26 Januari 2026

Yang Menyatakan



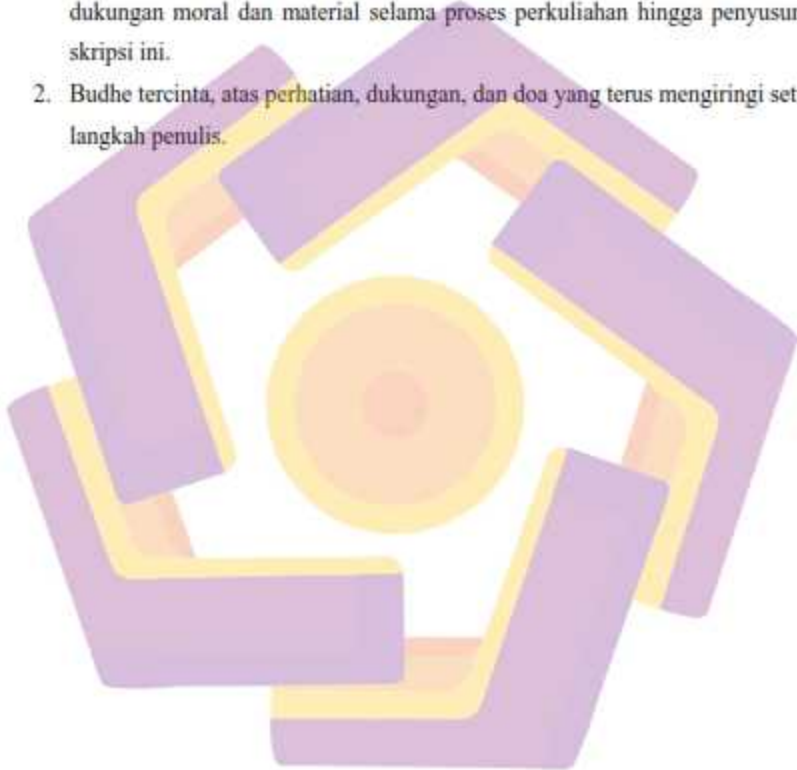
EFAX84757203

Aditya Dwi Suryo HardiYanto

HALAMAN PERSEMBAHAN

Dengan memanjatkan puji dan syukur ke hadirat Allah SWT, penulis mempersembahkan skripsi ini kepada:

1. Ibu dan Ayah tercinta, yang senantiasa memberikan doa, kasih sayang, serta dukungan moral dan material selama proses perkuliahan hingga penyusunan skripsi ini.
2. Budhe tercinta, atas perhatian, dukungan, dan doa yang terus mengiringi setiap langkah penulis.



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga skripsi ini dapat diselesaikan dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada program studi yang penulis tempuh.

Penulis menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Ibu dan Ayah penulis, atas doa, kasih sayang, perhatian, serta dukungan moral yang senantiasa diberikan selama masa perkuliahan hingga penyusunan skripsi ini.
2. Ibu Anna Baita, S.Kom., M.Kom., selaku Dosen Pembimbing, yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta masukan yang sangat berharga sejak tahap awal hingga penyelesaian skripsi ini.
3. Budhe(*Simbok*) Rubinem, atas perhatian, doa, dan dukungan yang senantiasa menyertai penulis selama proses perkuliahan dan penyusunan skripsi ini.
4. Sigit Surya Bagus Setiawan, selaku kakak laki-laki penulis, yang telah memberikan dukungan material dan motivasi sehingga proses penyusunan skripsi ini dapat berjalan dengan lancar.

Yogyakarta, 24 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.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6

2.2	Dasar Teori.....	15
2.2.1	Serangan Jantung	15
2.2.2	<i>Machine Learning</i>	15
2.2.3	<i>Logistic Regression</i>	19
2.2.4	<i>Hyperparameter Tuning</i>	22
2.2.5	<i>Grid Search</i>	23
2.2.6	<i>Randomsearch</i>	24
2.2.7	<i>Bayesian Optimization</i>	26
2.2.8	<i>XAI</i>	27
2.2.9	<i>SHAP</i>	28
2.2.1	<i>Metric Evaluation</i>	30
BAB III METODE PENELITIAN		35
3.1	Objek Penelitian	35
3.2	Alur Penelitian.....	35
3.3	Alat dan Bahan	39
BAB IV HASIL DAN PEMBAHASAN		40
4.1	Pengumpulan data	40
4.1.1	Informasi Umum Dataset	40
4.2	Pemahaman dan Eksplorasi Data (<i>EDA</i>)	42
4.2.1	Inspeksi Awal dan Profiling Dataset.....	42
4.2.2	Identifikasi Anomali	43
4.2.3	Analisis Distribusi Fitur dan Variabel Target	45
4.3	<i>Data Preprocessing</i>	63
4.3.1	Penanganan Anomali	63
4.3.2	Transformasi Fitur (<i>Encoding</i>).....	64

4.3.3	Analisis Korelasi Fitur	66
4.3.4	Splitting Data dan Penskalaan (<i>Scaling</i>).....	68
4.4	Eksperimen dan Evaluasi Model <i>Baseline</i>	69
4.4.1	Konfigurasi Model <i>Baseline</i>	70
4.4.2	Analisis Metrik Evaluasi <i>Baseline</i>	70
4.4.3	Analisis Confusion Matrix.....	72
4.5	Optimasi <i>Hyperparameter</i> dan Komparasi <i>Search Space</i>	74
4.5.1	Konfigurasi Ruang Pencarian (<i>Search Space</i>).....	75
4.5.2	Hasil Optimasi <i>Hyperparameter</i>	75
4.5.3	Perbandingan Model <i>Baseline</i> dan Model <i>Tuning</i>	77
4.6	Interpretasi Model dengan <i>SHAP (XAI)</i>	81
4.6.1	Interpretasi Global (<i>Feature Importance</i>).....	82
4.6.2	Interpretasi Lokal (<i>Waterfall Plot</i> pada Kasus Spesifik)	84
4.7	Pembahasan Hasil dan Temuan Klinis.....	88
4.8	<i>Deployment</i> Model	90
4.8.1	<i>Deployment</i> Model Berbasis <i>Web</i>	90
4.8.2	<i>Repository Notebook</i>	92
4.8.3	<i>Repository Model</i>	92
BAB V	PENUTUP	93
5.1	Kesimpulan.....	93
5.2	Saran.....	94
REFERENSI		95
LAMPIRAN SHAP LOCAL		101

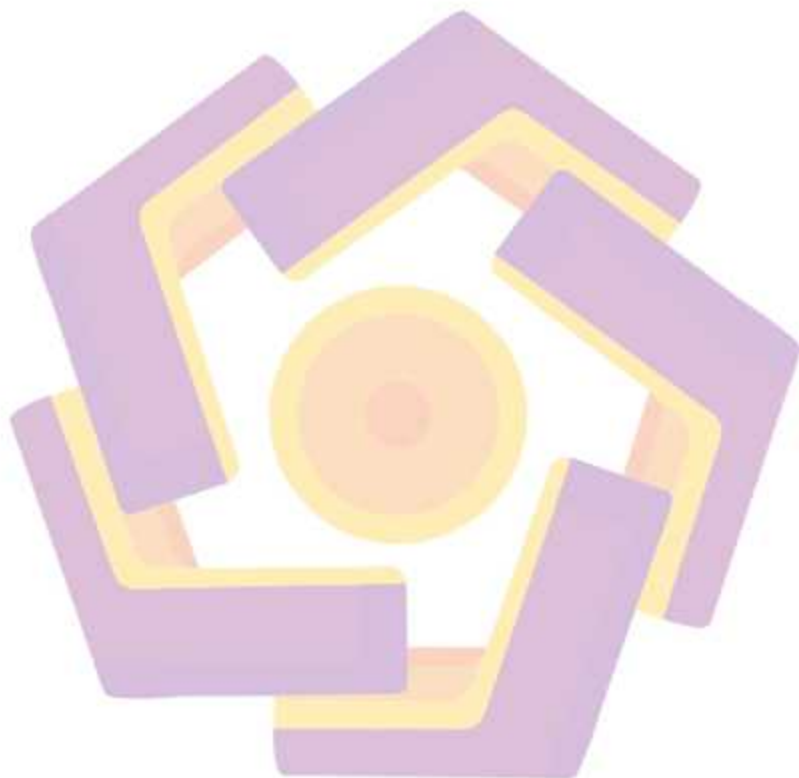
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 4.1 Informasi Umum	40
Tabel 4.2 Deskriptif Fitur Numerik	42
Tabel 4.3 Jumlah dan Persentase Nilai 0	44
Tabel 4.4 <i>Encoding</i>	65
Tabel 4.5 Konfigurasi Model	70
Tabel 4.6 Evaluasi Metrik	71
Tabel 4.7 Pasien <i>FN Model Baseline</i>	73
Tabel 4.8 Konfigurasi <i>Search Space</i>	75
Tabel 4.9 Hasil Optimasi <i>Hyperparameter</i>	76
Tabel 4.10 Peningkatan Kinerja	78
Tabel 4.11 Ringkasan Representatif SHAP <i>Local</i>	86

DAFTAR GAMBAR

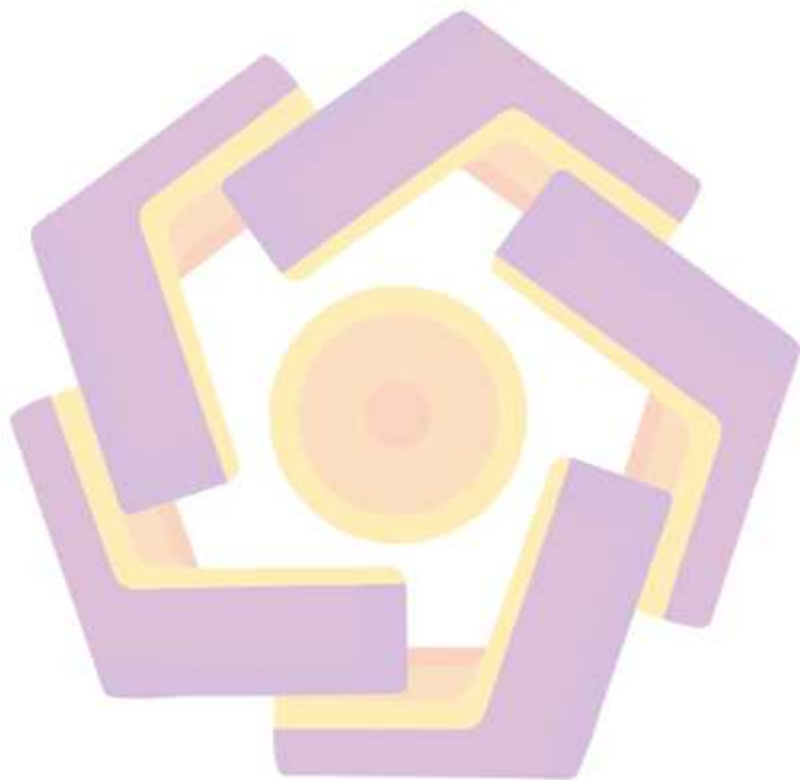
Gambar 2.1 Alur Proses Explainable Artificial Intelligence	28
Gambar 2.2 <i>Confusion Matrix</i>	31
Gambar 2.3 Grafik <i>ROC</i> menunjukkan lima pengklasifikasi	33
Gambar 2.4 Grafik <i>AUC</i> menunjukkan pengklasifikasi diskrit.....	34
Gambar 3.1 Alur Penelitian	36
Gambar 4.1 Distribusi Jenis Kelamin	46
Gambar 4.2 Distribusi Pasien Sakit dan Sehat.....	46
Gambar 4.3 Distribusi Kategori Chest Pain Type	47
Gambar 4.4 Distribusi <i>ChestPainType</i> Pada Kategori Pasien	48
Gambar 4.5 Distribusi <i>RestingECG</i>	49
Gambar 4.6 Distribusi <i>RestingECG</i> Pada Kategori Pasien.....	50
Gambar 4.7 Distribusi Usia.....	51
Gambar 4.8 Distribusi Usia Pada Kategori Pasien	52
Gambar 4.9 Distribusi Heart Disease.....	53
Gambar 4.10 Distribusi Kolesterol	54
Gambar 4.11 Histogram MaxHR.....	54
Gambar 4.12 Distribusi <i>MaxHR</i> Pada Kategori Pasien.....	55
Gambar 4.13 Distribusi <i>FastingBS</i>	57
Gambar 4.14 Distribusi <i>FastingBS</i> Pada Kategori Pasien.....	57
Gambar 4.15 Distribusi <i>Exercise Angina</i>	58
Gambar 4.16 Distribusi <i>ExerciseAngina</i> Pada Kategori Pasien	59
Gambar 4.17 Distribusi <i>Oldpeak</i>	60
Gambar 4.18 Distribusi <i>Oldpeak</i> Pada Kategori Pasien	61
Gambar 4.19 Distribusi <i>ST Slope</i>	62
Gambar 4.20 Distribusi <i>ST Slope</i> Pada Kategori Pasien.....	62
Gambar 4.21 Correlation Matrix Antar Fitur.....	67
Gambar 4.22 Kurva <i>ROC</i>	71
Gambar 4.23 Hasil <i>Confusion Matrix</i>	73
Gambar 4.24 <i>ROC</i> Gabungan	80

Gambar 4.25 <i>SHAP</i> Global.....	82
Gambar 4.26 <i>Interface</i> Input Data Klinis Pasien.....	91
Gambar 4.27 Interpretasi <i>SHAP</i>	91

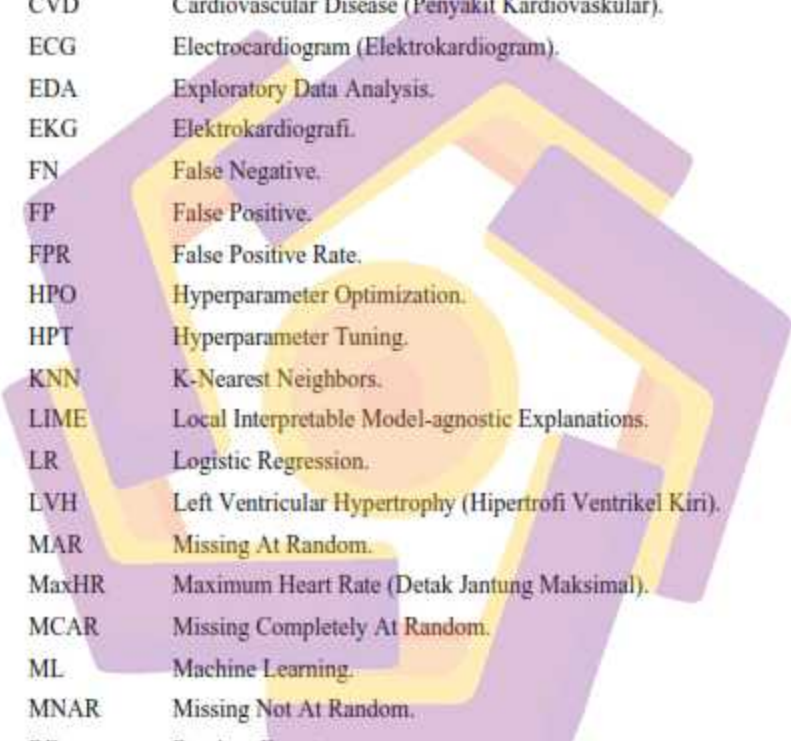


DAFTAR LAMPIRAN

LAMPIRAN SHAP LOCAL	101
---------------------------	-----

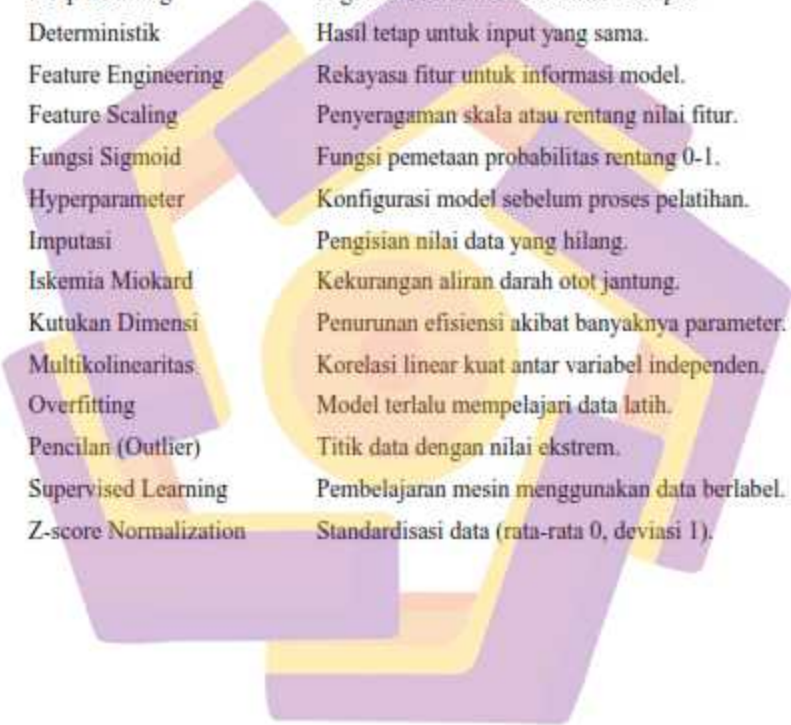


DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence (Kecerdasan Buatan).
ANN	Artificial Neural Network (Jaringan Saraf Tiruan).
AUC	Area Under the Curve.
CAD	Coronary Artery Disease (Penyakit Jantung Koroner).
CVD	Cardiovascular Disease (Penyakit Kardiovaskular).
ECG	Electrocardiogram (Elektrokardiogram).
EDA	Exploratory Data Analysis.
EKG	Elektrokardiografi.
FN	False Negative.
FP	False Positive.
FPR	False Positive Rate.
HPO	Hyperparameter Optimization.
HPT	Hyperparameter Tuning.
KNN	K-Nearest Neighbors.
LIME	Local Interpretable Model-agnostic Explanations.
LR	Logistic Regression.
LVH	Left Ventricular Hypertrophy (Hipertrofi Ventrikel Kiri).
MAR	Missing At Random.
MaxHR	Maximum Heart Rate (Detak Jantung Maksimal).
MCAR	Missing Completely At Random.
ML	Machine Learning.
MNAR	Missing Not At Random.
RF	Random Forest.
ROC	Receiver Operating Characteristic.
SHAP	SHapley Additive exPlanations.
TN	True Negative.
TP	True Positive.
TPR	True Positive Rate.
XAI	Explainable Artificial Intelligence.

DAFTAR ISTILAH



Threshold	Batas nilai penentu klasifikasi prediksi.
Biner	Data dengan dua kemungkinan nilai (0 atau 1).
Brute-force	Pengujian seluruh kombinasi secara menyeluruh.
Data Leakage	Kebocoran informasi data uji ke data latih.
Deep Learning	Algoritma saraf tiruan struktur berlapis.
Deterministik	Hasil tetap untuk input yang sama.
Feature Engineering	Rekayasa fitur untuk informasi model.
Feature Scaling	Penyeragaman skala atau rentang nilai fitur.
Fungsi Sigmoid	Fungsi pemetaan probabilitas rentang 0-1.
Hyperparameter	Konfigurasi model sebelum proses pelatihan.
Imputasi	Pengisian nilai data yang hilang.
Iskemia Miokard	Kekurangan aliran darah otot jantung.
Kutukan Dimensi	Penurunan efisiensi akibat banyaknya parameter.
Multikolinearitas	Korelasi linear kuat antar variabel independen.
Overfitting	Model terlalu mempelajari data latih.
Pencilan (Outlier)	Titik data dengan nilai ekstrem.
Supervised Learning	Pembelajaran mesin menggunakan data berlabel.
Z-score Normalization	Standardisasi data (rata-rata 0, deviasi 1).

INTISARI

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, termasuk di Indonesia. Deteksi dini yang akurat menjadi tantangan penting karena kesalahan klasifikasi, khususnya *false negative*, dapat berdampak serius terhadap keselamatan pasien. Model klasifikasi berbasis *machine learning* banyak digunakan untuk membantu proses diagnosis, namun performanya sangat bergantung pada konfigurasi *hyperparameter* yang digunakan. Model *baseline* yang tidak dioptimasi berpotensi menghasilkan kinerja yang kurang optimal, terutama dalam mendeteksi pasien yang benar-benar menderita penyakit jantung.

Penelitian ini bertujuan untuk meningkatkan kinerja model klasifikasi penyakit jantung menggunakan algoritma *Logistic Regression* melalui optimasi *hyperparameter*. Proses optimasi dilakukan dengan membandingkan tiga metode, yaitu *GridSearchCV*, *RandomizedSearchCV*, dan *Bayesian Optimization*, menggunakan skema *cross-validation*. Evaluasi model dilakukan berdasarkan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Selain itu, pendekatan *Explainable Artificial Intelligence (XAI)* menggunakan *SHAP (SHapley Additive Explanations)* diterapkan untuk menginterpretasikan keputusan model secara global dan lokal.

Hasil penelitian menunjukkan bahwa optimasi *hyperparameter* mampu meningkatkan kinerja model, khususnya pada metrik *Recall* sebesar 5,43% dan *F1-Score* sebesar 0,91% dibandingkan model *baseline*. Namun demikian, terjadi penurunan kinerja pada metrik *Precision* sebesar 3,09% dan *ROC-AUC* sebesar 0,62%. Interpretasi *SHAP* mengungkap bahwa fitur klinis seperti *ST Slope* pada *EKG* dan *Chest Pain Type* memiliki pengaruh dominan terhadap prediksi model, serta menunjukkan konsistensi dengan pengetahuan medis.

Kata kunci: Penyakit Jantung, *Logistic Regression*, *Hyperparameter Tuning*, *SHAP*, *Machine Learning*

ABSTRACT

Heart disease is one of the leading causes of death worldwide, including in Indonesia. Early and accurate detection is a critical challenge, as misclassification particularly false negatives can have serious consequences for patient safety. Machine learning-based classification models are widely used to support medical diagnosis; however, their performance highly depends on the selected hyperparameter configuration. Baseline models without optimization may produce suboptimal results, especially in identifying patients with heart disease.

This study aims to improve the performance of a heart disease classification model using Logistic Regression through hyperparameter optimization. Three optimization methods were evaluated, namely GridSearchCV, RandomizedSearchCV, and Bayesian Optimization, employing a cross-validation scheme. Model performance was assessed using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics. In addition, Explainable Artificial Intelligence (XAI) was applied using SHAP (SHapley Additive Explanations) to provide global and local interpretability of the model's predictions.

The results indicate that hyperparameter optimization improves model performance, particularly in Recall, which increases by 5.43%, and F1-Score, which increases by 0.91% compared to the baseline model. However, a decrease is observed in Precision by 3.09% and ROC-AUC by 0.62%. SHAP-based interpretation reveals that clinical features such as ST Slope from electrocardiogram (ECG) results and Chest Pain Type have a dominant influence on the model's predictions, demonstrating consistency with established medical knowledge.

Keywords: Heart Disease, Logistic Regression, Hyperparameter Tuning, SHAP, Machine Learning