

**IMPLEMENTASI LOGISTIC REGRESSION UNTUK
PREDIKSI PENYAKIT JANTUNG PADA
DATASET CLEVELAND UCI**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Sistem Informasi



disusun oleh

SAYIDAH AISAH AL-QODIRY

21.12.1919

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**IMPLEMENTASI LOGISTIC REGRESSION UNTUK
PREDIKSI PENYAKIT JANTUNG PADA
DATASET CLEVELAND UCI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Sistem Informasi



disusun oleh

SAYIDAH AISAH AL-QODIRY

21.12.1919

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI LOGISTIC REGRESSION UNTUK
PREDIKSI PENYAKIT JANTUNG PADA
DATASET CLEVELAND UCI**

yang disusun dan diajukan oleh

Sayidah Aisah Al-Qodiry

21.12.1919

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Januari 2025

Dosen Pembimbing,



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PENGESAHAN
SKRIPSI
IMPLEMENTASI LOGISTIC REGRESSION UNTUK
PREDIKSI PENYAKIT JANTUNG PADA
DATASET CLEVELAND UCI

yang disusun dan diajukan oleh

Sayidah Aisah Al-Qodiry

21.12.1919

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Januari 2025

Susunan Dewan Penguji

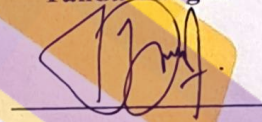
Nama Penguji

Donni Prabowo, S.Kom., M.Kom.
NIK. 190302253

Ninik Tri Hartanti, S.Kom., M.Kom.
NIK. 190302330

Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Sayidah Aisah Al-Qodiry
NIM : 21.12.1919

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Logistic Regression untuk Prediksi Penyakit Jantung pada Dataset Cleveland UCI

Dosen Pembimbing : Hanif Al Fatta, S.Kom., M.Kom., Ph.D.

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, 21 Januari 2025

Yang Menyatakan,



Sayidah Aisah Al-Qodiry

HALAMAN PERSEMBAHAN

Pertama-tama, penulis mengucapkan puji syukur kehadiran Allah SWT atas segala nikmat-Nya berupa kesehatan, kekuatan, dan inspirasi yang sangat besar dalam proses penyelesaian laporan tugas akhir ini. Shalawat dan salam senantiasa tercurah kepada Nabi Muhammad SAW. Laporan tugas akhir ini penulis persembahkan sebagai wujud semangat dan usaha yang tak henti-hentinya, serta sebagai ungkapan cinta dan kasih sayang penulis kepada orang-orang yang sangat berarti dalam hidup penulis.

Untuk karya sederhana ini, maka penulis persembahkan untuk:

1. Pintu surgaku, Ibu Sabini. Beliau sangat berperan penting dalam proses menyelesaikan program studi penulis, beliau juga tidak sempat merasakan Pendidikan di bangku perkuliahan, namun beliau tidak henti memberi semangat, serta do'a yang selalu mengiringi langkah penulis sehingga bisa menyelesaikan program studi ini sampai selesai.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dosen Pembimbing, terima kasih telah memberi bimbingan, nasehat, dan dukungan sehingga bisa menyelesaikan laporan tugas akhir ini.
3. Bapak Andriyan Dwi Putra, M.Kom., selaku Dosen Wali penulis yang telah memberikan bantuan, nasehat, dan dukungan yang sangat berarti selama perkuliahan ini, serta dalam proses penyelesaian laporan tugas akhir ini.
4. Kakak-kakak tercinta Muhammad Mikail Al-Qodiry, Muhammad Jibril Babasamasi, Rini Wijaya, Azizul Khakim dan keponakan saya Casilda Zarifa Khakim. Terima kasih atas dukungan moral maupun materiil yang telah diberikan, serta doa dan asih sayang yang luar biasa.
5. Teman-teman Sistem Informasi Angkatan 2021, terutama Kelas 02, terima kasih atas motivasi, semangat, dan dukungan tanpa henti yang telah diberikan, sehingga secara tidak langsung membantu penulis dalam menyelesaikan laporan tugas akhir ini. Teruntuk Dhea, Iqoh, Riya, Sindi, Yane, Lovina, Diez, Citta, Fanny, Kiara, Dina, Kiki, Tania, dan Siska, kalian adalah orang-orang pilihan yang selalu berada di balik layar, membersamai perjuangan ini, dan selalu siap membantu penulis. Terima kasih, semoga perjuangan kita dilancarkan hingga akhir dan kita semua dapat meraih kesuksesan.

6. Untuk seseorang yang belum dapat ditulis dengan jelas namanya disini, namun sudah tertulis jelas di *Lauhul Mahfudz* untukku. Terimakasih sudah menjadi salah satu sumber motivasi penulis dalam menyelesaikan laporan tugas akhir ini. Meskipun saat ini penulis tidak tahu keberadaanmu entah di bumi bagian mana dan menggenggam tangan siapa. Seperti kata Bj Habibie “Kalau memang dia dilahirkan untuk saya, kamu jungkir balik pun saya yang dapat”.
7. Terakhir, terima kasih kepada perempuan sederhana yang memiliki keinginan tinggi namun terkadang sulit dimengerti isi kepalanya, sang penulis yaitu diri saya sendiri, Sayidah Aisah Al-Qodiry, yang telah bekerja keras dan berjuang sejauh ini. Mampu mengendalikan diri dari berbagai tekanan diluar keadaan dan tak pernah memutuskan menyerah sesulit apapun proses penyusunan laporan tugas akhir ini dengan menyelesaikan sebaik dan semaksimal mungkin, ini merupakan pencapaian yang patut dibanggakan untuk diri sendiri.

Dengan segala kerendahan hati, penulis berharap semoga laporan tugas akhir ini dapat bermanfaat bagi pihak yang membutuhkan dan dapat dijadikan referensi yang berguna untuk penelitian-penelitian selanjutnya. Penulis juga mengharapkan kritik dan saran yang membangun untuk perbaikan di masa yang akan datang.

Yogyakarta, 21 Januari 2025

Penulis

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir ini dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

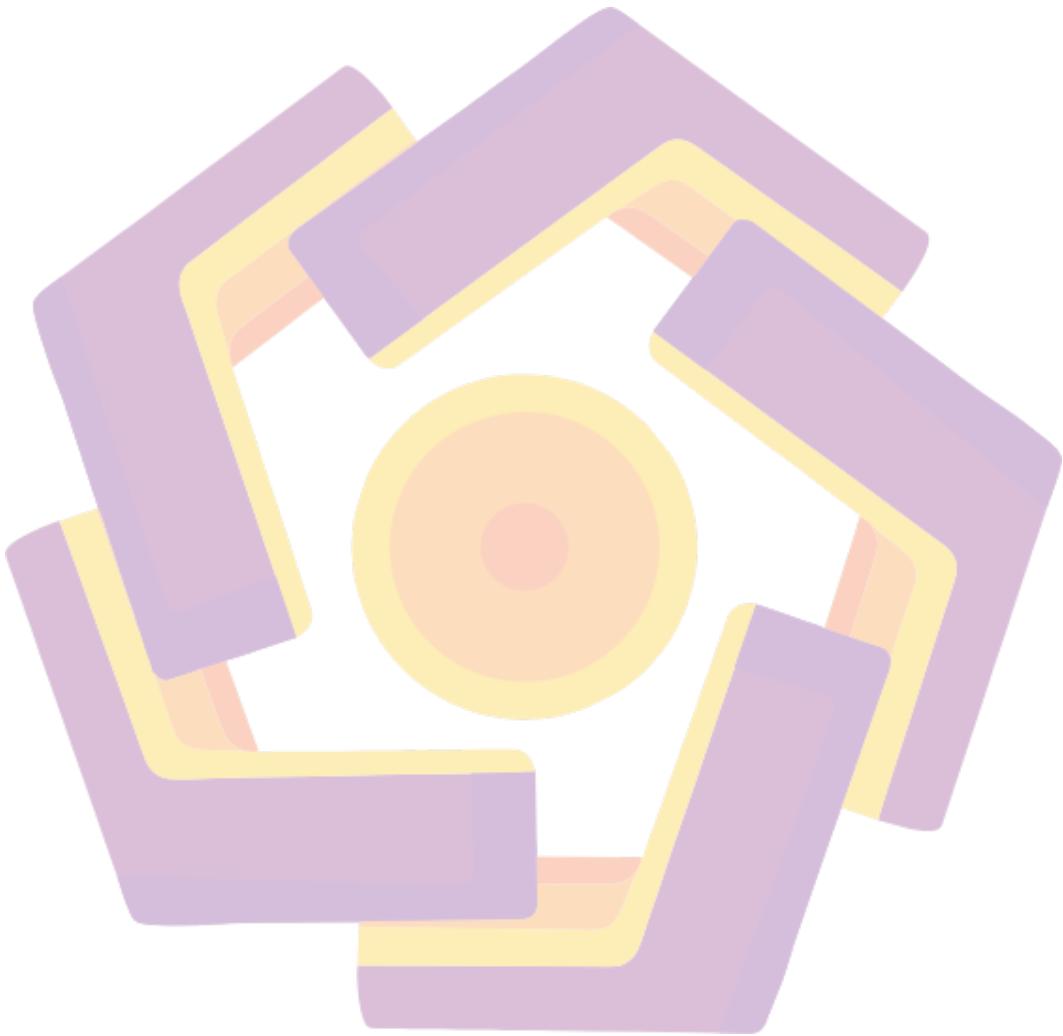
1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer sekaligus Dosen Pembimbing, yang telah memberikan arahan, saran, dan motivasi kepada penulis
3. Bapak Anggit Dwi Hartanto, M.Kom., selaku Ketua Program Studi Sistem Informasi
4. Bapak Andriyan Dwi Putra, M.Kom., selaku Dosen Wali penulis.
5. Kedua orang tua, keluarga besar, dan teman-teman tercinta yang memberikan semangat dan doa kepada penulis.
6. Semua pihak yang telah membantu terselesaikannya penulisan laporan tugas akhir ini.

Yogyakarta, 21 Januari 2025

Penulis

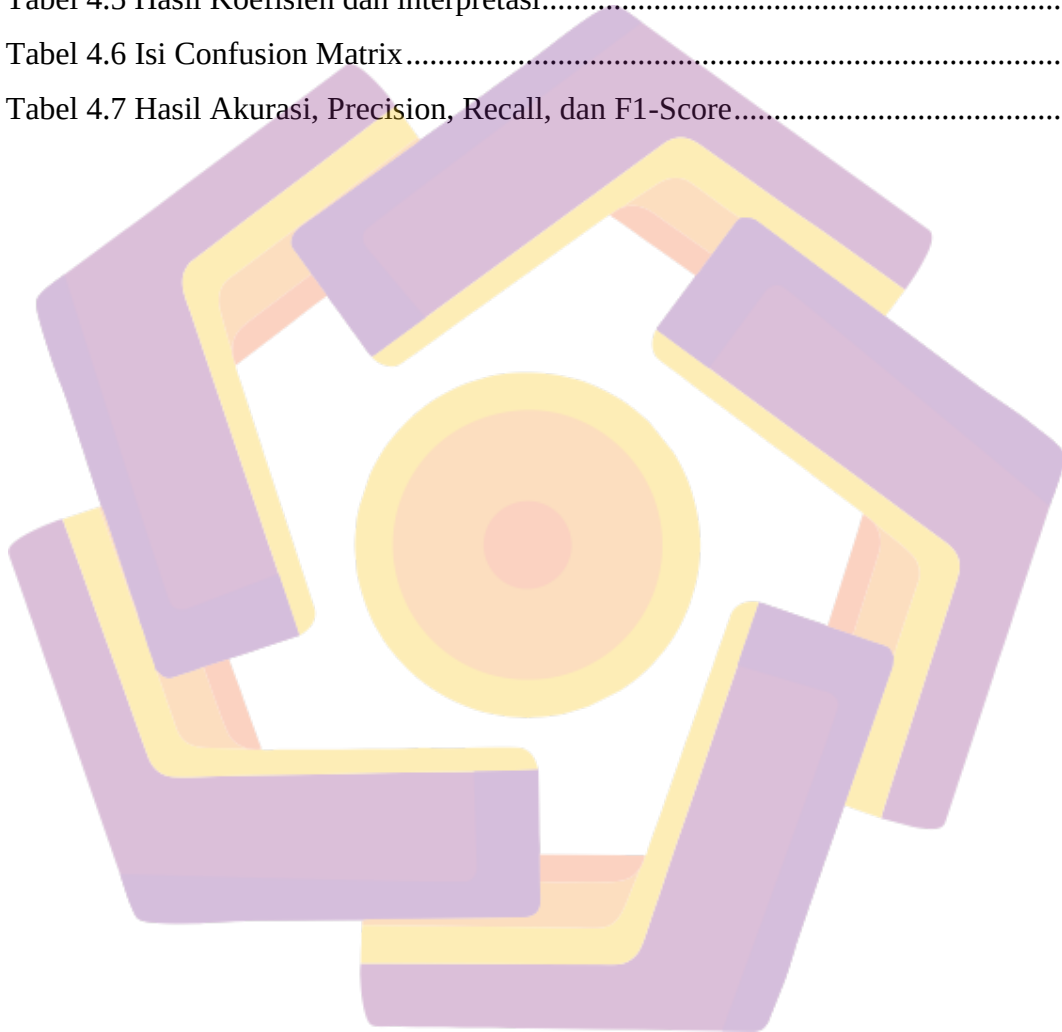
DAFTAR ISI	
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN.....	xiv
DAFTAR ISTILAH	xv
INTISARI.....	xvii
ABSTRACT.....	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	1
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penelitian.....	2
BAB II TINJAUAN PUSTAKA.....	4
2.1 Studi Literatur	4
2.2 Dasar Teori	11
2.2.1 Penyakit jantung.....	11
2.2.2 Data mining.....	11
2.2.3 Machine learning.....	12
2.2.4 Logistic regression.....	12
2.2.5 Confusion Matrix	13
2.2.6 Akurasi (Accuracy).....	14
2.2.7 Presisi (Precision)	14
2.2.8 Recall	15

2.2.9	F1-Score.....	15
BAB III METODE PENELITIAN		16
3.1	Objek Penelitian	16
3.2	Alur Penelitian	16
3.2.1	Data Selection.....	17
3.2.2	Pre-Processing dan Cleaning Data	17
3.2.3	Transformasi Data	17
3.2.4	Classification	17
3.2.5	Interpretation	18
3.2.6	Knowledge Presentation.....	18
3.3	Alat dan Bahan.....	18
3.3.1	Dataset Penelitian.....	18
3.3.2	Alat	18
BAB IV HASIL DAN PEMBAHASAN		20
4.1	Deskripsi Dataset.....	20
4.2	Pre-processing dan Cleaning Data	21
4.3	Transformasi Data.....	22
4.4	Classification.....	27
4.4.1	Confussion matrix	27
4.4.2	Akurasi, precision, recall, dan f1-score.....	30
4.4.3	Visualisasi kurva ROC dan perhitungan nilai AUC.....	31
4.5	Interpretation.....	35
4.6	Knowledge Presentation.....	37
4.6.1	Confussion matrix	37
4.6.2	Akurasi, precision, recall, dan f1-score.....	38
4.6.3	ROC Curve dan Nilai AUC.....	39
4.6.4	Koefisien Model.....	40
BAB V PENUTUP.....		41
5.1	Kesimpulan	41
5.2	Saran	41
REFERENSI.....		43



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	6
Tabel 4.1 Contoh Dataset Cleveland UCI	21
Tabel 4.2 Contoh Output y_{test} dan y_{pred}	28
Tabel 4.3 Hasil Confusion Matrix dalam Heatmap	29
Tabel 4.4 FPR dan TPR	32
Tabel 4.5 Hasil Koefisien dan interpretasi.....	35
Tabel 4.6 Isi Confusion Matrix.....	37
Tabel 4.7 Hasil Akurasi, Precision, Recall, dan F1-Score.....	38

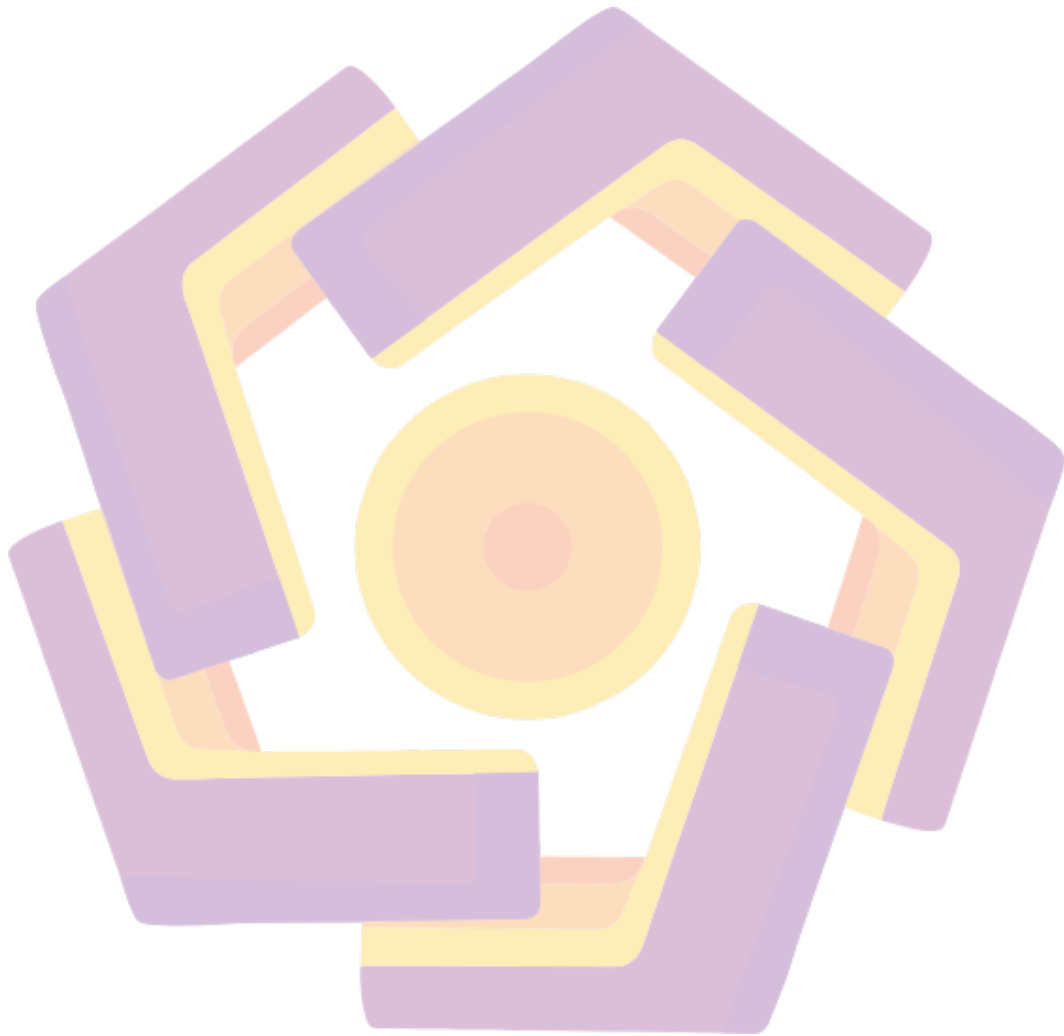


DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian KDD dalam Data Mining ..	18
Gambar 4.1 Script Pre-processing dan Cleaning Data	21
Gambar 4.2 Script Pembagian Data.....	23
Gambar 4.3 Script Balancing Data menggunakan SMOTE	24
Gambar 4.4 Contoh Dataset Balancing.....	26
Gambar 4.5 Script Implementasi Logistic Regression	26
Gambar 4.6 Script Confusion Matrix	27
Gambar 4.7 Hasil Output Confusion Matrix	28
Gambar 4.8 Script Akurasi, Precision, Recall, dan F1-Score.....	30
Gambar 4.9 Hasil Output Akurasi, Precision, Recall, dan F1-Score.....	30
Gambar 4.10 Script Kurva ROC dan Nilai AUC.....	31
Gambar 4.11 Hasil Output Kurva ROC dan Nilai AUC.....	34
Gambar 4.12 Script Koefisien Model	35
Gambar 4.13 Script Koefisien Model Fitur yang Paling Berpengaruh.....	36
Gambar 4.14 Visualisasi Heatmap Confusion Matrix	38
Gambar 4.15 ROC Curve dan AUC	40

DAFTAR LAMPIRAN

Source Code Program	46
---------------------------	----



DAFTAR LAMBANG DAN SINGKATAN



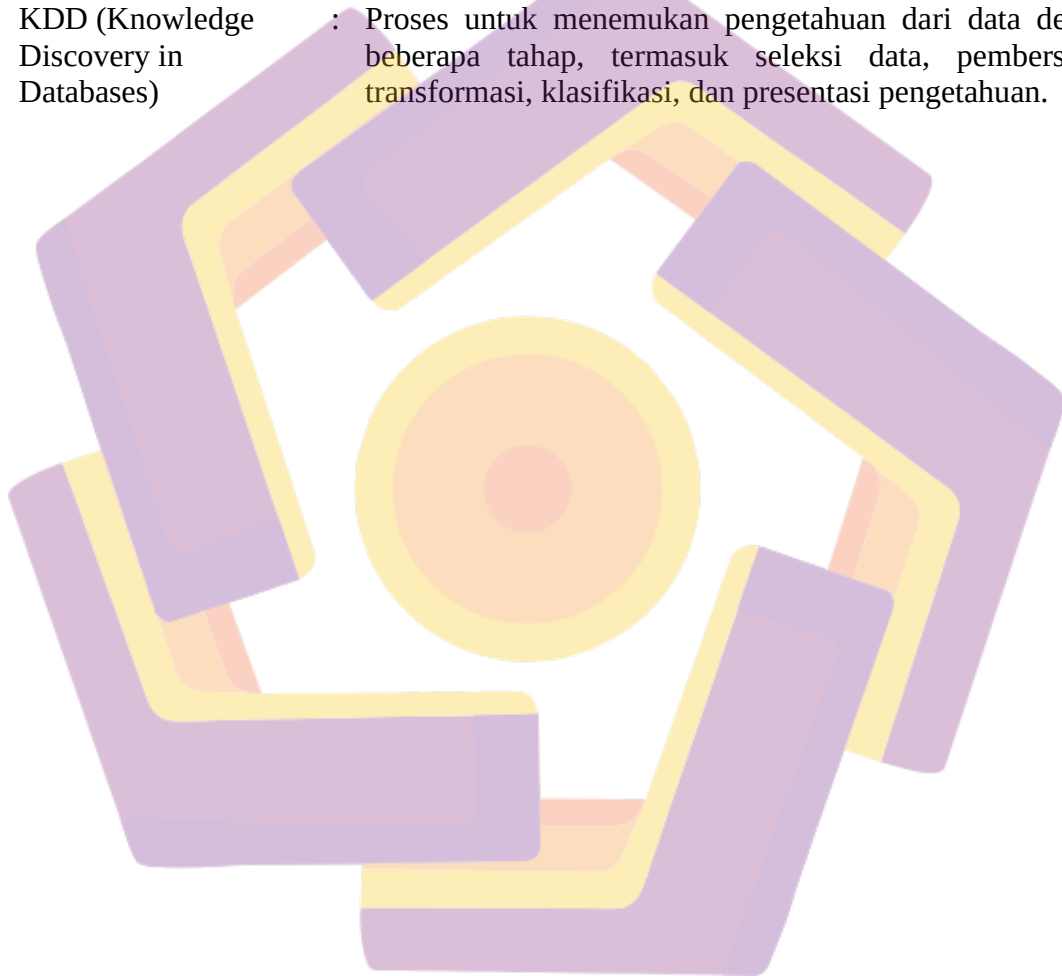
TP	: True Positive
FP	: False Positive
TN	: True Negative
FN	: False Negative
KDD	: Knowledge Discovery in Databases
SMOTE	: Synthetic Minority Over-sampling Technique
ROC	: Receiver Operating Characteristic
AUC	: Area Under the Curve
FBS	: Fasting Blood Sugar
ECG	: Electrocardiogram
CP	: Chest Pain
Trestbps	: Resting Blood Pressure
Chol	: Cholesterol
Thalach	: Maximum Heart Rate Achieved
Exang	: Exercise Induced Angina
Oldpeak	: Pengurangan segmen ST saat olahraga
CA	: Number of Major Vessels Colored by Fluoroscopy
Thal	: Thalassemia Status
FPR	: False Positive Rate
TPR	: True Positive Rate
Σ	: Sigma

DAFTAR ISTILAH

Dataset Cleveland UCI	: Dataset yang berisi data pasien dan fitur medis untuk memprediksi risiko penyakit jantung.
Pre-processing:	: Proses mempersiapkan data untuk analisis lebih lanjut, termasuk pengisian nilai kosong dan normalisasi data.
Missing values	: Nilai yang hilang dalam dataset yang perlu ditangani.
Mean Imputation	: Teknik untuk mengganti nilai kosong dengan rata-rata nilai dalam kolom tersebut.
Normalisasi	: Proses transformasi data untuk memastikan nilai fitur berada dalam rentang yang sama, misalnya [0, 1] dengan teknik seperti Min-Max Scaler.
Min-Max Scaler	: Teknik normalisasi untuk mengubah nilai fitur ke dalam rentang [0, 1].
Data Latih	: Bagian dari dataset yang digunakan untuk melatih model.
Data Uji	: Bagian dari dataset yang digunakan untuk menguji model setelah dilatih.
Stratified Sampling	: Teknik pembagian data yang memastikan distribusi label target tetap seimbang antara data latih dan data uji.
SMOTE (Synthetic Minority Over-sampling Technique)	: Teknik untuk menangani ketidakseimbangan data dengan menambah data sintetis pada kelas minoritas.
Logistic Regression	: Algoritma statistik yang digunakan untuk memprediksi probabilitas suatu kejadian berdasarkan variabel independen.
Confusion Matrix	: Matriks yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan prediksi dan nilai sebenarnya.
Akurasi	: Metrik untuk mengukur seberapa baik model dalam melakukan prediksi secara keseluruhan.
Precision	: Metrik yang mengukur akurasi prediksi positif.
Recall	: Metrik yang mengukur kemampuan model untuk mendeteksi kasus positif yang sebenarnya.
F1-Score	: Kombinasi dari precision dan recall untuk memberikan gambaran lebih lengkap mengenai kinerja model.
ROC Curve	: Kurva yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan True Positive Rate dan

False Positive Rate.

- AUC (Area Under the Curve) : Metrik yang mengukur seberapa baik model dapat membedakan antara kelas positif dan negatif.
- True Positive (TP) : Jumlah prediksi benar untuk kelas positif.
- False Positive (FP) : Jumlah prediksi yang salah untuk kelas positif.
- True Negative (TN) : Jumlah prediksi benar untuk kelas negatif.
- False Negative (FN) : Jumlah prediksi yang salah untuk kelas negatif.
- KDD (Knowledge Discovery in Databases) : Proses untuk menemukan pengetahuan dari data dengan beberapa tahap, termasuk seleksi data, pembersihan, transformasi, klasifikasi, dan presentasi pengetahuan.



INTISARI

Penyakit jantung merupakan salah satu penyebab utama kematian di Indonesia, dengan tingkat kematian yang terus meningkat setiap tahun. Metode tradisional dalam diagnosis penyakit jantung memerlukan waktu dan tidak selalu efisien, sehingga diperlukan pendekatan berbasis teknologi yang lebih akurat dan cepat. Penelitian ini menggunakan algoritma Logistic Regression untuk membangun model prediksi risiko penyakit jantung. Dataset Cleveland UCI, yang terdiri dari 297 data pasien dengan 13 atribut medis, digunakan sebagai basis penelitian ini. Dataset ini diolah melalui proses Knowledge Discovery in Database (KDD), yang meliputi data selection, pre-processing dan cleaning data, transformasi data, classification, interpretasi, dan knowledge presentation.

Metode yang digunakan mencakup normalisasi data, penyeimbangan kelas menggunakan SMOTE, serta evaluasi model melalui metrik seperti akurasi, precision, recall, F1-score, dan ROC-AUC. Hasil penelitian menunjukkan bahwa model Logistic Regression mampu memberikan prediksi dengan akurasi sebesar 91,67%, precision 100% untuk kelas pasien dengan penyakit jantung, recall 82,14%, F1-score 90,23%, dan nilai AUC sebesar 0,95. Hasil ini mengindikasikan kemampuan model dalam membedakan antara pasien yang terkena dan tidak terkena penyakit jantung secara efektif.

Penelitian ini memberikan kontribusi dalam pengembangan metode prediksi berbasis data untuk membantu diagnosis dini penyakit jantung. Model ini dapat dimanfaatkan oleh tenaga medis dan institusi kesehatan untuk meningkatkan efisiensi dan akurasi dalam pengambilan keputusan. Peneliti selanjutnya disarankan untuk mengeksplorasi algoritma lain atau meningkatkan dataset untuk mengoptimalkan performa model.

Kata Kunci: Penyakit Jantung, Logistic Regression, Prediksi, SMOTE, KDD

ABSTRACT

Heart disease is one of the leading causes of death in Indonesia, with the death rate continuing to increase every year. Traditional methods in diagnosing heart disease are time-consuming and not always efficient, so a more accurate and faster technology-based approach is needed. This study uses the Logistic Regression algorithm to build a heart disease risk prediction model. The Cleveland UCI dataset, consisting of 297 patient data with 13 medical attributes, is used as the basis for this study. This dataset is processed through the Knowledge Discovery in Database (KDD) process, which includes data selection, pre-processing and cleaning data, data transformation, classification, interpretation, and knowledge presentation.

The methods used include data normalization, class balancing using SMOTE, and model evaluation through metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The results showed that the Logistic Regression model was able to provide predictions with an accuracy of 91.67%, precision of 100% for the class of patients with heart disease, recall of 82.14%, F1-score of 90.23%, and an AUC value of 0.95. These results indicate the model's ability to effectively distinguish between patients with and without heart disease.

This study contributes to the development of data-based prediction methods to assist in early diagnosis of heart disease. This model can be utilized by medical personnel and health institutions to improve efficiency and accuracy in decision making. Further researchers are advised to explore other algorithms or increase the dataset to optimize model performance.

Keywords: Heart Disease, Logistic Regression, Prediction, SMOTE, KDD