



**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK MEMPREDIKSI
PENYAKIT JANTUNG**

LAPORAN PENELITIAN

**MUHAMMAD FITRA RHOMADON
211420070**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
PALEMBANG
2025**



**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK MEMPREDIKSI
PENYAKIT JANTUNG**

**MUHAMMAD FITRA RHOMADON
211420070**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana
Komputer**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
PALEMBANG
2025**

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
MEMPREDIKSI PENYAKIT JANTUNG**

**MUHAMMAD FITRA RHOMADON
211420070**

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Teknik Informatika

Palembang, 15 Agustus 2025
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,

Pembimbing



WydyantoMM., M.Kom., Ph.D



Dr. Tata Sutabri, S.Kom., MMSI., MKM

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN


Skripsi Berjudul "PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK MEMPREDIKSI PENYAKIT JANTUNG" Oleh "Muhammad Fitra Rhomadon", telah dipertahankan di depan komisi penguji pada hari Jumat tanggal 15 Agustus 2025.

Komisi Penguji

1. Ketua : WydyantoMM., M.Kom.,Ph.D
2. Anggota : Dr. A. Haidar Mirza, S.T., M.Kom.
3. Anggota : Nurul Huda, M.Kom.



(.....)


(.....)


(.....)

Mengetahui,
Program Studi Teknik Informatika
Fakultas Sains Teknologi
Universitas Bina Darma
Ketua,


Universitas Bina Darma
Fakultas Sains Teknologi


Alek Wijaya, S.Kom., M.I.T.

SURAT PERNYATAAN

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Muhammad Fitra Rhomadon
NIM : 211420070

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, Agustus 2025
Yang membuat pernyataan,



Muhammad Fitra Rhomadon
NIM : 211420070

MOTTO DAN PERSEMBAHAN

MOTTO

"Kesuksesan bukanlah milik mereka yang pintar, tetapi milik mereka yang mau berusaha dengan sungguh-sungguh."

PERSEMBAHAN

Skripsi ini penulis persembahkan dengan penuh rasa syukur dan cinta kepada:

1. Allah SWT, atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua tercinta, yang selalu mendoakan, mendukung, dan menjadi sumber semangat dalam setiap langkah hidup penulis.
3. Dosen pembimbing pak Wydyanto MM., M.Kom.,Ph.D dan seluruh dosen pengajar, yang telah memberikan ilmu, bimbingan, serta motivasi selama masa perkuliahan.
4. Teman-teman seperjuangan, yang telah menjadi sahabat dalam suka dan duka selama masa studi.
5. Almamater tercinta, tempat penulis bertumbuh, belajar, dan meraih impian.

ABSTRAK

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di dunia, termasuk di Indonesia. Deteksi dini terhadap risiko penyakit jantung sangat penting untuk mencegah komplikasi yang lebih parah dan meningkatkan kualitas hidup pasien. Penelitian ini bertujuan untuk menerapkan algoritma *Logistic Regression* dalam membangun model prediksi penyakit jantung berbasis data. Dataset yang digunakan berasal dari Kaggle, dengan jumlah 1.025 data pasien dan 14 atribut yang mencakup faktor-faktor risiko seperti usia, jenis kelamin, tekanan darah, kolesterol, denyut jantung maksimum, dan lain-lain. Proses penelitian dilakukan menggunakan pendekatan CRISP-DM yang mencakup tahapan pemahaman bisnis, eksplorasi data, preprocessing, pemodelan, evaluasi, dan uji coba model. Tahapan preprocessing meliputi pembersihan data, encoding variabel kategorikal, standarisasi data numerik, dan pembagian data menjadi data latih dan data uji. Model dikembangkan menggunakan bahasa pemrograman Python dan pustaka scikit-learn, lalu dievaluasi menggunakan metrik akurasi, precision, recall, F1-score, confusion matrix, serta ROC-AUC. Hasil evaluasi menunjukkan bahwa model *Logistic Regression* mampu memberikan hasil prediksi yang baik, dengan nilai akurasi sebesar 0,93, precision 0,93, recall 0,96, dan F1-score 0,95. Dengan performa tersebut, model ini dapat digunakan sebagai alat bantu untuk tenaga medis dalam mendeteksi dini risiko penyakit jantung dan mendukung pengambilan keputusan secara lebih efektif dan efisien.

Kata Kunci: *Logistic Regression*, Penyakit Jantung, Prediksi, Machine Learning, Data Mining.

ABSTRACT

Heart disease is one of the leading causes of death worldwide, including in Indonesia. Early detection of heart disease risk is crucial to prevent more severe complications and improve patients' quality of life. This study aims to apply the Logistic Regression algorithm to build a data-driven heart disease prediction model. The dataset used is from Kaggle, with 1,025 patient data and 14 attributes covering risk factors such as age, gender, blood pressure, cholesterol, maximum heart rate, and others. The research process was conducted using the CRISP-DM approach, which includes business understanding, data exploration, preprocessing, modeling, evaluation, and model testing. The preprocessing stage includes data cleaning, encoding categorical variables, standardizing numeric data, and dividing the data into training and test data. The model was developed using the Python programming language and the scikit-learn library, then evaluated using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. The evaluation results showed that the Logistic Regression model was able to provide good prediction results, with an accuracy of 0.93, a precision of 0.93, a recall of 0.96, and an F1-score of 0.95. With this performance, this model can be used as a tool for medical personnel in early detection of heart disease risk and supporting more effective and efficient decision-making.

Keywords: Logistic Regression, Heart Disease, Prediction, Machine Learning, Data Mining.

KATA PENGANTAR

Puji syukur penulis panjatkan atas ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul **“Penerapan Algoritma Logistic Regression Untuk Memprediksi Penyakit Jantung”** ini sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Bina Darma Palembang.

Dalam proses penyusunan skripsi ini, penulis menyadari sepenuhnya bahwa tanpa bantuan, bimbingan, serta doa dari berbagai pihak, skripsi ini tidak akan terselesaikan dengan baik. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala nikmat, kekuatan, dan kesehatan yang telah diberikan selama proses pengerjaan skripsi ini.
2. Kedua orang tua tercinta, atas doa, dukungan moral dan material, serta semangat yang tiada henti dalam mendampingi setiap langkah penulis.
3. Bapak Wydyanto, MM., M.Kom., Ph.D., selaku dosen pembimbing yang telah dengan sabar memberikan arahan, bimbingan, dan motivasi kepada penulis selama penyusunan skripsi ini.
4. Ibu Prof. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang
5. Bapak Dr. Tata Sutabri, S.Kom., MMSI., MKM selaku Dekan Fakultas Sains Teknologi Universitas Bina Darma Palembang
6. Bapak Alek Wijaya, S.Kom., M.I.T. selaku Ketua Program Studi Teknik Informatika Universitas Bina Darma Palembang
7. Bapak Dr. A. Haidar Mirza, S.T., M.Kom. selaku Dosen Penguji
8. Ibu Nurul Huda, M.Kom. selaku Dosen Penguji
9. Seluruh dosen dan staf pengajar di Program Studi Teknik Informatika Universitas Bina Darma yang telah memberikan bekal ilmu dan pengalaman yang sangat berarti.
10. Teman-teman seperjuangan, atas kebersamaan, dukungan, dan bantuan yang diberikan selama masa perkuliahan maupun saat penyusunan skripsi.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis dengan rendah hati menerima segala bentuk kritik dan saran yang bersifat membangun demi penyempurnaan karya ini di masa mendatang.

Palembang, Agustus 2025

Penulis

DAFTAR ISI

HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN	iv
SURAT PERNYATAAN	v
MOTTO DAN PERSEMBAHAN	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	3
1.4 Ruang Lingkup dan Batasan Masalah	3
1.4.1 Ruang Lingkup	3
1.4.2 Batasan Masalah	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penyakit Jantung	5
2.2 Machine Learning	6
2.3 Data Mining	7
2.3.1 Proses Data Mining	7
2.3.2 Teknik-Teknik Data Mining	8
2.3.3 Studi Perbandingan Algoritma	8
2.3.4 Implikasi Penerapan Data Mining	8
2.4 Logistic Regression	8
2.4.1 Kelebihan Logistic Regression	9
2.4.2 Kekurangan Logistic Regression	9
2.4.3 Aplikasi Logistic Regression dalam Prediksi Penyakit Jantung	9
2.5 Preprocessing Data	10
2.6 Evaluation Model	11
2.7 Google Colaboratory	12
2.8 Python	13

2.9 Penelitian Terdahulu	14
BAB III METODOLOGI PENELITIAN	18
3.1 Pendekatan Penelitian	18
3.2 Metode Penelitian	18
3.3 Objek Penelitian.....	20
3.4 Teknik Pengumpulan Data.....	21
3.4.1 Jenis Data	21
3.4.2 Sumber Data	21
3.5 Teknik Analisis Data.....	22
3.6 Dataset Yang Digunakan	24
3.7 Alur Penelitian.....	27
BAB IV HASIL DAN PEMBAHASAN	30
4.1 Exploratory Data Analysis (EDA).....	30
4.1.1 Heatmap	30
4.2 Preprocessing Data.....	33
4.2.1 Cek Outlier.....	33
4.2.2 Menghapus Outlier Menggunakan Metode IQR (Interquartile Range).....	34
4.2.3 Cek Jumlah Data Setelah Outlier Dihapus & Cek Data Duplikat.....	36
4.2.4 Hapus Data Duplikat.....	36
4.2.5 Cek Missing Value	37
4.3 Modeling	37
4.3.1 Membagi Data Menjadi Fitur (X) & Target (y).....	38
4.3.2 Membagi Data Menjadi Data Pelatihan & Data Uji.....	38
4.3.3 Standarisasi Data	38
4.3.4 Pelatihan Model.....	39
4.3.5 Evaluasi Model	40
4.4 Membangun Sistem Prediksi.....	46
4.5 Deployment Ke Streamlit & Uji Coba Model	48
BAB V KESIMPULAN DAN SARAN	51
5.1 Kesimpulan	51
5.2 Saran	52
DAFTAR PUSTAKA	53
LAMPIRAN.....	57

DAFTAR GAMBAR

Gambar 3. 1 Metode CRISP-DM	18
Gambar 3. 2 Teknik Analisis Data.....	22
Gambar 3. 3 Info Dataset.....	25
Gambar 3. 4 5 Data Teratas Dataset	26
Gambar 3. 5 5 Data Terbawah Dataset.....	26
Gambar 3. 6 Deskriptif Dataset.....	26
Gambar 3. 7 Diagram Alur Penelitian	27
Gambar 4. 1 Heatmap	31
Gambar 4. 2 Boxplot Sebelum Outlier Dihapus.....	34
Gambar 4. 3 Menghapus Outlier.....	35
Gambar 4. 4 Boxplot Setelah Outlier Dihapus	35
Gambar 4. 5 Jumlah Data Setelah Outlier Dihapus & Jumlah Data Duplikat.....	36
Gambar 4. 6 Menghapus Data Duplikat.....	36
Gambar 4. 7 Cek Missing Value.....	37
Gambar 4. 8 Pembuatan & Pelatihan Model Logistic Regression.....	39
Gambar 4. 9 Akurasi Model Logistic Regression	40
Gambar 4. 10 Confusion Matrix.....	42
Gambar 4. 11 Classification Report.....	43
Gambar 4. 12 ROC-AUC Score	45
Gambar 4. 13 ROC Curve	45
Gambar 4. 14 Membangun Sistem Prediksi.....	47
Gambar 4. 15 Halaman Utama Project Prediksi Penyakit Jantung	49
Gambar 4. 16 Uji Coba Model Pada Data Orang Yang Tidak Ada Penyakit Jantung....	49
Gambar 4. 17 Uji Coba Model Pada Data Orang Yang Ada Penyakit Jantung	50

DAFTAR TABEL

Tabel 3. 1 Penjelasan Variabel Dalam Dataset	20
Tabel 4. 1 Tahapan Umum Dalam Preprocessing Data.....	33
Tabel 4. 2 Metrik Evaluasi Yang Umum Digunakan	40
Tabel 4. 3 Penjelasan TP, TN, FP & FN.....	41
Tabel 4. 4 Penjelasan Metrik Classification Report.....	43
Tabel 4. 5 Komponen ROC Curve	44
Tabel 4. 6 Interpretasi Nilai AUC	46
Tabel 4. 7 Langkah - Langkah Membangun Sistem Prediksi	47
Tabel 4. 8 Proses Umum Deployment.....	48



DAFTAR LAMPIRAN

1. Surat Keterangan LULUS Ujian Sarjana Skripsi
2. SK Pembimbing
3. Lembar Perbaikan Seminar Proposal
4. Lembar Perbaikan Seminar Hasil Penelitian
5. Lembar Pengajuan Judul
6. Lembar Penggantian Judul
7. Lembar Konsultasi Seminar Proposal
8. Lembar Konsultasi Skripsi
9. Turnitin
10. LOA Jurnal
11. Formulir Kelayakan Jilid Skripsi

