

**KLASIFIKASI PENYAKIT JANTUNG DENGAN  
ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC  
REGRESSION**

**SKRIPSI SARJANA INFORMATIKA**

Oleh

I KADEK AGGA SUGITHA

207064516133



**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA  
UNIVERSITAS NASIONAL  
2024**

**HALAMAN PENGESAHAN TUGAS AKHIR**

**KLASIFIKASI PENYAKIT JANTUNG DENGAN  
ALGORITMA K-NEAREST NEIGHBOR DAN  
LOGISTIC REGRESSION**



**I Kadek Agga Sugitha**

**207064516133**

Pembimbing I

A handwritten signature in black ink, appearing to be 'AT' with a flourish.

(Dr. Agung Triayudi, S.Kom., M.Kom)

NIDN. 0419068604

Pembimbing II

A handwritten signature in black ink, appearing to be 'E. Handayani' with a flourish.

(Ir. Endah Tri Esti Handayani, MMSI)

NIDN. 0314106802

## HALAMAN PERYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : I Kadek Agga Sugitha

NIM : 207064516133

Tanda Tangan :

Tanggal : 26 Agustus 2024

Mengetahui

Pembimbing 1 : Dr. Agung Triayudi, S.Kom., M.Kom. ( )

Pembimbing 2 : Ir. Endah Tri Esti Handayani, MMSI. ( )



**HALAMAN PENGESAHAN**  
**TUGAS SARJANA**  
**KLASIFIKASI PENYAKIT JANTUNG DENGAN ALGORITMA K-**  
**NEAREST NEIGHBOR DAN LOGISTIC REGRESSION**

Oleh:

I Kadek Agga Sugitha

207064516133

Program Studi Informatika  
Fakultas Teknologi Komunikasi dan Informatika

Disetujui pada tanggal: 26 Agustus 2024

Pembimbing I

Pembimbing II

Dr. Agung Triayudi, S.Kom., M.Kom.

NIDN. 0419068604

Ir. Endah Tri Esti Handayani, MMSI

NIDN. 0314106802

Penguji 1

Penguji 2

Dr. Andrianingsih, S.Kom., MMSI

NIDN. 0303097902

Agus Iskandar, S.Kom., M.Kom.

NIDN. 0310087503

UNIVERSITAS NASIONAL

**HALAMAN PERNYATAAN PERSETUJUAN  
PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN  
AKADEMIS**

Sebagai sivitas akademik Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika, Saya yang bertanda tangan di bawah ini:

Nama : I Kadek Agga Sugitha

NIM : 207064516133

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (Non-exclusive Royalti Free Right) atas karya ilmiah saya yang berjudul:

**Klasifikasi Penyakit Jantung Dengan Algoritma K-Nearest  
Neighbor Dan Logistic Regression**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Fakultas Teknologi Komunikasi dan Informatika berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 26 Agustus 2024

Yang menyatakan



(I Kadek Agga Sugitha)

## PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul:

### **KLASIFIKASI PENYAKIT JANTUNG DENGAN ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION**

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber Informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta 27 Agustus 2024



I Kadek Agga Sugitha

207064516133

## HALAMAN PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul:

### KLASIFIKASI PENYAKIT JANTUNG DENGAN ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Genap 2023 – 2024 pada tanggal 27 Agustus 2024

Pembimbing I



(Dr. Agung Triayudi, S.Kom., M.Kom)

NIDN. 0419068604

Pembimbing II



(Ir. Endah Tri Esti Handayani, MMSI)

NIDN. 0314106802

Ketua Program Studi Informatika



(Ratih Titi Komala Sari, ST., MM., MMSI)

NIDN. 0301038302

**HALAMAN PERSETUJUAN JUDUL  
YANG TIDAK ATAU YANG DIREVISI**

Nama : I Kadek Agga Sugitha  
NPM : 207064516133  
Fakultas / Akademi : Fakultas Teknologi Komunikasi dan Informatika  
Program Studi : Informatika  
Tanggal Sidang : 21 Agustus 2024

JUDUL DALAM BAHASA INDONESIA:

**KLASIFIKASI PENYAKIT JANTUNG DENGAN  
ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC  
REGRESSION**

JUDUL DALAM BAHASA INGGRIS:

***CLASSIFICATION OF HEART DISEASE USING K-NEAREST  
NEIGHBOR AND LOGISTIC REGRESSION ALGORITHM***

**TANDA TANGAN DAN TANGGAL**

<b>Pembimbing 1</b>	<b>Ka. Prodi</b>	<b>Mahasiswa</b>
TGL: 27/08/2024	TGL: 27/08/2024	TGL: 27/08/2024
		



**HALAMAN PERSETUJUAN JUDUL  
YANG TIDAK ATAU YANG DIREVISI**

Nama : I Kadek Agga Sugitha  
NPM : 207064516133  
Fakultas / Akademi : Fakultas Teknologi Komunikasi dan Informatika  
Program Studi : Informatika  
Tanggal Sidang : 21 Agustus 2024

JUDUL DALAM BAHASA INDONESIA:

**KLASIFIKASI PENYAKIT JANTUNG DENGAN  
ALGORITMA K-NEAREST NEIGHBOR DAN LOGISTIC  
REGRESSION**

JUDUL DALAM BAHASA INGGRIS:

**CLASSIFICATION OF HEART DISEASE USING K-NEAREST  
NEIGHBOR AND LOGISTIC REGRESSION ALGORITHM**

**TANDA TANGAN DAN TANGGAL**

<b>Pembimbing 2</b>	<b>Ka. Prodi</b>	<b>Mahasiswa</b>
TGL: 27/08/2024	TGL: 27/08/2024	TGL: 27/08/2024
		

## KATA PENGANTAR

Segala rasa syukur disampaikan kepada Tuhan Yang Maha ESA atas limpahan rahmat, karunia, dan petunjuk-Nya yang telah diberikan kepada penulis, memungkinkan penyelesaian skripsi ini dengan sukses. Mengenai judul penelitian ini, yaitu “Klasifikasi Penyakit Jantung Dengan Algoritma K-Nearest Neighbor Dan Logistic Regression”. Salah satu tujuan penelitian ini sebagai syarat memperoleh gelar Sarjana Komputer (S.Kom) dari Program Studi Informatika. di Universitas Nasional.

Skripsi ini dibuat berdasarkan hasil penelitian yang dilakukan dan Penulis menyadari jika tanpa bimbingan dan dukungan dari semua pihak dalam proses penelitian ini, tidak dapat selesai tepat pada waktunya dan dengan hasil yang baik. Oleh karena itu, izinkanlah penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Ayah, mama, dan kakak yang telah memberikan dukungan moril dan materil, serta doa kepada penulis sehingga skripsi ini dapat selesai dengan baik.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom. Selaku pembimbing 1 dan Ibu Ir. Endah Tri Esthi Handayani, MMSI. Selaku dosen pembimbing 2 yang telah menyediakan waktu dan pikirannya dalam membantu penulis menyelesaikan skripsi ini.
3. Teman-teman saya yang saling membantu, menyemangati dan berjuang bersama dalam mengerjakan skripsi.
4. Seluruh dosen FTKI Universitas Nasional yang telah memberikan pelajaran selama menempuh pendidikan.
5. Seluruh staff dan karyawan FTKI Universitas Nasional yang telah melayani penulis selama kuliah. Begitu juga dengan semua pihak yang tidak dapat disebutkan satu persatu, yang telah turut berkontribusi pada terwujudnya penulisan skripsi ini. Penulis memahami bahwa penulisan ini masih memiliki

kekurangan. Sebagai penutup, diharapkan skripsi ini dapat memberikan manfaat bagi penulis dan semua pihak yang terlibat dan pembacanya.

Jakarta, 26 Agustus 2014

Penulis



I Kadek Agga Sugitha

207064516133



## ABSTRAK

Penyakit jantung merupakan penyebab utama kematian di seluruh dunia, termasuk di Indonesia, dengan peningkatan prevalensi dan angka kematian yang menempatkan beban yang signifikan terhadap pelayanan kesehatan dan masyarakat. Kurangnya informasi gejala awal berkontribusi besar terhadap tantangan ini. Penelitian ini bertujuan untuk mencegah penyakit jantung melalui diagnosis dini dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dan Regresi Logistik. Dataset, yang bersumber dari Kaggle.com, terdiri dari 15 atribut klinis untuk diagnosis penyakit jantung. Pengujian menunjukkan bahwa metode K-NN dengan  $k = 3$  mencapai kinerja optimal pada data uji (30%), dengan akurasi 90%, presisi 93%, recall 87%, dan f1-score 90%. Sebagai perbandingan, Regresi Logistik dengan sigmoid mencapai akurasi 86%, presisi 83%, recall 90%, dan f1-score 86% pada data uji yang sama. Hasil ini menunjukkan bahwa K-Nearest Neighbor mengungguli Regresi Logistik sebagai algoritma klasifikasi untuk dataset penyakit jantung. Menerapkan temuan ini dalam kerangka kerja Streamlit berbasis web diharapkan dapat meningkatkan efisiensi dan ketepatan waktu pemantauan dan diagnosis penyakit jantung.

**Kata Kunci:** Penyakit Jantung, K-Nearest Neighbors, Logistic Regression, Klasifikasi, Perbandingan, Streamlit

## ABSTRACT

Heart disease is the leading cause of death globally, including in Indonesia, with increasing prevalence and mortality rates placing a significant burden on healthcare and society. A lack of early symptom information contributes substantially to this challenge. This study aims to prevent heart disease through early diagnosis using the K-Nearest Neighbor (K-NN) and Logistic Regression algorithms. The dataset, sourced from Kaggle.com, comprises 15 clinical attributes for heart disease diagnosis. Testing revealed that the K-NN method with  $k = 3$  achieved optimal performance on the test data (30%), with 90% accuracy, 93% precision, 87% recall, and 90% f1-score. In comparison, Logistic Regression with sigmoid achieved 86% accuracy, 83% precision, 90% recall, and an 86% f1-score on the same test data. These results indicate that K-Nearest Neighbor outperforms Logistic Regression as a classification algorithm for this heart disease dataset. Implementing these findings in the web-based Streamlit framework is expected to enhance the efficiency and timeliness of heart disease monitoring and diagnosis.

**Keywords:** Heart Disease, K-Nearest Neighbors, Logistic Regression, Classification, Comparison, Streamlit

# DAFTAR ISI

<b>KATA PENGANTAR</b> .....	i
<b>ABSTRAK</b> .....	iii
<b>ABSTRACT</b> .....	iv
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR GAMBAR</b> .....	vii
<b>DAFTAR TABLE</b> .....	viii
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Identifikasi Masalah</b> .....	3
<b>1.3 Tujuan Penelitian</b> .....	3
<b>1.4 Batasan Masalah</b> .....	3
<b>1.5 Kontribusi</b> .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	5
<b>2.1 Studi Literatur</b> .....	5
<b>2.2 Landasan Teori</b> .....	14
<b>2.2.1 Penyakit Jantung</b> .....	14
<b>2.2.2 Machine Learning</b> .....	15
<b>2.2.3 K-Nearest Neighbor</b> .....	16
<b>2.2.4 Logistic Regression</b> .....	18
<b>2.2.5 Kaggle</b> .....	19
<b>2.2.6 Python</b> .....	20
<b>2.2.7 Streamlit</b> .....	21
<b>2.2.8 VS Code</b> .....	21
<b>BAB III METODE PENELITIAN</b> .....	23
<b>3.1 Objek Penelitian</b> .....	23
<b>3.2 Fokus Penelitian</b> .....	23
<b>3.3 Sumber Data</b> .....	23

<b>3.4 Tahap Penelitian</b> .....	24
<b>3.4.1 Identifikasi masalah</b> .....	24
<b>3.4.2 Studi Literatur</b> .....	25
<b>3.4.2 Pengumpulan Data</b> .....	26
<b>3.4.4 Analisis Sistem</b> .....	29
<b>3.4.5 Model Klasifikasi</b> .....	30
<b>3.4.6 Perancangan Sistem</b> .....	34
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	47
<b>4.1 Perhitungan Manual</b> .....	47
<b>4.1.1 Algoritma K-Nearest Neighbors (K-NN)</b> .....	47
<b>4.1.2 Algoritma Logistic Regression</b> .....	50
<b>4.2 Hasil Evaluasi</b> .....	55
<b>4.2.1 Algoritma K-NN</b> .....	55
<b>4.2.2 Algoritma Logistic Regression</b> .....	64
<b>4.2 Hasil Perbandingan</b> .....	71
<b>4.3 Tampilan Interface</b> .....	74
<b>BAB V KESIMPULAN DAN SARAN</b> .....	83
<b>5.1 Kesimpulan</b> .....	83
<b>5.2 Saran</b> .....	83
<b>Daftar Pustaka</b> .....	85
<b>Lampiran</b> .....	90

## DAFTAR GAMBAR

<b>Gambar 2. 1</b> Anatomi Jantung.....	15
<b>Gambar 2. 2</b> Model K-Nearest Neighbor .....	16
<b>Gambar 2. 3</b> Model Logistic Regression.....	18
<b>Gambar 3. 1</b> Alur Tahapan Penelitian .....	24
<b>Gambar 3. 2</b> Dataset Penyakit Jantung.....	27
<b>Gambar 3. 3</b> Flowchart Sistem .....	34
<b>Gambar 3. 4</b> Korelasi Matrix Atribut .....	37
<b>Gambar 4. 1</b> Confusion Matrix Algoritma K-NN .....	55
<b>Gambar 4. 2</b> Grafik Cross-Validation K-NN.....	60
<b>Gambar 4. 3</b> Kurva ROC K-NN .....	62
<b>Gambar 4. 4</b> Confusion Matrix Logistic Regression.....	64
<b>Gambar 4. 5</b> Grafik Cross-Validation Logistic Regression.....	68
<b>Gambar 4. 6</b> Kurva ROC Logistic Regression .....	70
<b>Gambar 4. 7</b> Tampilan Dataset Pada web.....	75
<b>Gambar 4. 8</b> Tampilan Metode Klasifikasi .....	76
<b>Gambar 4. 9</b> Tampilan Pengujian Klasifikasi.....	77
<b>Gambar 4. 10</b> Tampilan Hasil Pengujian K-NN.....	79
<b>Gambar 4. 11</b> Tampilan Hasil Pengujian Logistic Regression.....	80
<b>Gambar 4. 12</b> Tampilan Hasil Perbandingan.....	81
<b>Gambar 4. 13</b> Tampilan Prediksi Jantung .....	82



## DAFTAR TABLE

<b>Table 3. 1</b> Variable Penelitian .....	27
<b>Table 3. 2</b> Confusion Matrix .....	42
<b>Table 4. 1</b> Sample Data .....	47
<b>Table 4. 2</b> Iterasi .....	54

