

SKRIPSI

**ANALISIS PERBANDINGAN PREDIKSI PENYAKIT
JANTUNG MENGGUNAKAN ALGORITMA *LOGISTIC*
REGRESSION DAN *K-NEAREST NEIGHBOR***



ABYAN OZAGA

1519620026

**PROGRAM STUDI
SISTEM DAN TEKNOLOGI INFORMASI**

FAKULTAS TEKNIK

UNIVERSITAS NEGERI JAKARTA

2025

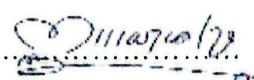

LEMBAR PENGESAHAN

Judul : Analisis Perbandingan Prediksi Penyakit Jantung Menggunakan
Algoritma Klasifikasi Logistic Regression Dan Algoritma Klasifikasi
K-Nearest Neighbors.

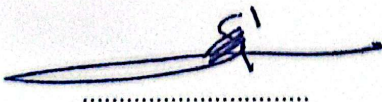

Penyusun : Abyan Ozaga

NIM : 1519620026

Disetujui Oleh:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Murien Nugraheni, S.T., M.Cs (Dosen Pembimbing 1)		5 September 2024
Dr. Widodo, S.Kom., M.Kom (Dosen Pembimbing 2)		11 September 2024

PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Lipur Sugiyanta, Ph.D (Dosen Penguji 1)		3 Februari 2025
Irma Permata Sari, M.Eng. (Dosen Penguji 2)		23 Desember 2024

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 10 Februari 2025
Yang Membuat Pernyataan



Abyan Ozaga
NIM. 1519620026



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Abyan Ozaga

NIM : 1519620026

Fakultas/Prodi : Fakultas Teknik/Sistem dan Teknologi Informasi

Alamat email : ABYANOZAGA_1519620026@mhs.unj.ac.id

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi Tesis Disertasi Lain-lain (.....)

yang berjudul :

Analisis Perbandingan Prediksi Penyakit Jantung Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 10 Februari 2025

Penulis

(Abyan Ozaga)

KATA PENGANTAR

Puji dan syukur tercurahkan atas kehadiran Allah SWT, Tuhan Yang Maha Esa yang telah memberikan rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisi Perbandingan Prediksi Penyakit Jantung Menggunakan Algoritma Klasifikasi Logistic Regression Dan Algoritma Klasifikasi K-Nearest Neighbors” yang merupakan salah satu syarat untuk meraih gelar Sarjana Program Studi Sistem Dan Teknologi Informasi pada Fakultas Teknik, Universitas Negeri Jakarta. Shalawat serta salam tak lupa penulis curahkan kepada junjungan Nabi Besar Muhammad SAW. beserta keluarga, sahabat serta kerabat, semoga selalu mendapatkan syafa’at dari beliau. Aamiin.

Dalam proses penyusunan dan penulisan skripsi ini dari awal hingga selesai dapat terwujud berkat bimbingan, dorongan, bantuan serta saran-saran dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua, yaitu Bapak Mirsa Mandong dan Ibu Ita Ghaula, serta Kakak Indira Naratisa yang selalu mendoakan yang terbaik, memberikan semangat, dorongan materi dan moral kepada penulis dalam menyelesaikan skripsi ini.
2. Bapak Lipur Sugiyanta, Ph.D. selaku Koordinator Program Studi Sistem Dan Teknologi Informasi Fakultas Teknik Universitas Negeri Jakarta.
3. Ibu Murien Nugraheni, S.T., M.Cs. selaku dosen pembimbing I yang selalu memberikan kesediaan waktunya kapanpun, memberikan saran dan arahan, memberikan ide, bimbingan serta motivasi dalam penulisan skripsi ini sehingga penulis dapat menyelesaikan skripsi dengan lancar.
4. Bapak Dr. Widodo, S.Kom., M.Kom. selaku dosen pembimbing II yang selalu memberikan kesediaan waktunya kapanpun, memberikan saran yang membangun dalam bimbingan, kesempatan dan ide serta masukan-masukan dalam penyusunan dan pembuatan skripsi ini sehingga penulis dapat menyelesaikannya dengan lancar.

5. Teman-teman STI 2020, 2021, 2022, 2023 yang senantiasa membantu dan memberikan saran, motivasi, pengalaman dan doa dalam penyusunan skripsi.
6. Teruntuk teman-teman dekat saya yaitu: Arya, Ryan, Kharil, Fauzi, dan Ramadhan yang senantiasa membantu dan memberikan saran kepada penulis dalam penyusunan skripsi ini.
7. Seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa skripsi ini jauh dari kesempurnaan, untuk itu penulis mohon maaf apabila terdapat kekurangan dan kesalahan, sehingga kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan skripsi ini. Akhir kata, penulis berharap agar penelitian dan penyusunan skripsi ini dapat bermanfaat bagi pihak yang membaca dan khususnya bagi penulis.

Jakarta, 10 Februari 2025

Yang Membuat Pernyataan



Abyan Ozaga

NIM. 1519620026

Intelligentia - Dignitas

ABSTRAK

Abyan Ozaga, Analisis Perbandingan Prediksi Penyakit Jantung Menggunakan Algoritma Klasifikasi Logistic Regression Dan Algoritma Klasifikasi K-Nearest Neighbors. Skripsi. Program Studi Sistem Dan Teknologi Informasi, Fakultas Teknik Universitas Negeri Jakarta 2025. Dosen Pembimbing : Murien Nungraheni S.T.,M.Cs dan Dr. Widodo, S.Kom., M.Kom.

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, sehingga deteksi dini sangat krusial dalam upaya pencegahan dan pengobatan. Penelitian ini bertujuan membandingkan performa algoritma *K-Nearest Neighbors* (KNN) dan Logistic Regression dalam memprediksi risiko penyakit jantung berdasarkan data medis pasien yang diperoleh dari *UCI Heart Disease Dataset*. Dataset ini mencakup berbagai parameter klinis seperti tekanan darah (*trestbps*), kadar kolesterol (*chol*), detak jantung maksimum (*thalach*), serta indikator lainnya yang berhubungan dengan kondisi jantung. Metode *K-Fold Cross Validation* (5 fold) digunakan untuk membagi data guna memastikan hasil yang lebih akurat dan menghindari bias, sementara *confusion matrix* digunakan sebagai metrik evaluasi untuk menilai efektivitas model dalam mengklasifikasikan pasien dengan dan tanpa penyakit jantung. Hasil pengujian menunjukkan bahwa *Logistic Regression* memperoleh akurasi 89%, sedangkan *K-Nearest Neighbors* (KNN) menghasilkan akurasi 80%, yang menunjukkan bahwa Logistic Regression lebih unggul dalam mengenali pola data yang terkait dengan penyakit jantung. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan sistem prediksi berbasis machine learning yang lebih akurat dan dapat membantu tenaga medis dalam mendeteksi potensi penyakit jantung secara lebih dini dan efisien.

Kata Kunci: *Machine Learning*, Penyakit Jantung, *K-Nearest Neighbors*, *Logistic Regression*, *K-Fold Cross Validation*, *Confusion Matrix*, Prediksi Penyakit

ABSTRACT

Abyan Ozaga, *Comparative Analysis of Heart Disease Prediction Using Logistic Regression and K-Nearest Neighbor Algorithms*. Thesis. Information Systems And Technology Study Program, Faculty of Engineering, State University of Jakarta 2025. Supervisor : Murien Nungraheni S.T.,M.Cs and Dr. Widodo, S.Kom.,. M.Kom.

Heart disease is one of the leading causes of death worldwide, making early detection crucial for prevention and treatment. This study aims to compare the performance of the K-Nearest Neighbors (KNN) and Logistic Regression (LogReg) algorithms in predicting heart disease risk based on medical patient data obtained from the UCI Heart Disease Dataset. The dataset includes various clinical parameters such as blood pressure (trestbps), cholesterol levels (chol), maximum heart rate (thalach), and other indicators related to heart conditions. The K-Fold Cross Validation (5-fold) method was applied to partition the data, ensuring more accurate results and reducing bias, while the confusion matrix was used as an evaluation metric to assess the effectiveness of the models in classifying patients with and without heart disease. The results show that Logistic Regression achieved an accuracy of 89%, while K-Nearest Neighbors (KNN) obtained 80%, indicating that Logistic Regression outperforms KNN in recognizing data patterns associated with heart disease. Thus, this study contributes to the development of a more accurate machine learning-based prediction system that can assist healthcare professionals in detecting potential heart disease cases earlier and more efficiently.

Keywords: Machine Learning, Heart Disease, K-Nearest Neighbors, Logistic Regression, K-Fold Cross Validation, Confusion Matrix, Disease Prediction.

Intelligentia - Dignitas

DAFTAR ISI

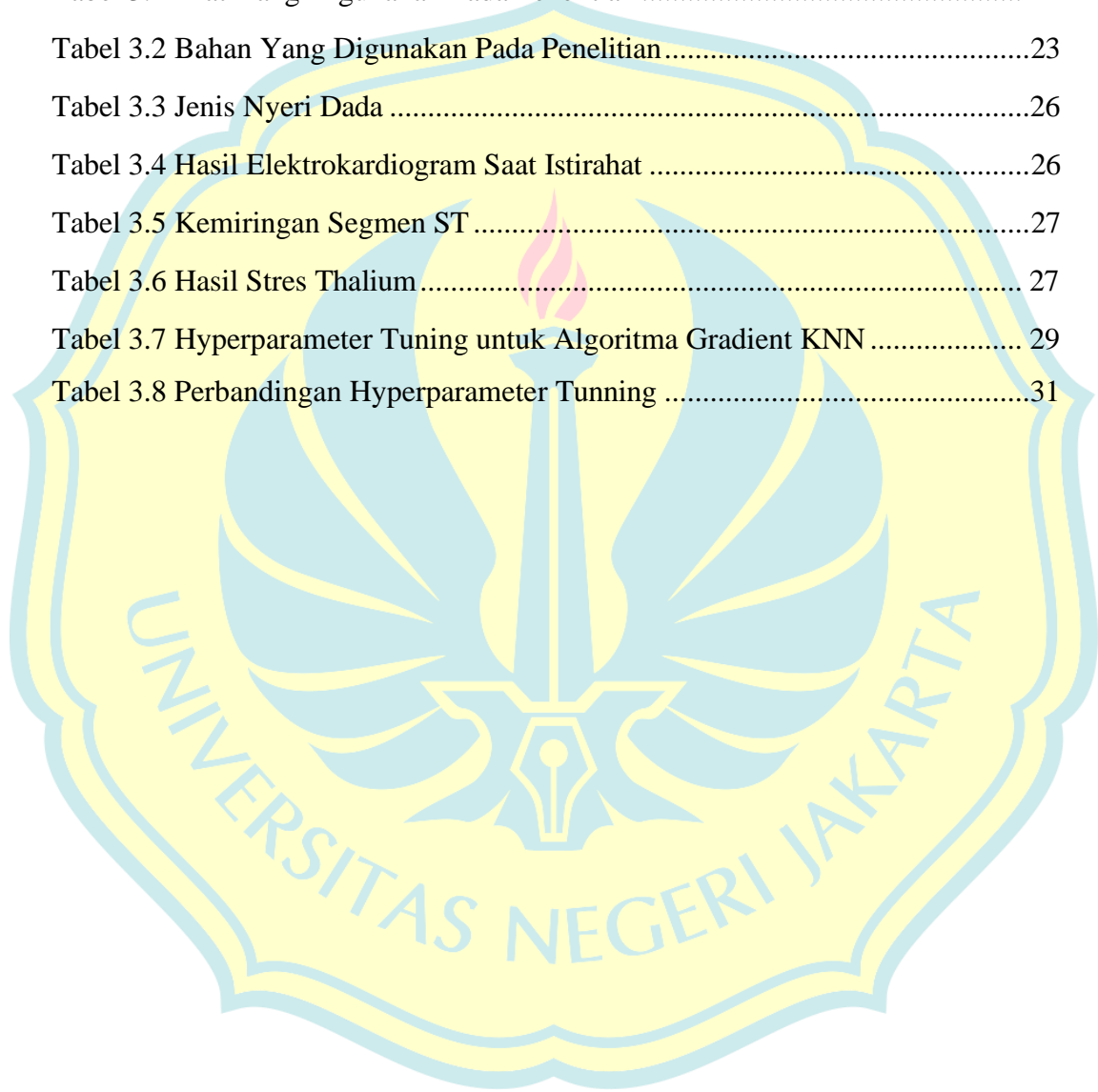
LEMBAR PENGESAHAN SKRIPSI.....	ii
LEMBAR PERNYATAAN	iii
KATA PENGANTAR.....	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah.....	3
1.3 Pembatasan Masalah.....	4
1.4 Perumusan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian.....	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Kerangka Teoritik	7
2.2 Penelitian Relevan.....	15
2.3 Kerangka Berpikir.....	17
2.4 <i>State of the art</i>	21
BAB III METODE PENELITIAN.....	22
3.1 Alat dan Bahan Tugas Akhir.....	22
3.1.1 Alat	22
3.1.2 Bahan.....	23
3.2 Alur Penelitian	24
3.3 Penjabaran Langkah Penelitian	25
3.3.1 Pengumpulan Dataset.....	25
3.3.2 Pembersihan Data.....	27
3.3.3 Pembagian Data.....	28
3.3.4 Pembangunan Model Klasifikasi.....	28
3.3.5 Penyetelan Parameter (<i>Hyperparameter Tuning</i>).....	29
3.3.6 Evaluasi	33

BAB IV HASIL DAN PEMBAHASAN.....	34
4.1 Deskripsi Data.....	34
4.1.1 Data Load.....	34
4.1.2 Pembersihan Data.....	35
4.1.2.3 Pembersihan Outliers.....	36
4.1.2.4 Penyeimbangan Data.....	37
4.1.2.5 Feature Encoding.....	38
4.1.3 Exploratory Data Analysis.....	39
4.1.3.1 Mengelaborasi features dan tipe data.....	40
4.1.3.2 Penggunaan Mean, Median, dan Modus.....	41
4.1.3.3 Mendeskripsikan Statistical Five Summaries.....	41
4.1.3.4 Mendeskripsikan Distribusi Data.....	42
4.1.4 Penerapan Feature Engineering.....	44
4.2 Pengujian Penelitian.....	45
4.2.1 Menyiapkan K-Fold Cross Validation.....	46
4.2.2 Melakukan Pelatihan Dan Evaluasi Model.....	46
4.2.3 Menghitung Rata-Rata Akurasi.....	48
4.2.4 Penyetelan Hyperparameter Tuning.....	48
4.2.5 Menampilkan Classification Report Dan Confusion Matrix.....	49
BAB V KESIMPULAN DAN SARAN.....	51
5.1 Kesimpulan.....	51
5.2 Saran.....	51
DAFTAR PUSTAKA.....	52
LAMPIRAN-LAMPIRAN.....	55

Intelligentia - Dignitas

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	12
Tabel 2.2 Penelitian Yang Relevan	15
Tabel 3.1 Alat Yang Digunakan Pada Penelitian	22
Tabel 3.2 Bahan Yang Digunakan Pada Penelitian.....	23
Tabel 3.3 Jenis Nyeri Dada	26
Tabel 3.4 Hasil Elektrokardiogram Saat Istirahat	26
Tabel 3.5 Kemiringan Segmen ST	27
Tabel 3.6 Hasil Stres Thalium.....	27
Tabel 3.7 Hyperparameter Tuning untuk Algoritma Gradient KNN	29
Tabel 3.8 Perbandingan Hyperparameter Tuning	31



Intelligentia - Dignitas

DAFTAR GAMBAR

Gambar 2.1 Kerangka Berpikir	17
Gambar 3.1 Alur Penelitian	24
Gambar 3.2 Penerapan KNN	29
Gambar 3.3 Penerapan Logistic Regression	29
Gambar 4.1 Load Dataset	34
Gambar 4.2 Output load dataset	35
Gambar 4.3 Pembersihan Missing Value	35
Gambar 4.4 Total Duplikat Pada Data	36
Gambar 4.5 Penghapusan Data Duplikat	36
Gambar 4.6 Penghapusan Outliers	37
Gambar 4.7 Penyeimbangan Data Menggunakan Random Sample	38
Gambar 4.8 Feature Encoding	39
Gambar 4.9 Pencarian Mean, Median, Dan Modus Pada Setiap Atribut	41
Gambar 4.10 Statistical Five Summaries	41
Gambar 4.11 Distribusi Data Umur	42
Gambar 4.12 Distribusi Data Tekanan Darah Saat Istirahat	42
Gambar 4.13 Distribusi Data Detak Jantung Maximum	43
Gambar 4.14 Distribusi Data Serum Kolesterol	43
Gambar 4.15 Distribusi Data Depresi ST yang Dipicu Olahraga	44
Gambar 4.16 Feature Engineering	45
Gambar 4.17 Penerapan K-Fold	46
Gambar 4.18 Pelatihan Dan Evaluasi Model	47
Gambar 4.19 Hasil Akurasi Dari Setiap Fold	47
Gambar 4.20 Rata-Rata Akurasi Dari Setiap Fold	48
Gambar 4.21 Hyperparameter Tuning Untuk Kedua Algoritma	48
Gambar 4.22 Classification Report dan Confusion Matrix	49
Gambar 4.23 Hasil Algoritma Logistic Regression	49

DAFTAR LAMPIRAN

Lampiran 1. Surat Tugas Dosen Pembimbing.....	55
Lampiran 2. Lembar Konsultasi Pembimbing 1.....	56
Lampiran 3. Lembar Konsultasi Pembimbing 2.....	57
Lampiran 4. Surat Pernyataan Dosen Pembimbing 1.....	58
Lampiran 5. Surat Pernyataan Dosen Pembimbing 2.....	59
Lampiran 6. Sumber Dataset Yang Digunakan.....	60
Lampiran 7. Sampel Dataset.....	61
Lampiran 8. Bukti Submit Paper.....	62



Intelligentia - Dignitas