

**KLASTERISASI DATA PENYAKIT JANTUNG
MENGUNAKAN *K-MEANS* DALAM SISTEM INFORMASI
KESEHATAN**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh:

Saskia Adinda

217006516064



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2024

**KLASTERISASI DATA PENYAKIT JANTUNG
MENGUNAKAN *K-MEANS* DALAM SISTEM INFORMASI
KESEHATAN**

SKRIPSI SARJANA SISTEM INFORMASI

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh:

Saskia Adinda

217006516064



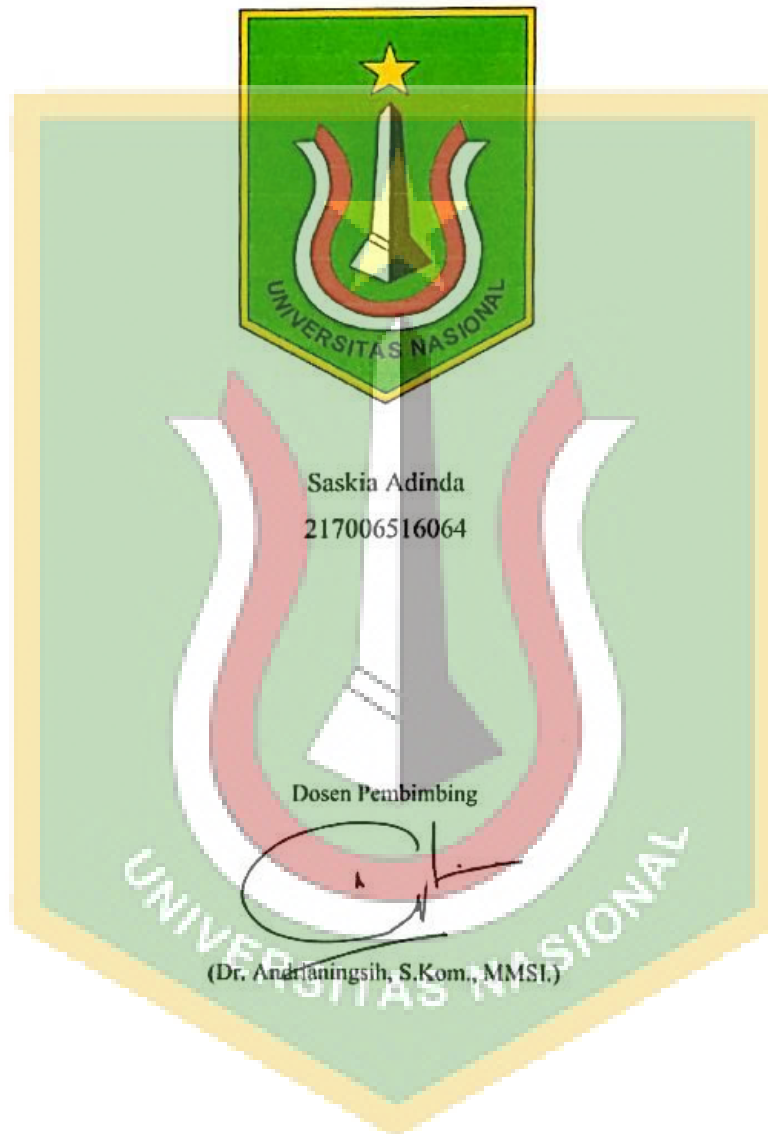
**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2024

HALAMAN PENGESAHAN

TUGAS AKHIR

KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN K- MEANS DALAM SISTEM INFORMASI KESEHATAN



HALAMAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN *K-MEANS* DALAM SISTEM INFORMASI KESEHATAN

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi 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.



LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

KLASTERISASI DATA PENYAKIT JANTUNG MENGUNAKAN K-MEANS DALAM SISTEM INFORMASI KESEHATAN

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang **Review Akhir Semester Ganjil 2024-2025** pada tanggal 28 Februari Tahun 2025



LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Saskia Adinda
NPM : 217006516064
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 26 Februari 2025

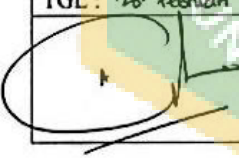
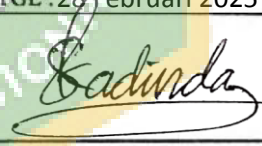
JUDUL DALAM BAHASA INDONESIA :

**KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN K-MEANS
DALAM SISTEM INFORMASI KESEHATAN**

JUDUL DALAM BAHASA INGGRIS :

***CLUSTERING HEART DISEASE DATA USING K-MEANS IN HEALTH
INFORMATION SYSTEMS***

TANDA TANGAN DAN TANGGAL

Pembimbing I	Pembimbing II / Prodi	Mahasiswa
TGL : 26 Februari 2025	TGL : 26 Februari 2025	TGL : 28 Februari 2025
		

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang senantiasa melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul "KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN *K-MEANS* DALAM SISTEM INFORMASI KESEHATAN". Skripsi ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.

Skripsi ini bertujuan untuk mengembangkan analisis berbasis data dalam sistem informasi kesehatan guna membantu klasterisasi pasien berdasarkan tingkat risiko penyakit jantung. Dengan memanfaatkan algoritma *K-Means Clustering*, penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap pengelolaan data kesehatan yang lebih efisien dan mendukung pengambilan keputusan medis yang lebih tepat.

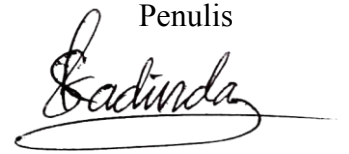
Penyelesaian skripsi ini tentunya tidak terlepas dari doa, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas segala nikmat kesehatan, kekuatan, dan kemudahan yang diberikan selama proses penyelesaian skripsi ini.
2. Orang tua tercinta ibu Elinda Roza dan bapak Agus Mahyudin serta kedua kakak penulis Sarah Amanda, S.Kom dan Salsa Bila Ayunda, S.IKom, yang dengan penuh kasih sayang selalu mendoakan, mendukung, dan memberikan semangat tanpa henti kepada penulis di setiap langkah kehidupan.
3. Diri sendiri yang telah berjuang dengan penuh kesabaran dan ketekunan untuk menyelesaikan skripsi ini meskipun menghadapi berbagai tantangan.
4. Ibu Dr. Andrianingsih, S.Kom., MMSI., selaku Ketua Program Studi Sistem Informasi sekaligus dosen pembimbing tunggal, yang telah dengan sabar memberikan bimbingan, arahan, dan masukan yang sangat berarti selama proses penyusunan skripsi ini.
5. Rekan-rekan mahasiswa dan teman seperjuangan, yang telah memberikan dukungan moral, saran, kebersamaan, serta semangat dan bantuan baik langsung maupun tidak langsung dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat bagi semua pihak, khususnya dalam pengembangan sistem informasi kesehatan di Indonesia.

Jakarta, 20 November 2025

Penulis



Saskia Adinda



ABSTRAK

Penyakit jantung merupakan salah satu isu kesehatan utama yang memerlukan perhatian serius dalam pengelolaannya, terutama dalam menentukan skala prioritas penanganan pasien. Penelitian ini bertujuan untuk menerapkan metode *K-Means Clustering* dalam analisis data pasien penyakit jantung. Data yang digunakan diperoleh dari *Heart Disease UCI Dataset* yang tersedia di *Kaggle*, mencakup berbagai atribut seperti usia, jenis kelamin, jenis nyeri dada, tekanan darah, kadar kolesterol, denyut jantung, perubahan depresi segmen ST, kemiringan segmen ST, serta jumlah pembuluh darah yang terdeteksi.

Tahapan penelitian ini diawali dengan proses pembersihan data (*data cleaning*), normalisasi, dan seleksi fitur yang relevan. Setelah itu, algoritma *K-Means* diterapkan untuk mengelompokkan pasien ke dalam beberapa klaster berdasarkan kemiripan karakteristik medis mereka. Pendekatan ini bertujuan untuk menghasilkan klaster yang dapat membantu tenaga medis dalam menentukan prioritas penanganan pasien berdasarkan tingkat risiko penyakit jantung yang mereka miliki.

Metode *K-Means Clustering* dipilih dalam penelitian ini karena kemampuannya dalam mengelompokkan data secara efisien berdasarkan karakteristik tertentu. Data pasien penyakit jantung yang telah dikumpulkan diproses untuk menghasilkan klasterisasi yang dapat merepresentasikan pola tertentu dalam dataset. Selain itu, penelitian ini juga mengeksplorasi bagaimana penerapan *K-Means Clustering* dalam sistem informasi kesehatan rumah sakit dapat meningkatkan efisiensi pengelolaan data pasien dan membantu tenaga medis dalam mengambil keputusan yang lebih cepat dan akurat.

Hasil penelitian menunjukkan bahwa metode *K-Means Clustering* mampu mengelompokkan pasien dengan baik berdasarkan tingkat risiko penyakit jantung. Hasil klasterisasi ini dapat diintegrasikan ke dalam sistem informasi kesehatan rumah sakit, sehingga tenaga medis dapat lebih mudah dalam memprioritaskan penanganan pasien. Dengan demikian, penelitian ini memberikan kontribusi dalam peningkatan kualitas layanan kesehatan dengan memanfaatkan teknologi analitik untuk pengelolaan data pasien penyakit jantung.

Kata kunci: *K-Means Clustering*, Penyakit Jantung, Sistem Informasi Kesehatan, Klasterisasi, *Heart Disease UCI Dataset*.

ABSTRACT

Heart disease is one of the major health issues that requires serious attention in its management, especially in determining the priority scale of patient treatment. This study aims to apply the K-Means Clustering method in the analysis of heart disease patient data. The data used is obtained from the Heart Disease UCI Dataset available on Kaggle, including various attributes such as age, gender, type of chest pain, blood pressure, cholesterol levels, heart rate, changes in ST segment depression, ST segment slope, and the number of blood vessels detected.

The stages of this research began with data cleaning, normalization, and selection of relevant features. After that, the K-Means algorithm was applied to group patients into clusters based on the similarity of their medical characteristics. This approach aims to generate clusters that can assist medical personnel in prioritizing patient treatment based on their level of heart disease risk.

The K-Means Clustering method was chosen in this study due to its ability to efficiently group data based on certain characteristics. The collected heart disease patient data is processed to produce a clustering that can represent certain patterns in the dataset. In addition, this study also explores how the application of K-Means Clustering in hospital health information systems can improve the efficiency of patient data management and assist medical personnel in making faster and more accurate decisions.

The results show that the K-Means Clustering method is able to cluster patients well based on the risk level of heart disease. The results of this clustering can be integrated into the hospital's health information system, so that medical personnel can more easily prioritize patient treatment. Thus, this research contributes to improving the quality of health services by utilizing analytical technology for the management of heart disease patient data.

Keywords: K-Means Clustering, Heart Disease, Health Information System, Clustering, Heart Disease UCI Dataset.

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
HALAMAN KEASLIAN TUGAS AKHIR	ii
LEMBAR PERSETUJUAN TUGAS AKHIR.....	iii
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	iv
KATA PENGANTAR.....	v
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	vii
DAFTAR RUMUS	viii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	4
1.5 Kontribusi Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Landasan Teori	6
2.2 Research Positioning.....	24
2.3 Technology Information System	25
2.3.1 Data Mining	25
2.3.2 Klasterisasi	27
2.3.3 K-Means.....	27
2.3.4 Python	28
2.3.5 Google Colab	29

2.3.6	Penyakit Jantung	29
2.3.7	Sistem Manajemen Informasi Kesehatan.....	30
2.3.8	Streamlit.....	30
BAB III METODOLOGI PENELITIAN		31
3.1	Tahap Penelitian	31
3.2	Waktu Penelitian	31
3.3	Penentuan Subjek Penelitian	32
3.4	Fokus Penelitian	33
3.5	Sumber Data	33
3.6	Desain Penelitian	33
3.7	Pengumpulan Data.....	37
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Implantasi Tahapan Penelitian	39
4.1.1	Business Understanding	39
4.1.2	Data Understanding	39
4.1.3	Data Preparation.....	40
4.1.4	Modeling	42
4.1.5	Evaluation.....	44
4.1.6	Deployment.....	47
4.2	Hitungan Manual	52
BAB V KESIMPULAN		75
5.1	Kesimpulan	75
5.2	Saran.....	75
DAFTAR PUSTAKA.....		76
LAMPIRAN.....		78

DAFTAR GAMBAR

Gambar 2. 1. Research Positioning.....	24
Gambar 2. 2. CRISP-DM.....	25
Gambar 2. 3 Google Colab	29
Gambar 2. 4 Streamlit.....	30
Gambar 3. 1. Waktu Penelitian	31
Gambar 3. 2 Alur Tahapan Penelitian.....	34
Gambar 3. 3 Penerapan K-Means Clustering	36
Gambar 4. 1 Missing Value.....	40
Gambar 4. 2 Normalisasi Data.....	41
Gambar 4. 3 Atribut yang digunakan.....	42
Gambar 4. 4 Proses Klasterisasi	43
Gambar 4.5. 13 variable dengan 3 cluster.....	44
Gambar 4.6. 13 variable dengan 4 cluster.....	44
Gambar 4.7. 13 variable 2 cluster.....	45
Gambar 4. 8. 9 variable 2 cluster.....	45
Gambar 4. 9 Tampilan Dashboard	47
Gambar 4. 10 Tampilan Dataset	48
Gambar 4. 11 Tampilan Hasil Klasterisasi	49
Gambar 4. 12 Tampilan Statistik setiap Klaster	50
Gambar 4. 13 Tampilan Visualisasi Klasterisasi.....	51

DAFTAR TABEL

Tabel 2. 1. Reference Journal.....	6
Tabel 4. 1. 10 Dataset Penyakit Jantung	52
Tabel 4. 2. Dataset yang diperlukan.....	52
Tabel 4. 3. Dataset Setelah Normalisasi	64
Tabel 4. 4. Hasil Clustering.....	70

DAFTAR RUMUS

(1).....	27
(2).....	27
(3).....	28
(2).....	36
(3).....	36
(4).....	53
(5).....	53

