

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

Saskia Adinda^{1*}, Andrianingsih Andrianingsih²

^{1,2}Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Indonesia

¹saskiaadinda2021@student.unas.ac.id, ²andrianingsih@civitas.unas.ac.id

Received: 25-03-2025

Revised: 25-04-2025

Approved: 30-05-2025

ABSTRACT

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 menggunakan Heart Disease UCI Dataset dari Kaggle, yang mencakup berbagai atribut medis seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan jumlah pembuluh darah yang terdeteksi. Tahapan penelitian mencakup pembersihan data, normalisasi, seleksi fitur, serta penerapan algoritma K-Means untuk mengelompokkan pasien berdasarkan kemiripan karakteristik medis mereka. Hasil klusterisasi ini bertujuan untuk membantu tenaga medis dalam menentukan prioritas penanganan pasien berdasarkan tingkat risiko penyakit jantung. Evaluasi menunjukkan bahwa metode K-Means Clustering mampu mengelompokkan pasien dengan baik dan dapat diintegrasikan ke dalam sistem informasi kesehatan rumah sakit untuk meningkatkan efisiensi pengelolaan data serta mempercepat pengambilan keputusan medis. Dengan demikian, penelitian ini berkontribusi dalam pemanfaatan teknologi analitik guna meningkatkan kualitas layanan kesehatan dalam menangani pasien penyakit jantung.

Keywords: K-Means Clustering, Penyakit Jantung, Sistem Informasi Kesehatan, Klusterisasi, Heart Disease UCI Dataset.

PENDAHULUAN

Penyakit jantung masih menjadi salah satu faktor utama penyebab kematian di berbagai negara. Berdasarkan laporan World Health Organization (WHO), sekitar 17,9 juta jiwa meninggal setiap tahunnya akibat penyakit kardiovaskular, yang mencakup sekitar 31% dari total kematian global (WHO, 2020). Deteksi dini memiliki peran penting dalam menekan angka kematian ini. Oleh sebab itu, pengembangan sistem prediksi yang akurat dan efisien dapat membantu dalam proses diagnosis serta pencegahan penyakit jantung agar tidak berkembang ke tahap yang lebih parah.

Kemajuan teknologi informasi, terutama dalam pembelajaran mesin, telah membawa dampak besar bagi sektor kesehatan, khususnya dalam upaya prediksi penyakit. Salah satu algoritma yang sering digunakan dalam proses pengelompokan data pasien adalah K-Means Clustering. Metode ini berfungsi untuk mengelompokkan pasien berdasarkan berbagai faktor risiko seperti usia, tekanan darah, kadar kolesterol, serta indikator medis lainnya. K-Means Clustering bekerja dengan membagi dataset ke dalam beberapa kelompok berdasarkan kedekatan jarak antara titik data dan pusat kluster (Ramesh & Anusha, 2022).

Penelitian yang dilakukan oleh Hassan et al. (2023) mengungkapkan bahwa penerapan K-Means Clustering dalam kombinasi dengan teknik pembelajaran mesin lainnya, seperti Naïve Bayes atau Logistic Regression, dapat meningkatkan akurasi prediksi. Studi lain menunjukkan bahwa menggabungkan K-Means dengan Decision Tree menghasilkan identifikasi faktor risiko utama penyakit jantung yang lebih akurat serta mengurangi tingkat kesalahan klasifikasi (Sandeep et al., 2021).

Lebih lanjut, pendekatan hybrid yang mengintegrasikan K-Means dengan teknik lain, sebagaimana diusulkan oleh Alazab et al. (2021), terbukti meningkatkan ketepatan prediksi dengan menggabungkan berbagai metode pembelajaran mesin. Sistem ini memungkinkan pengelompokan pasien ke dalam kategori risiko rendah, sedang, dan tinggi, sehingga sangat bermanfaat dalam pengambilan keputusan medis.

Seiring meningkatnya kebutuhan prediksi penyakit yang lebih cepat dan akurat, K-Means Clustering menjadi salah satu pendekatan yang efektif dalam menentukan risiko penyakit jantung. Metode ini mampu menangani data dalam jumlah besar dengan berbagai karakteristik pasien yang berbeda (Bashar et al., 2020).

Studi yang dilakukan di RS Jantung Harapan Kita pada tahun 2020 mencatat adanya lonjakan jumlah pasien penyakit jantung yang memerlukan perawatan intensif. Data dari rumah sakit menunjukkan bahwa antara tahun 2017 hingga 2020, jumlah pasien dengan penyakit jantung koroner meningkat sekitar 15% setiap tahunnya. Kelompok usia produktif (30-50 tahun) mendominasi kasus ini, dengan faktor risiko utama meliputi hipertensi, obesitas, dan gaya hidup yang tidak sehat.

Untuk mengatasi lonjakan pasien ini, RS Jantung Harapan Kita mulai mengadopsi sistem manajemen informasi kesehatan yang lebih terstruktur. Namun, analisis data pasien masih menjadi tantangan karena sistem tradisional kurang efisien dalam menangani volume data yang terus meningkat. Oleh karena itu, penelitian ini berfokus pada penerapan K-Means Clustering untuk mengelompokkan pasien berdasarkan riwayat kesehatan mereka, sehingga dapat meningkatkan akurasi diagnosis serta efektivitas penanganan medis.

Berbagai penelitian telah membuktikan efektivitas K-Means Clustering dalam pengelolaan data medis. Misalnya, studi oleh Putra et al (2020) yang menerapkan metode ini pada data pasien rumah sakit di Surabaya menunjukkan bahwa pengelompokan pasien berdasarkan gejala dan riwayat keluarga mampu mempercepat proses pengambilan keputusan medis dan mendukung perencanaan perawatan yang lebih tepat sasaran. Sementara itu, Sari et al (2020) menemukan bahwa penerapan K-Means Clustering dalam analisis hasil pemeriksaan EKG dan tekanan darah di rumah sakit Bandung efektif dalam mengidentifikasi pasien berisiko tinggi yang memerlukan penanganan segera.

Penelitian di Johns Hopkins Hospital (2021) juga menunjukkan bahwa metode ini dapat digunakan untuk memprediksi kebutuhan perawatan intensif bagi pasien yang berisiko mengalami serangan jantung. Melalui pengelompokan data klinis, beban kerja tenaga medis dapat dikurangi, sementara efisiensi layanan kesehatan meningkat. Teknik pengelompokan seperti K-Means sangat efektif dalam mengelompokkan titik-titik data dengan karakteristik yang sama, memberikan wawasan yang dapat ditindaklanjuti untuk pengambilan keputusan (Yusriana Chusna Fadilah, 2025).

Di National Heart Centre Singapore, studi yang dilakukan pada tahun 2020 menemukan bahwa penerapan K-Means Clustering untuk mengelompokkan pasien penyakit jantung yang memerlukan rawat inap menghasilkan sistem perawatan yang lebih efisien. (Chong et al, 2020) melaporkan bahwa metode ini digunakan untuk mengelompokkan pasien berdasarkan riwayat penyakit, hasil tes darah, serta tingkat aktivitas fisik. Hasilnya menunjukkan bahwa pasien dengan faktor risiko serupa dapat menerima perawatan yang lebih terfokus dan sesuai dengan kebutuhan medis mereka.

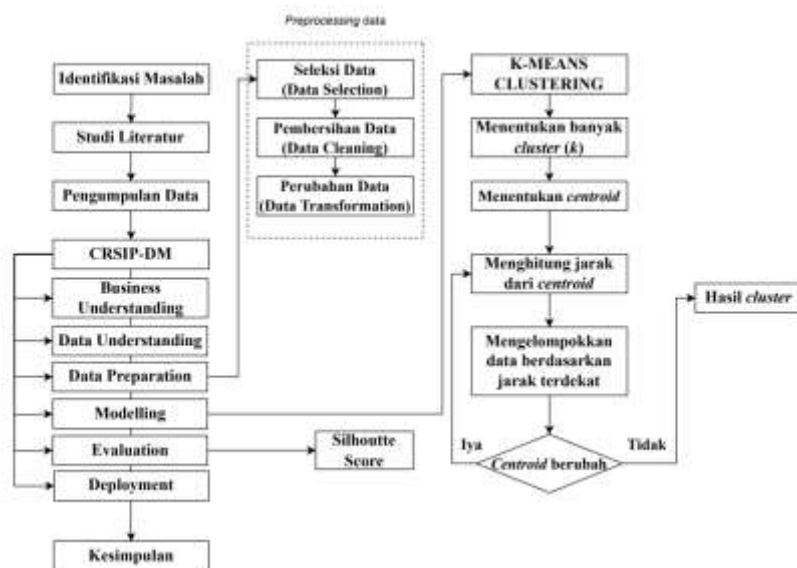
Studi lain oleh Jones et al (2021) di Cleveland Clinic, Amerika Serikat, menunjukkan bahwa penerapan K-Means Clustering dalam pengelompokan pasien berdasarkan riwayat keluarga dan hasil tes genetik dapat membantu dalam memprediksi risiko komplikasi jangka panjang. Dengan adanya klusterisasi ini, dokter

dapat memberikan rekomendasi pencegahan yang lebih spesifik bagi pasien yang memiliki kecenderungan genetik terhadap penyakit jantung.

Di Indonesia, penelitian oleh Yusuf et al (2023) di Universitas Gadjah Mada membahas penerapan K-Means Clustering untuk menganalisis data pasien penyakit jantung dari berbagai puskesmas di Yogyakarta. Studi ini mengolah data yang mencakup hasil tes darah, tekanan darah, kadar gula darah, dan riwayat kesehatan lainnya. Hasil penelitian menunjukkan bahwa metode K-Means Clustering mampu mengelompokkan pasien dengan tingkat risiko tinggi lebih akurat dibandingkan metode konvensional, sehingga memungkinkan dokter puskesmas untuk merujuk pasien dengan cepat ke rumah sakit besar jika diperlukan.

Dengan mempertimbangkan keberhasilan berbagai penelitian sebelumnya, penerapan K-Means Clustering dalam sistem informasi kesehatan dapat menjadi solusi yang efektif dalam mengelola data medis yang semakin kompleks. Teknologi ini tidak hanya membantu dalam mengelompokkan pasien berdasarkan tingkat risiko penyakit jantung, tetapi juga berkontribusi dalam meningkatkan kualitas layanan kesehatan dengan mempercepat proses diagnosa dan pengambilan keputusan medis. Penelitian ini bertujuan untuk menggali lebih dalam potensi K-Means Clustering dalam sistem manajemen informasi kesehatan, khususnya dalam pengelolaan data pasien penyakit jantung di Indonesia.

METODE PENELITIAN



Gambar 1. Alur Penelitian

Gambar di atas merupakan diagram alur proses klusterisasi data menggunakan metode K-Means Clustering dalam konteks CRISP-DM (Cross-Industry Standard Process for Data Mining). Proses dimulai dengan Identifikasi Masalah, kemudian dilakukan Studi Literatur untuk mendukung penelitian. Setelah itu, data dikumpulkan dan masuk dalam tahapan CRISP-DM yang mencakup Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment.

Pada tahap Preprocessing Data, dilakukan Seleksi Data (Data Selection), Pembersihan Data (Data Cleaning), dan Perubahan Data (Data Transformation) untuk memastikan kualitas data sebelum digunakan dalam proses klusterisasi. Selanjutnya, dalam tahapan K-Means Clustering, jumlah kluster (k) ditentukan, lalu dilakukan inialisasi centroid. Algoritma K-Means kemudian menghitung jarak setiap titik data

ke centroid terdekat dan mengelompokkannya sesuai dengan jarak terdekat tersebut.

Proses ini berlangsung secara iteratif, di mana sistem memeriksa apakah centroid berubah setelah pengelompokan data. Jika centroid berubah, maka proses penghitungan jarak dan pengelompokan ulang dilakukan kembali. Jika centroid tidak berubah, maka proses berhenti dan menghasilkan hasil kluster. Evaluasi hasil kluster dilakukan menggunakan Silhouette Score untuk menilai kualitas klusterisasi sebelum diimplementasikan atau digunakan lebih lanjut dalam tahap Deployment. Akhirnya, kesimpulan dari proses ini diambil berdasarkan hasil yang diperoleh.

HASIL DAN PEMBAHASAN

1. Business Understanding

Penelitian ini bertujuan untuk mengelompokkan pasien berdasarkan karakteristik data penyakit jantung dengan menggunakan metode K-Means Clustering menjadi akut dan tidak akut. Dengan memahami pola dari data pasien, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan pengelolaan dan pengambilan keputusan di bidang kesehatan, khususnya terkait penyakit jantung. Selain itu, penelitian ini bertujuan untuk mengeksplorasi apakah metode K-Means Clustering mampu menghasilkan kluster yang jelas dan dapat diinterpretasikan, sehingga mempermudah tenaga medis dalam memahami kondisi pasien. Output yang diharapkan adalah kluster pasien yang dapat dihubungkan dengan tingkat risiko atau kondisi kesehatan mereka

2. Data Understanding

Tahapan ini dilakukan untuk memahami struktur dan karakteristik data yang digunakan dalam penelitian. Dataset yang digunakan diambil dari Kaggle, berisi 1025 baris data terkait penyakit jantung dengan 13 atribut utama, di antaranya: usia (age), jenis kelamin (sex), nyeri dada (cp), tekanan darah (trestbps), kolesterol (chol), denyut jantung (thalach), perubahan depresi segmen (oldpeak), kemiringan segmen (slope), dan jumlah pembuluh darah (ca). Setiap atribut memberikan informasi penting terkait kondisi kesehatan pasien. Sebagai langkah awal, dilakukan visualisasi data untuk memahami distribusi masing-masing atribut. Histogram, scatter plot, dan heatmap korelasi digunakan untuk menemukan pola awal dalam dataset. Beberapa outlier juga terdeteksi, terutama pada atribut chol dan trestbps, yang akan dipertimbangkan selama proses analisis. Dengan memahami karakteristik data ini, penelitian dapat mempersiapkan langkah-langkah berikutnya, seperti normalisasi dan seleksi fitur, agar hasil klusterisasi lebih akurat dan relevan.

3. Data Preparation

Pada tahap ini, dilakukan persiapan data untuk memastikan kualitasnya sesuai untuk analisis menggunakan metode K-Means Clustering. Data dari dataset Kaggle terlebih dahulu diubah menjadi format yang dapat diproses oleh algoritma. Proses ini meliputi normalisasi data menggunakan StandardScaler untuk menyelaraskan skala atribut seperti age, sex, trestbps, chol, dan lainnya. Semua nilai diubah sehingga memiliki rata-rata 0 dan standar deviasi 1 agar algoritma K-Means dapat bekerja secara optimal.

```
[8] # Normalisasi data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data)
scaled_data = np.round(scaled_data, 2)
```

▶ scaled_data

```
array([[ -0.79,  0.5 ,  0.  , -0.36, -0.78,  1.54, -0.66,  1.29,  0.69],
       [-0.63,  0.5 ,  0.  ,  0.51, -0.97,  0.87,  0.92, -1.29, -1.03],
       [ 2.11,  0.5 ,  0.  ,  0.8 , -1.61, -0.68,  0.54, -1.29, -1.03],
       [ 0.66,  0.5 ,  0.  ,  0.98, -0.97,  1.18, -1.42,  1.29, -0.17],
       [ 0.82, -2.  ,  0.  ,  0.4 ,  1.01, -1.67,  0.02,  0.  ,  1.54],
       [ 0.18, -2.  ,  0.  , -1.82,  0.01, -0.84, -0.66,  0.  , -1.03],
       [ 0.18,  0.5 ,  0.  , -1.  ,  1.54,  0.09,  1.9 , -1.29,  1.54],
       [-0.31,  0.5 ,  0.  ,  1.68,  0.9 ,  0.35, -0.82,  0.  , -0.17],
       [-1.75,  0.5 ,  0.  , -0.65,  0.03,  0.3 , -0.82,  1.29, -1.03],
       [-0.47,  0.5 ,  0.  , -0.54,  0.84, -1.15,  1.  ,  0.  ,  0.69]])
```

Gambar 2. Normalisasi Data

Pada kode di gambar, data awal yang terdapat dalam variabel data akan dinormalisasi menggunakan StandardScaler(). Fungsi fit_transform() dari StandardScaler digunakan untuk menghitung nilai rata-rata dan standar deviasi dari setiap fitur pada dataset, kemudian mentransformasikan data tersebut dengan mengurangi rata-rata dan membagi dengan standar deviasi. Hal ini memastikan bahwa setiap fitur memiliki rata-rata 0 dan standar deviasi 1.

Hasil dari normalisasi ditampilkan dalam bentuk array yang disebut scaled_data. Data ini adalah data yang sudah dinormalisasi, di mana setiap nilai telah dikurangi dengan rata-rata dan dibagi dengan standar deviasi dari masing-masing fitur. Nilai-nilai dalam array tersebut lebih konsisten dan berada dalam rentang yang seragam. Nilai-nilai dalam baris tersebut menunjukkan bahwa data untuk setiap fitur telah disesuaikan agar memiliki mean 0 dan standar deviasi 1. Misalnya, nilai pada fitur pertama (-0.79) menunjukkan bahwa data asli untuk fitur tersebut berada lebih rendah dari rata-rata.

```
▶ # Pilih atribut yang akan digunakan untuk clustering
selected_features = ['age', 'sex', 'cp', 'trestbps', 'chol', 'thalach', 'oldpeak', 'slope', 'ca']
data = df[selected_features]
```

[7] data

	age	sex	cp	trestbps	chol	thalach	oldpeak	slope	ca
0	52	1	0	125	212	168	1.0	2	2
1	53	1	0	140	203	155	3.1	0	0
2	70	1	0	145	174	125	2.6	0	0
3	61	1	0	148	203	161	0.0	2	1
4	62	0	0	138	294	106	1.9	1	3
5	58	0	0	100	248	122	1.0	1	0
6	58	1	0	114	318	140	4.4	0	3
7	55	1	0	160	289	145	0.8	1	1
8	46	1	0	120	249	144	0.8	2	0
9	54	1	0	122	286	116	3.2	1	2

Gambar 3. Pemilihan Atribut

Selain itu, atribut-atribut yang relevan dipilih berdasarkan analisis awal, yaitu age, sex, cp, trestbps, chol, thalach, oldpeak, slope, dan ca. Proses ini memastikan bahwa data bersih, terstruktur, dan siap digunakan untuk tahap modeling.

4. Modeling

Tahapan ini menggunakan algoritma K-Means Clustering untuk mengelompokkan data pasien penyakit jantung. Nilai awal centroid ditentukan secara acak, dan proses iterasi dilakukan untuk memperbarui centroid hingga mencapai konvergensi, yaitu ketika perubahan posisi centroid tidak lagi signifikan.

Jumlah kluster ditentukan sebanyak dua ($k = 2$), yang merepresentasikan pengelompokkan risiko pasien menjadi dua kategori utama: risiko rendah dan risiko tinggi. Setelah klusterisasi selesai, hasilnya dianalisis untuk memastikan bahwa setiap kluster memiliki karakteristik yang jelas dan berbeda. Hasil ini diintegrasikan dalam sistem informasi kesehatan untuk membantu pengambilan keputusan.

```
# Inisialisasi model K-Means dengan jumlah kluster = 2
kmeans = KMeans(n_clusters=2, random_state=42)

# Fit model ke data
X_scaled = scaler.fit_transform(data)
kmeans.fit(X_scaled)

# Tambahkan label kluster ke dataset
df['Cluster'] = kmeans.labels_

# Tampilkan centroid
centroids = kmeans.cluster_centers_
print("Centroid Klaster:\n", centroids)

# Tampilkan hasil klusterisasi
print(df[['age', 'Cluster']].head(10))
```

```
Centroid Klaster:
[[ 0.76231829  0.54227086  0.58867366  0.83587015  0.42826551  0.39564581
  0.72743678 -0.22772225  0.54756615]
 [-0.50821219 -0.36151391 -0.39244911 -0.55724677 -0.28551034 -0.26376387
 -0.48495786  0.15181483 -0.3650441 ]]
age Cluster
0  52      1
1  53      0
2  70      1
3  61      0
4  62      0
5  58      1
6  58      1
7  55      1
8  46      1
9  54      0
```

Gambar 4. Klusterisasi

Gambar di atas menunjukkan hasil dari proses klusterisasi menggunakan K-Means, dengan informasi tentang centroid klaster dan klusterisasi data.

Centroid Klaster:

- Bagian pertama dari gambar menampilkan nilai centroid untuk dua kluster yang telah dihasilkan dari algoritma K-Means. Setiap baris dalam centroid mewakili

nilai rata-rata dari fitur-fitur dalam data untuk kluster tersebut.

- Centroid adalah titik pusat dari kluster yang mewakili posisi rata-rata dari data dalam kluster tersebut. Centroid dihitung untuk setiap kluster berdasarkan fitur-fitur yang ada.

Data Klasterisasi:

- Baris 0: age = 52, dan Cluster = 1. Ini berarti data dengan usia 52 dimasukkan ke dalam kluster 1.
- Baris 1: age = 53, dan Cluster = 0. Data dengan usia 53 dimasukkan ke dalam kluster 0.
- Setiap baris mencerminkan data yang telah dikelompokkan berdasarkan kesamaan fitur, seperti yang ditentukan oleh centroid kluster.

5. Evaluation

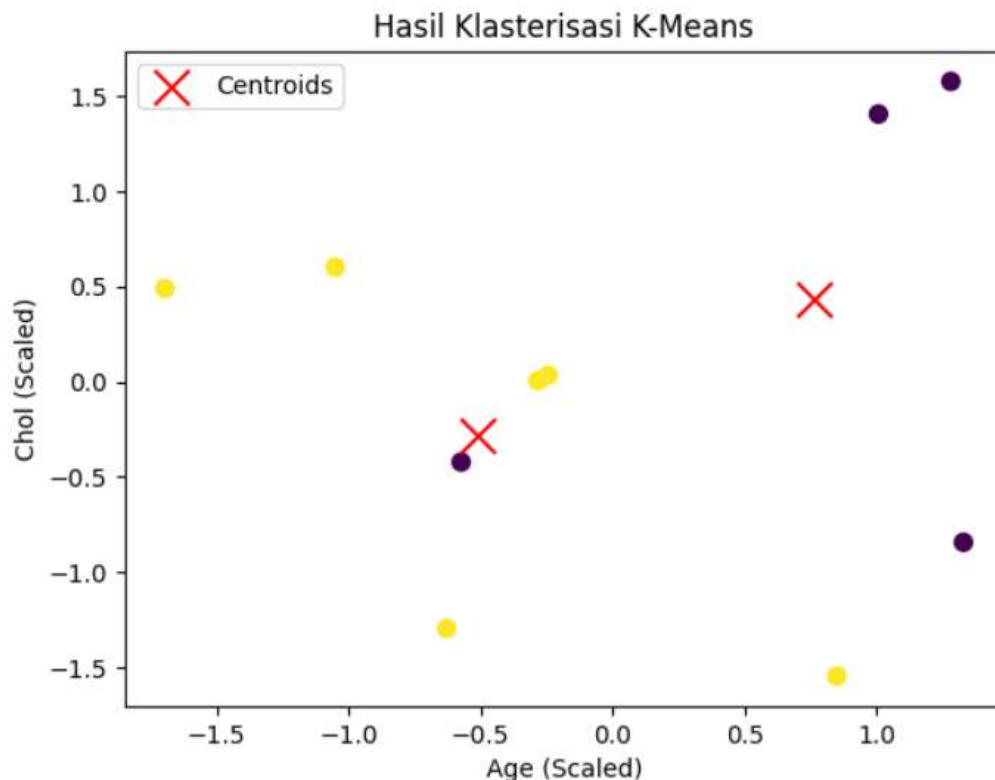
Hasil klasterisasi dievaluasi menggunakan Silhouette Score untuk mengukur seberapa baik kluster yang terbentuk. Nilai Silhouette Score yang tinggi menunjukkan bahwa data dalam satu kluster memiliki kesamaan yang tinggi, sementara data antar kluster berbeda secara signifikan. Selain itu, dilakukan analisis visual menggunakan scatter plot untuk memverifikasi distribusi data dalam kluster. Evaluasi ini bertujuan memastikan bahwa metode K-Means menghasilkan kluster yang bermakna dan dapat diinterpretasikan oleh pihak medis

```
# Evaluasi menggunakan Silhouette Score
silhouette_avg = silhouette_score(X_scaled, kmeans.labels_)
print(f"Silhouette Score: {silhouette_avg:.2f}")

# Visualisasi hasil klasterisasi
import matplotlib.pyplot as plt

# Plot hasil clustering menggunakan dua dimensi (contoh: age dan chol)
plt.scatter(X_scaled[:, 0], X_scaled[:, 4], c=kmeans.labels_, cmap='viridis', s=50)
plt.scatter(centroids[:, 0], centroids[:, 4], c='red', marker='x', s=200, label='Centroids')
plt.title('Hasil Klasterisasi K-Means')
plt.xlabel('Age (Scaled)')
plt.ylabel('Chol (Scaled)')
plt.legend()
plt.show()
```

Silhouette Score: 0.11



Gambar 5. Silhoutte Score

Gambar 4.4 adalah hasil visualisasi klasterisasi K-Means menggunakan dua fitur yang telah diskalakan, yaitu Age (Usia) dan Chol (Kolesterol). Berikut penjelasan elemen-elemen yang ada dalam gambar tersebut:

Tampilan Grafik:

- Sumbu X: Mewakili fitur Age (Usia) yang telah diskalakan. Data ini menunjukkan usia yang telah dinormalisasi.
- Sumbu Y: Mewakili fitur Chol (Kolesterol) yang telah diskalakan. Data ini menunjukkan tingkat kolesterol yang juga telah dinormalisasi.
- Titik-titik yang terlihat di grafik adalah data yang telah dikelompokkan (setelah dilakukan klasterisasi) berdasarkan kedekatannya dengan centroid klaster.

Titik Data:

- Titik berwarna kuning menunjukkan data yang termasuk ke dalam klaster 0.
- Titik berwarna ungu menunjukkan data yang termasuk ke dalam klaster 1.
- Titik-titik ini menunjukkan hasil klasterisasi, di mana setiap titik mewakili data individu yang telah dikelompokkan berdasarkan kedekatannya dengan centroid.

Centroid:

- Titik merah menunjukkan posisi centroid untuk masing-masing klaster.
- Centroid adalah titik pusat dari setiap klaster yang dihitung berdasarkan rata-rata fitur-fitur dari data dalam klaster tersebut. Ada dua centroid yang ditunjukkan di sini: satu untuk klaster 0 (berwarna kuning) dan satu lagi untuk klaster 1 (berwarna ungu).

Silhouette Score:

- Di bagian atas kiri gambar, terdapat Silhouette Score sebesar 0.11.

- Silhouette Score adalah metrik untuk menilai kualitas kluster yang dihasilkan oleh K-Means. Nilai ini menunjukkan seberapa baik setiap data dikelompokkan dengan data lain dalam kluster yang sama dibandingkan dengan data di kluster lain. Nilai ini berkisar antara -1 hingga 1, dengan nilai mendekati 1 menunjukkan kluster yang lebih baik dan nilai mendekati -1 menunjukkan kluster yang buruk.
- Dalam hal ini, nilai 0.11 menunjukkan bahwa klusterisasi tidak terlalu baik, dan mungkin perlu perbaikan dalam pemilihan jumlah kluster atau pemilihan fitur.

6. Deployment

Pada tahap Deployment dalam proyek K-Means Clustering, tujuan utamanya adalah membuat hasil clustering dapat diakses dan divisualisasikan dengan cara yang mudah dimengerti dan interaktif. Untuk ini, kita menggunakan Streamlit, sebuah framework Python yang memudahkan pembuatan aplikasi web interaktif dengan cepat, tanpa perlu menulis banyak kode frontend.



Gambar 6. Dashboard

Gambar di atas menunjukkan antarmuka awal aplikasi Streamlit untuk Klusterisasi Data Penyakit Jantung Menggunakan K-Means.

Berikut adalah elemen yang terlihat:

1. Fitur Unggah Dataset: Aplikasi menyediakan fitur untuk mengunggah file dataset dalam format CSV, dengan batas ukuran maksimum hingga 200MB.
2. Tombol Browse Files: Pengguna dapat memilih file dataset dari komputer lokal menggunakan tombol ini atau dengan cara drag and drop.
3. Pemilihan Fitur untuk Klusterisasi: Pengguna dapat memilih minimal 2 fitur dari dataset yang akan digunakan untuk proses klusterisasi.
4. Pemilihan Jumlah Klusterisasi: Aplikasi menyediakan opsi untuk menentukan jumlah kluster yang diinginkan sesuai kebutuhan analisis.
5. Judul Aplikasi: "KLASTERISASI DATA PENYAKIT JANTUNG MENGGUNAKAN K-MEANS" ditampilkan secara jelas sebagai identitas utama aplikasi.
6. Deskripsi Aplikasi: Aplikasi ini dirancang untuk melakukan klusterisasi data penyakit jantung dengan memanfaatkan algoritma K-Means Clustering.
7. Pesan Informasi Awal: "Silakan upload file dataset dalam format CSV untuk memulai." Pesan ini memberikan arahan kepada pengguna untuk mengunggah dataset sebagai langkah awal sebelum melakukan analisis lebih lanjut.

Dataset

Dataset yang telah diunggah:

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca
0	52	1	0	125	212	0	1	168	0	1	2	2
1	53	1	0	140	203	1	0	155	1	3.1	0	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0
3	61	1	0	148	203	0	1	161	0	0	2	1
4	62	0	0	138	294	1	1	106	0	1.9	1	3
5	58	0	0	100	248	0	0	122	0	1	1	0
6	58	1	0	114	318	0	2	140	0	4.4	0	3
7	55	1	0	160	289	0	0	145	1	0.8	1	1
8	46	1	0	120	249	0	0	144	0	0.8	2	0
9	54	1	0	122	286	0	0	116	1	3.2	1	2

Gambar 7. Dataset pada Dashboard

Gambar di atas menunjukkan tampilan dataset yang telah berhasil diunggah menggunakan fitur unggah dataset. Dataset ditampilkan dengan semua atribut yang masih lengkap, tanpa adanya proses seleksi atau transformasi data pada tahap ini.

Hasil Klasterisasi

Dataset setelah klasterisasi dengan 2 klaster:

	age	sex	cp	trestbps	chol	thalach	oldpeak	slope	ca	Cluster
0	52	1	0	125	212	168	1	2	2	1
1	53	1	0	140	203	155	3.1	0	0	0
2	70	1	0	145	174	125	2.6	0	0	0
3	61	1	0	148	203	161	0	2	1	1
4	62	0	0	138	294	106	1.9	1	3	0
5	58	0	0	100	248	122	1	1	0	0
6	58	1	0	114	318	140	4.4	0	3	0
7	55	1	0	160	289	145	0.8	1	1	0
8	46	1	0	120	249	144	0.8	2	0	1
9	54	1	0	122	286	116	3.2	1	2	0

Gambar 8. Hasil Klasterisasi pada Dashboard

Gambar di atas menampilkan dataset yang telah difilter berdasarkan atribut yang dipilih melalui fitur pemilihan atribut untuk klusterisasi. Selain itu, hasil klusterisasi dengan algoritma K-Means juga ditampilkan pada dataset, dengan setiap data dikelompokkan ke dalam kluster yang sesuai.

Statistik Tiap Kluster

Kluster 0:

	age	sex	cp	trestbps	chol	thalach	oldpeak	slope	ca
count	419	419	419	419	419	419	419	419	419
mean	59.0764	0.7446	0.4391	135.6563	253.6874	131.537	1.9274	0.9308	1.1766
std	7.216	0.4366	0.8711	18.6775	58.4059	19.3758	1.1917	0.4456	1.1164
min	40	0	0	100	131	71	0	0	0
25%	55	0	0	122	212	117	1	1	0
50%	59	1	0	132	249	132	1.8	1	1
75%	64	1	0	145	286	145	2.6	1	2
max	77	1	3	200	564	170	6.2	2	4

Kluster 1:

	age	sex	cp	trestbps	chol	thalach	oldpeak	slope	ca
count	606	606	606	606	606	606	606	606	606
mean	51.2244	0.6617	1.2904	128.8152	240.6848	161.2673	0.4797	1.6997	0.462
std	8.8368	0.4735	0.9866	16.0982	45.5977	16.5685	0.7055	0.5164	0.8534
min	29	0	0	94	126	96	0	0	0
25%	44	0	0	120	210	152	0	1	0
50%	52	1	1	130	234	162	0	2	0
75%	58	1	2	138	266	173	0.8	2	1
max	71	1	3	192	417	202	3.8	2	4

Gambar 9. Statistik tiap Kluster pada Dashboard

Gambar di atas menampilkan hasil klusterisasi dengan 2 kluster, sesuai dengan jumlah kluster yang telah ditentukan melalui fitur pemilihan jumlah kluster. Pada tampilan ini, statistik dari masing-masing kluster ditampilkan berdasarkan atribut yang telah dipilih, sehingga pengguna dapat melihat karakteristik utama dari setiap kluster secara lebih mendetail.

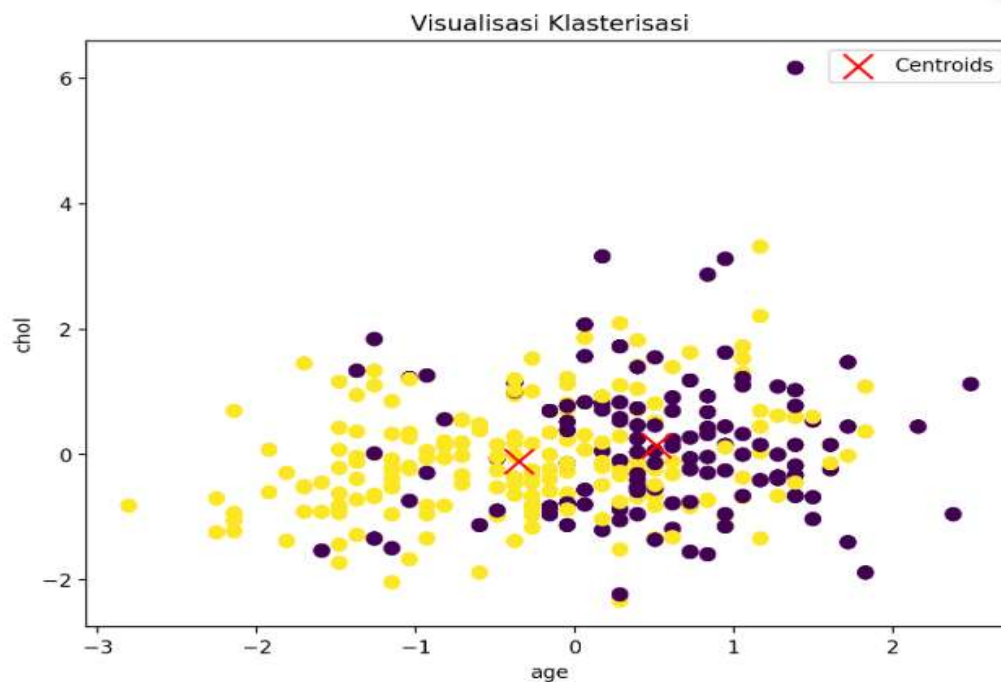
Visualisasi Klasterisasi

Pilih fitur X:

age

Pilih fitur Y:

chol



Gambar 10. Visualisasi Klasterisasi pada Dashboard

Gambar di atas menampilkan visualisasi hasil klasterisasi berdasarkan fitur X dan fitur Y yang dapat dipilih sesuai dengan atribut yang telah ditentukan melalui fitur pemilihan atribut untuk klasterisasi. Visualisasi ini memungkinkan pengguna untuk melihat pembagian klaster secara visual, berdasarkan kombinasi atribut yang dipilih.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Means berhasil mengelompokkan data pasien penyakit jantung ke dalam dua klaster, yaitu klaster pasien dengan kondisi akut dan klaster pasien dengan kondisi tidak akut. Atribut-atribut seperti umur, jenis kelamin, nyeri dada, tekanan darah, kolesterol, denyut jantung, perubahan depresi segmen, kemiringan segmen, dan jumlah pembuluh darah. Evaluasi hasil klasterisasi menunjukkan bahwa metode K-Means memberikan performa yang cukup baik dengan jarak antar centroid yang signifikan dan konsistensi hasil klasterisasi pada data uji. Selanjutnya dapat diintegrasikan ke dalam sistem informasi kesehatan rumah sakit untuk membantu tenaga medis dalam mendukung deteksi dini pasien dengan risiko tinggi untuk dapat memprioritaskan penanganan pasien. Dengan mengimplementasikan metode ini, tenaga medis dapat membuat keputusan yang lebih tepat, mengalokasikan sumber daya dengan lebih efisien, dan meningkatkan kualitas perawatan pasien secara keseluruhan.

DAFTAR PUSTAKA

- [1] A. S. A. A. Yusriana Chusna Fadilah, "APPLYING K-MEANS CLUSTERING FOR GROUPING PAPUA'S DISTRICTS BASED ON POVERTY INDICATORS ANALYSIS," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, pp. 2685-8223, 2025.
- [2] F. Yunita, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *SISTEMASI*, 2020.
- [3] D. A. D. N. A. B. N. S. M. J. F. N. Dhewayani, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, pp. 2088-2270, 2022.
- [4] N. S. G. D. L. Adi Zulkarnaen Saputra, "Klasterisasi Nilai Ujian Sekolah Menggunakan Metode Algoritma KMeans," *Jurnal Janitra Informatika dan Sistem Informasi*, pp. 1-9, 2023.
- [5] A. ., A. T. ., C. B. P.-A. ., D. O. N. ., N. D. F. ., S. R. B. T. Nurul Maulida Surbakti, "Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel," *Jurnal Matematika, Ilmu pengetahuan Alam, Kebumihan dan Angkasa*, pp. 98-107, 2024.
- [6] R. Nazar, "IMPLEMENTASI PEMROGRAMAN PYTHON MENGGUNAKAN GOOGLE COLAB," *Jurnal Informatika dan Komputer*, p. 2089 – 4384, 2024.
- [7] Khoirudin, "Klasifikasi Penyakit Jantung Menggunakan Perbandingan Algoritma Decision Tree," *Jurnal Ilmiah Manajemen Informatika dan Komputer*, pp. 19-25, 2024.
- [8] I. S. Sitompul, "Analisis Pelaksanaan Sistem Informasi Kesehatan Dinas Kesehatan Provinsi Riau," *Journal of Healthcare Technology and Medicine*, 2024.
- [9] D. D. Darmawan, "METODOLOGI PENELITIAN DALAM PENGEMBANGAN APLIKASI STREAMLIT: STUDI KASUS VISUALISASI DATA GEMPA BUMI DI INDONESIA," *researchgate*, 2024.
- [10] W. H. Organization, "Penyakit kardiovaskular (PKV)," 11 Juni 2021. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
- [11] R. & S. D. Hidayat, "Penerapan Metode Data Mining untuk Memprediksi Penyakit Jantung Menggunakan Algoritma K-Means," *Jurnal Teknologi dan Sistem Komputer*, pp. 1-8, 2020.
- [12] R. & S. D. Hidayat, "Penerapan Metode Data Mining untuk Memprediksi Penyakit Jantung Menggunakan Algoritma K-Means," *Jurnal Teknologi dan Sistem Komputer*, pp. 1-8, 2020.
- [13] R. F. & N. A. Sari, "Implementasi Data Mining untuk Analisis Sentimen pada Ulasan Produk Menggunakan Algoritma Naïve Bayes," *Jurnal Sistem Informasi*, pp. 23-30, 2022.
- [14] A. & R. A. Setiawan, "Penerapan Data Mining untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Klasifikasi," *Jurnal Informatika*, pp. 15-22, 2023.
- [15] A. & W. A. Pramudito, "Klasterisasi Data Pelanggan Menggunakan Algoritma K-Means untuk Meningkatkan Strategi Pemasaran," *Jurnal Ilmiah Teknologi Informasi*, pp. 45-52, 2021.