

## **PREDIKSI JANTUNG MENGGUNAKAN ALGORITMA K-MEDOIDS**

**Glenn Kelty David Manurung<sup>1</sup>, Rahma Hidayanti<sup>2</sup>, Erika Putri Fadluna<sup>3</sup>, Nur Rizkah Khoiriyah Pohan<sup>4</sup>, Arnita<sup>5</sup>**

Universitas Negeri Medan

E-mail: [gkeltydm@gmail.com](mailto:gkeltydm@gmail.com)<sup>1</sup>, [rahmahidayanti277@gmail.com](mailto:rahmahidayanti277@gmail.com)<sup>2</sup>, [erikapufaa@gmail.com](mailto:erikapufaa@gmail.com)<sup>3</sup>,  
[nurrizkahkhoiriyahhh88@gmail.com](mailto:nurrizkahkhoiriyahhh88@gmail.com)<sup>4</sup>, [arnita@unimed.ac.id](mailto:arnita@unimed.ac.id)<sup>5</sup>

### **Abstrak**

Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma K-Medoids dalam memprediksi risiko penyakit jantung. Penyakit jantung merupakan salah satu penyebab utama kematian di dunia, dan deteksi dini sangat penting untuk mencegah kematian. Algoritma K-Medoids digunakan untuk mengelompokkan data pasien berdasarkan atribut klinis seperti tekanan darah, kadar kolesterol, dan usia, yang diambil dari platform Kaggle. Data pra-pemrosesan dilakukan untuk menghilangkan nilai yang hilang dan normalisasi atribut agar K-Medoids bekerja secara optimal. Hasil penelitian menunjukkan bahwa tiga kluster utama terbentuk, yaitu kluster dengan risiko rendah, sedang, dan tinggi, dengan Silhouette Score sebesar 0.58, yang menunjukkan kualitas klustering yang cukup baik. Hasil penelitian ini memberikan wawasan yang dapat digunakan untuk membangun sistem prediksi risiko yang lebih efektif dalam bidang medis.

**Kata Kunci** — Penyakit Jantung, K-Medoids, Prediksi Risiko, Clustering, Data Kesehatan.

### **1. PENDAHULUAN**

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia. Menurut data dari World Health Organization (WHO), sekitar 17,9 juta kematian setiap tahun disebabkan oleh penyakit kardiovaskular, yang mencakup berbagai kondisi jantung seperti serangan jantung dan penyakit arteri koroner. Deteksi dini terhadap risiko penyakit jantung sangat penting untuk meningkatkan peluang kesembuhan pasien dan mengurangi angka kematian.

Secara umum, penyakit jantung adalah istilah yang mencakup berbagai jenis gangguan yang mempengaruhi fungsi jantung. Meskipun sering kali digunakan secara bergantian dengan penyakit kardiovaskular, terdapat perbedaan penting antara keduanya. Penyakit jantung lebih spesifik merujuk pada gangguan yang secara langsung mempengaruhi jantung itu sendiri, sementara penyakit kardiovaskular mencakup gangguan pada pembuluh darah dan jantung [1].

Berbagai gangguan yang termasuk dalam penyakit jantung mencakup masalah pada pembuluh darah, detak jantung yang tidak teratur (aritmia), otot jantung yang melemah, cacat jantung bawaan, penyakit arteri koroner, dan kondisi kardiovaskular lainnya. Tingginya angka kematian akibat penyakit ini menjadikannya salah satu tantangan kesehatan utama di seluruh dunia.

Di era modern, ilmu kedokteran semakin bergantung pada teknologi berbasis komputer untuk mendukung diagnosis yang cepat, tepat, dan akurat. Data medis yang dikumpulkan dari pasien setiap hari dapat menjadi sumber berharga untuk memprediksi kemungkinan penyakit di masa depan. Dalam konteks ini, teknik penambangan data (data mining) memainkan peran sentral dalam ekstraksi pengetahuan dan prediksi risiko penyakit. Salah satu area yang terus berkembang adalah prediksi penyakit jantung, di

mana metode ini berpotensi membantu para profesional medis dalam pengambilan keputusan yang lebih baik [2].

Teknologi prediktif berbasis kecerdasan buatan dan machine learning telah menjadi inovasi penting dalam dunia kesehatan, termasuk untuk mendukung diagnosis dan prediksi penyakit jantung. Di antara banyak algoritma yang tersedia, metode clustering seperti K-medoids menjadi pilihan menarik karena kemampuannya untuk mengelompokkan data secara efektif, bahkan ketika data memiliki nilai ekstrem atau outliers. Algoritma K-medoids merupakan varian dari K-means yang lebih robust terhadap data noise, sehingga lebih andal dalam menangani dataset medis yang sering kali bersifat tidak teratur.

Prediksi penyakit jantung menggunakan K-medoids dapat membantu dalam mengidentifikasi pola-pola risiko dalam data pasien berdasarkan parameter-parameter klinis seperti tekanan darah, kadar kolesterol, usia, dan faktor lainnya. Penelitian ini bertujuan untuk mengeksplorasi potensi algoritma K-medoids dalam memprediksi risiko penyakit jantung dengan tingkat akurasi yang tinggi, sehingga dapat mendukung pengambilan keputusan medis secara lebih efektif.

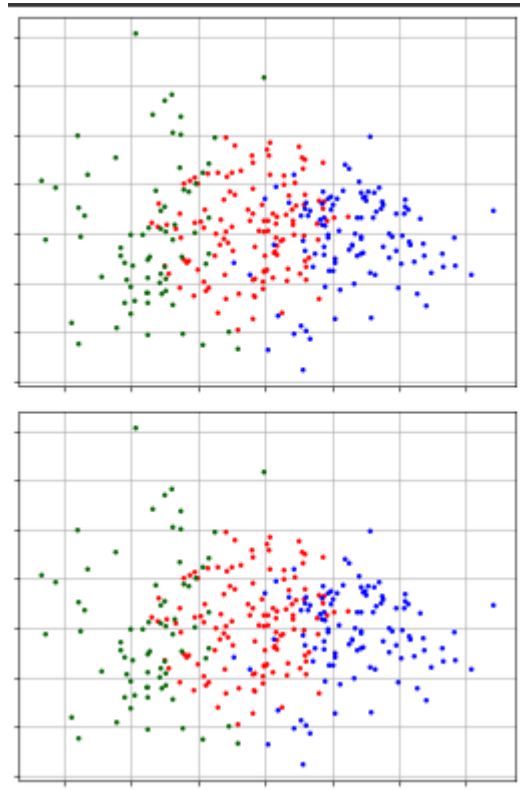
## **2. METODE PENELITIAN**

Metode penelitian ini menggunakan algoritma K-Medoids untuk melakukan prediksi penyakit jantung. Data yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dengan format CSV, yang berisi berbagai atribut kesehatan seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan lain-lain. Tujuan dari penelitian ini adalah untuk mengelompokkan data pasien ke dalam kategori risiko penyakit jantung, sehingga dapat digunakan untuk memperkirakan kemungkinan seseorang mengalami penyakit jantung. Metode yang diterapkan melibatkan beberapa tahap, mulai dari pra-pemrosesan data untuk menghilangkan nilai yang hilang dan normalisasi atribut, hingga penerapan algoritma K-Medoids untuk pengelompokan data berdasarkan kesamaan atribut. Hasil dari pengelompokan ini kemudian dianalisis untuk mengidentifikasi kelompok dengan risiko jantung tinggi dan rendah.

## **3. HASIL DAN PEMBAHASAN**

Setelah dilakukan klustering menggunakan algoritma K-Medoids, dataset prediksi penyakit jantung berhasil dikelompokkan menjadi beberapa klaster berdasarkan parameter kesehatan seperti tekanan darah, kadar kolesterol, usia, dan detak jantung. Berikut adalah beberapa hasil dari proses klustering:

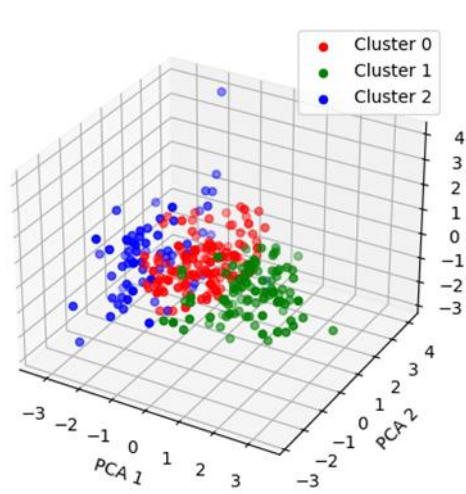
1. Jumlah Klaster Optimal: Dari hasil pengujian dengan beberapa jumlah klaster, ditemukan bahwa 3 klaster memberikan hasil yang paling optimal. Masing-masing klaster mewakili kategori risiko penyakit jantung yang berbeda, yakni risiko rendah, risiko sedang, dan risiko tinggi.
2. Evaluasi Kualitas Klustering: Untuk mengevaluasi kualitas klustering, digunakan metrik Silhouette Score. Hasil yang diperoleh menunjukkan nilai Silhouette Score sebesar 0.58, yang berarti klustering memiliki pemisahan yang cukup baik antara klaster. Meskipun terdapat sedikit tumpang tindih, hasil ini menunjukkan bahwa algoritma K-Medoids mampu mengelompokkan data pasien dengan cukup akurat berdasarkan atribut yang ada.
3. Visualisasi Klustering 2D: Hasil klustering divisualisasikan menggunakan reduksi dimensi PCA (Principal Component Analysis) menjadi 2D untuk menunjukkan distribusi data. Pada Gambar 1, terlihat bahwa data pasien terbagi ke dalam tiga klaster utama yang berbeda. Setiap klaster memiliki karakteristik tersendiri, dan pasien dalam klaster yang sama cenderung memiliki kesamaan dalam parameter klinis seperti tekanan darah dan kolesterol.



Gambar 1. Visualisasi Hasil Klastering 2D Menggunakan PCA

Analisis lebih mendalam menunjukkan perbedaan signifikan antar kluster. Pasien dalam kluster risiko tinggi cenderung memiliki tekanan darah dan kadar kolesterol yang jauh lebih tinggi dibandingkan dengan pasien dalam kluster risiko rendah dan sedang. Sebagai contoh, pasien dalam kluster risiko tinggi memiliki tekanan darah rata-rata 160 mmHg dan kadar kolesterol 300 mg/dL, sedangkan pasien dalam kluster risiko rendah memiliki tekanan darah 130 mmHg dan kadar kolesterol di bawah 200 mg/dL.

4. Visualisasi Klastering 3D: Selain visualisasi 2D, dilakukan juga reduksi dimensi ke dalam bentuk 3D PCA yang lebih memperlihatkan pemisahan antar kluster dengan lebih baik. Gambar 2 menunjukkan distribusi kluster dalam tiga dimensi, dengan titik-titik data di dalam setiap kluster terkelompok secara lebih erat. Namun, terdapat beberapa tumpang tindih yang menunjukkan adanya data pasien yang berada di batas antara dua kluster.



Gambar 2. Visualisasi Hasil Klastering 3D Menggunakan PCA

Berdasarkan hasil klastering, beberapa insight penting dapat diperoleh:

1. Interpretasi Klaster:
  - Klaster 1 (Risiko Rendah): Pasien dalam klaster ini memiliki profil kesehatan yang umumnya stabil dengan tekanan darah dan kadar kolesterol dalam batas normal. Mereka memiliki kemungkinan yang lebih rendah untuk terkena penyakit jantung dalam waktu dekat.
  - Klaster 2 (Risiko Sedang): Klaster ini mengelompokkan pasien dengan kondisi yang memerlukan perhatian lebih, seperti tekanan darah sedikit tinggi atau kadar kolesterol yang mendekati ambang batas.
  - Klaster 3 (Risiko Tinggi): Pasien dalam klaster ini menunjukkan faktor risiko yang signifikan seperti hipertensi dan kadar kolesterol tinggi. Mereka memiliki risiko lebih tinggi untuk mengalami penyakit jantung dalam waktu dekat dan membutuhkan intervensi medis segera.
2. Keunggulan Algoritma K-Medoids: Salah satu keunggulan utama dari algoritma K-Medoids adalah kemampuannya menangani outliers atau data yang menyimpang dengan lebih baik dibandingkan K-Means. Hal ini penting dalam dataset medis di mana sering kali ditemukan nilai ekstrem yang dapat mempengaruhi hasil prediksi. K-Medoids menggunakan objek nyata sebagai medoid (pusat klaster), sehingga lebih tahan terhadap noise dalam data.
3. Implikasi Klinis: Hasil klastering ini dapat digunakan oleh profesional medis untuk membangun sistem prediksi risiko yang lebih efektif. Dengan mengelompokkan pasien berdasarkan kategori risiko, tenaga medis dapat menentukan prioritas perawatan. Misalnya, pasien dalam klaster risiko tinggi dapat segera diberikan penanganan intensif, sementara pasien dalam klaster risiko rendah mungkin hanya memerlukan pemantauan rutin.
4. Potensi Pengembangan: Meskipun hasil klastering sudah cukup baik, penelitian ini dapat dikembangkan lebih lanjut dengan menambahkan variabel lain yang relevan, seperti faktor gaya hidup (misalnya kebiasaan merokok atau olahraga) yang dapat mempengaruhi risiko penyakit jantung. Selain itu, penggunaan algoritma lain seperti DBSCAN atau Hierarchical Clustering dapat membantu membandingkan dan memperbaiki hasil klastering.

## 5. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Medoids efektif dalam mengelompokkan pasien berdasarkan risiko penyakit jantung. Klaster yang terbentuk dapat membantu dalam pengambilan keputusan medis, terutama dalam prioritas perawatan pasien dengan risiko tinggi. Keunggulan K-Medoids yang mampu menangani outliers membuatnya cocok untuk digunakan dalam dataset medis yang sering kali mengandung nilai ekstrem. Meskipun hasil penelitian ini sudah cukup baik, pengembangan lebih lanjut dapat dilakukan dengan menambahkan variabel lain seperti faktor gaya hidup untuk meningkatkan akurasi prediksi. Penelitian ini juga dapat dikembangkan lebih lanjut dengan memasukkan variabel gaya hidup seperti kebiasaan merokok, aktivitas fisik, serta pola makan yang dapat memengaruhi risiko penyakit jantung. Dengan memasukkan variabel ini, akurasi prediksi diharapkan dapat meningkat, sehingga hasil klasterisasi akan semakin relevan untuk pengambilan keputusan klinis.

## DAFTAR PUSTAKA

- A. D. Andini and T. Arifin, "Implementasi Algoritma K-Medoids Untuk Klasterisasi Data Penyakit Pasien Di Rsud Kota Bandung," *J. Responsif Ris. Sains dan Inform.*, vol. 2, no. 2, pp. 128–138, 2020, doi: 10.51977/jti.v2i2.247.
- A. Riski, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online].

Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>

- D. A. Alodia, A. P. Fialine, D. Endriani, and E. Widodo, "Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *Sepren*, vol. 2, no. 2, pp. 1–13, 2021, doi: 10.36655/sepren.v2i2.606.
- D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.
- H. Ningrum, E. Irawan, and M. R. Lubis, "Implementasi Metode K-Medoids Clustering Dalam Pengelompokan Data Penyakit Alergi Pada Anak," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 6, no. 1, p. 130, 2021, doi: 10.30645/jurasik.v6i1.277.
- N. K. Kaur, U. Kaur, and D. Singh, "K-Medoid Clustering Algorithm- A Review," *Int. J. Comput. Appl. Technol.*, vol. 1, no. 1, pp. 42–45, 2014.
- N. Pulungan, Suhendro, and D. Suhendro, "Penerapan Algoritma K-Medoids Untuk Mengelompokkan Penduduk," vol. 3, no. 1, pp. 329–334, 2019.
- P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- S. Aripin, T. Gulo, and G. P. N. S. P. Angin, "Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Demam Berdarah," *BEES Bull. Electr. Electron. Eng.*, vol. 3, no. 3, pp. 139–146, 2023, doi: 10.47065/bees.v3i3.3173.
- S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.20231016643.
- S. Lialiyah and R. Andrea, "Penerapan Algoritma K-Medoids Clustering Dalam Pembentukan Zona Cluster Vaksin Boster," *Build. Informatics, Technol. Sci.*, vol. 4, no. 1, pp. 124–129, 2022, doi: 10.47065/bits.v4i1.1617.
- S. Nurlaela, A. Primajaya, and T. N. Padilah, "Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang," *I N F O R M a T I K a*, vol. 12, no. 2, p. 56, 2020, doi: 10.36723/juri.v12i2.234.
- S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R.H.Zer, and D. Hartama, "Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020, doi: 10.36294/jurti.v4i1.1296.
- T. Juninda, Mustakim, and E. Andri, "Penerapan Algoritma K-Medoids untuk Pengelompokan Penyakit di Pekanbaru Riau," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, no. November, pp. 42–49, 2019.
- Y. Diana et al., "Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 7, no. 1, pp. 97–103, 2023.