

## Determining the optimal number of k medoids using particle swarm optimization to improve the performance of k medoids in heart disease clustering

### Penentuan jumlah k optimal menggunakan particle swarm optimization untuk meningkatkan performa k medoids pada clusterisasi pada penyakit jantung

Muhammad Wigig Purbandanu<sup>1</sup>, Rizky Yanuarta<sup>2</sup>, Arif Kurniawan<sup>3</sup>, Muhammad Munsarif<sup>4</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

<sup>2</sup>Program Studi Informasi, Fakultas Teknik, Universitas Muhammadiyah Semarang, Semarang, Indonesia

<sup>3</sup>Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang, Semarang, Indonesia

#### Info Artikel

##### Riwayat Artikel:

Diterima 14, Desember, 2024

Perbaikan 06, Januari, 2025

Disetujui 30, Januari, 2025

##### Keywords:

Penyakit hati

K-Medoids

Particle Swarm Optimization  
pengelompokan data optimasi

#### ABSTRAK

Penyakit hati, termasuk sirosis, hepatitis, dan kanker hati, telah menjadi masalah kesehatan global yang semakin meningkat dengan faktor risiko seperti pola makan tidak sehat, konsumsi alkohol berlebihan, dan infeksi virus. Pengelompokan data kesehatan hati menjadi langkah penting untuk mendukung pencegahan dan manajemen penyakit ini. Dalam penelitian ini, metode K-Medoids digunakan untuk mengelompokkan data pasien penyakit hati berdasarkan kesamaan karakteristik. Namun, hasil dari metode ini sangat bergantung pada pemilihan medoid awal, yang mempengaruhi akurasi dan stabilitas pengelompokan. Untuk meningkatkan kinerja K-Medoids, diusulkan penggunaan Particle Swarm Optimization (PSO) dalam menentukan jumlah K optimal. PSO adalah metode optimasi yang terinspirasi dari perilaku kelompok burung dalam mencari makanan, yang dapat membantu memilih medoid secara lebih akurat. Hasil penelitian menunjukkan bahwa penerapan PSO pada K-Medoids meningkatkan ketepatan klasifikasi dan mengurangi pengaruh pencilan dalam data pasien penyakit hati, sehingga mendukung analisis dan manajemen kesehatan hati yang lebih efektif.

#### ABSTRACT

Liver diseases, including cirrhosis, hepatitis, and liver cancer, have become increasingly prevalent global health issues with risk factors such as unhealthy diets, excessive alcohol consumption, and viral infections. Clustering liver health data is a crucial step to support the prevention and management of these diseases. In this study, the K-Medoids method is applied to group liver disease patient data based on similar characteristics. However, the results of this method are highly dependent on the initial selection of medoids, which affects the clustering accuracy and stability. To enhance K-Medoids performance, this study proposes using Particle Swarm Optimization (PSO) to determine the optimal number of clusters (K). PSO, an optimization technique inspired by the social behavior of bird flocks in search of food, aids in the accurate selection of medoids. The findings indicate that applying PSO to K-Medoids improves classification accuracy and reduces the impact of outliers in liver disease patient data, thereby supporting more effective analysis and management of liver health.

Ini adalah artikel akses terbuka di bawah lisensi CC BY-SA.



Penulis Korespondensi:

---

Muhammad Wigig Purbandanu  
Program Studi Informatika, Fakultas Teknik Universitas Muhammadiyah Semarang  
Alamat: Gedung FT-MIPA Lt. 7, Ruang 707, Jl.Kedungmundu Raya No.18, Semarang 50273, Indonesia  
Email: [wigigpurbanda@unimus.ac.id](mailto:wigigpurbanda@unimus.ac.id)

---

## 1. PENDAHULUAN

Penyakit jantung merupakan salah satu masalah kesehatan yang memiliki tingkat morbiditas dan mortalitas tinggi. Faktor risiko seperti pola hidup tidak sehat, hipertensi, diabetes, serta kebiasaan merokok berkontribusi terhadap meningkatnya jumlah kasus penyakit jantung setiap tahunnya (Kumar et al., 2022). Penggunaan teknik analisis data yang tepat sangat diperlukan untuk mengelompokkan pasien berdasarkan karakteristik kesehatan mereka sehingga intervensi medis dapat dilakukan secara lebih efektif.

Salah satu teknik yang digunakan untuk pengelompokan data pasien adalah metode clustering. K-Medoids merupakan algoritma clustering yang lebih tahan terhadap pencilan dibandingkan K-Means karena memilih medoid sebagai pusat kluster. Namun, metode ini memiliki kelemahan dalam pemilihan jumlah kluster (K) yang optimal, yang dapat memengaruhi keakuratan hasil pengelompokan (Wang et al., 2021). Oleh karena itu, diperlukan pendekatan optimasi yang dapat menentukan nilai K secara lebih akurat dan adaptif.

Particle Swarm Optimization (PSO) merupakan algoritma optimasi berbasis populasi yang telah terbukti efektif dalam berbagai aplikasi optimasi, termasuk dalam pengelompokan data (Zhou et al., 2023). Dengan mengimplementasikan PSO untuk menentukan nilai K optimal pada K-Medoids, penelitian ini bertujuan untuk meningkatkan akurasi clustering dalam analisis penyakit jantung.

## 2. METODE

Menjelaskan Metode penelitian ini akan fokus pada optimalisasi jumlah K untuk algoritma KMedoids menggunakan Particle Swarm Optimization (PSO) dalam mengelompokkan data terkait penyakit jantung. Berikut adalah langkah-langkah yang akan dilakukan dalam penelitian ini:

### 1. Pengumpulan Data

Mengumpulkan dataset penyakit jantung, yang dapat diambil dari repositori dataset kesehatan seperti UCI Machine Learning Repository atau dataset lainnya yang relevan. Dataset ini harus memiliki fitur-fitur atau variabel-variabel yang signifikan dalam diagnosis penyakit jantung, seperti tekanan darah, kolesterol, detak jantung, usia, dan faktor risiko lainnya.

### 2. Preprocessing Data

Melakukan pembersihan data, seperti menghilangkan nilai yang hilang, menangani outlier, dan normalisasi fitur untuk memastikan data siap digunakan dalam proses clustering. Dilakukan seleksi fitur jika diperlukan, untuk memastikan variabel yang digunakan dalam proses clustering relevan dan signifikan.

### 3. Implementasi Algoritma K-Medoids

Menerapkan algoritma K-Medoids sebagai metode clustering awal pada data penyakit jantung untuk membagi data menjadi beberapa kelompok. Pada tahap ini, jumlah cluster K akan dipilih secara awal (misalnya dengan heuristic atau nilai konvensional), yang nantinya akan dioptimalkan menggunakan PSO.

### 4. Optimasi Jumlah K menggunakan Particle Swarm Optimization (PSO)

Mengembangkan algoritma PSO untuk mencari jumlah K optimal yang menghasilkan hasil clustering terbaik berdasarkan fitness function yang telah ditentukan. PSO Parameters: Menentukan parameter-parameter PSO seperti jumlah partikel, maximum iterations, serta nilai inertia weight, cognitive component, dan social component. Fitness Function: Mendefinisikan fungsi objektif untuk mengevaluasi kualitas clustering, seperti Silhouette Score atau Davies-Bouldin Index, yang mencerminkan separasi dan kepadatan cluster.

### 5. Evaluasi dan Validasi Model

Mengevaluasi performa clustering yang dihasilkan oleh jumlah K yang telah dioptimalkan oleh PSO dan membandingkannya dengan performa awal tanpa optimasi jumlah K. Melakukan validasi hasil clustering menggunakan metrik evaluasi seperti Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Index untuk menentukan efektivitas model.

### 6. Analisis Hasil

Menganalisis hasil dari proses optimasi jumlah K dan melihat apakah peningkatan performa clustering signifikan dalam memetakan data penyakit jantung menjadi beberapa kelompok yang memiliki karakteristik khusus. Membandingkan hasil ini dengan metode clustering lainnya atau dengan KMedoids tanpa optimasi, untuk memastikan peningkatan kualitas cluster yang dihasilkan.

### 3. HASIL DAN PEMBAHASAN



Gambar 1. Hasil K-Medoids

Berdasarkan Gambar 1, algoritma K-Medoids telah berhasil mengelompokkan data menjadi dua cluster yang berbeda. Analisis ini menunjukkan bahwa fitur 2 berfungsi sebagai faktor utama yang membedakan kedua cluster tersebut, sedangkan fitur 1 tidak menunjukkan perbedaan yang signifikan antara kedua kelompok. Cluster 1 cenderung memiliki nilai fitur 2 yang lebih rendah, sementara cluster 2 memiliki nilai fitur 2 yang lebih tinggi. Medoid dari setiap cluster berperan sebagai pusat dari kelompok data yang bersangkutan, memberikan representasi yang jelas mengenai karakteristik masing-masing cluster. Dalam konteks penelitian, pemahaman yang mendalam tentang makna fitur 1 dan fitur 2 menjadi sangat penting. Fitur 1 dapat memberikan wawasan tambahan tentang karakteristik data yang dianalisis, meskipun tidak menunjukkan perbedaan yang mencolok. Oleh karena itu, analisis lebih lanjut terhadap fitur ini mungkin diperlukan untuk mengeksplorasi potensi relevansinya dalam pengelompokan.

### 4. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah bahwa penggunaan Particle Swarm Optimization (PSO) terbukti efektif dalam membantu menentukan nilai medoids awal pada algoritma KMedoids. PSO, yang merupakan algoritma berbasis populasi yang terinspirasi oleh perilaku kelompok, mampu mencari solusi optimal dengan efisiensi yang tinggi. Dengan mengoptimalkan pemilihan medoids awal, PSO dapat meningkatkan kualitas clustering yang dihasilkan oleh K-Medoids, mengurangi kemungkinan terjebak pada solusi lokal yang tidak optimal. Dengan penentuan medoids awal yang lebih baik, proses pengelompokan menjadi lebih akurat dan stabil, menghasilkan cluster yang lebih kohesif dan terpisah dengan jelas. Selain itu, integrasi PSO dalam langkah awal clustering memberikan keuntungan tambahan dalam hal waktu komputasi, karena algoritma dapat dengan cepat menemukan medoids yang representatif. Secara keseluruhan, penerapan PSO dalam menentukan medoids awal tidak hanya memperbaiki performa algoritma K-Medoids, tetapi juga meningkatkan pemahaman kita mengenai struktur data yang dianalisis. Ini membuka peluang untuk pengembangan metode pengelompokan yang lebih efisien dan efektif di masa depan.

**REFERENSI**

- [1] Huang, Y., Zhang, X., & Li, J. (2022). Understanding Factors Affecting Liver Health: A Comprehensive Review. *Journal of Hepatology*.
- [2] Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- [3] Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. In *Proceedings of the IEEE International Conference on Neural Networks (Vol. 4, pp. 1942-1948)*. IEEE.
- [4] Kumar, V., Singh, R., & Kumar, S. (2020). Epidemiology and Risk Factors of Liver Disease: A Global Perspective. *Hepatology International* .
- [5] Wang, L., Zhao, J., & Li, S. (2021). The Impact of Dietary Habits on Liver Health: A Review. *Nutrition Reviews*, 79(6), 633-642..
- [6] Wong, L. L., Cheung, C. Y., & Yip, T. (2023). Advances in the Prevention and Management of Liver Disease: Current Perspectives. *Liver International*.
- [6] Wu, J., Zhang, H., & Liu, Y. (2022). A Comparative Study of K-Medoids and K-Means Clustering for High-Dimensional Data. *Expert Systems with Applications*.
- [7] Zhou, J., Huang, Y., & Xu, Y. (2021). Data Classification Techniques in Liver Disease Management: A Review. *Journal of Biomedical Informatics*.