

JURNAL STATISTIKA DAN SAINS DATA
Volume 4 Edisi 1, Oktober, 2026, 227-232
https://journal.student.uny.ac.id/index.php/jssd
https://dx.doi.org/10.21831/jssd.v4i1.25209

PENERAPAN *K-MEDOIDS* DAN *AGGLOMETARIVE HIERARCHICAL CLUSTERING* UNTUK MENGELOMPOKKAN KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN JUMLAH TENAGA KESEHATAN

Najla Nashirah*, Universitas Negeri Yogyakarta
Indira Ihnu Brilliant, Universitas Negeri Yogyakarta
*e-mail: najlanashirah.2020@student.uny.ac.id , indiraihnubrilliant@uny.ac.id

Abstrak. Distribusi tenaga kesehatan yang tidak merata antar wilayah di Provinsi Jawa Tengah menjadi tantangan dalam mewujudkan pelayanan kesehatan yang adil. Penelitian ini bertujuan untuk mengelompokkan 35 kabupaten/kota di Jawa Tengah berdasarkan jumlah tenaga kesehatan tahun 2023 menggunakan dua metode *clustering*, yaitu *K-medoids* dan *Agglomerative hierarchical clustering* (AHC). Analisis dimulai dengan tahap preprocessing data, uji asumsi kecukupan dan *multikolinearitas*, reduksi dimensi menggunakan Principal Component Analysis (PCA), serta penentuan jumlah *cluster* optimal menggunakan metode *Silhouette*. Hasil analisis menunjukkan bahwa metode *K-medoids* dan AHC berhasil membentuk tiga *cluster* wilayah. Kota Semarang secara konsisten membentuk *cluster* tersendiri karena memiliki jumlah tenaga kesehatan tertinggi, sedangkan wilayah lainnya tergabung dalam *cluster* yang mencerminkan tingkat ketersediaan tenaga kesehatan menengah hingga rendah. Temuan ini memberikan gambaran distribusi tenaga kesehatan yang dapat digunakan sebagai dasar perumusan kebijakan pemerataan SDM kesehatan.

***Kata kunci: Clustering*, *K-medoids*, *Agglomerative hierarchical clustering*, PCA, Tenaga Kesehatan**.

Abstract. *The uneven distribution of health workers across regions in Central Java Province presents a challenge in achieving equitable health services. This study aims to group 35 regencies/cities in Central Java based on the number of health workers in 2023 using two clustering methods: K-medoids and Agglomerative hierarchical clustering (AHC). The analysis included data preprocessing, assumption testing for adequacy and multicollinearity, dimensionality reduction through Principal Component Analysis (PCA), and optimal cluster determination using the Silhouette method. The results showed that both K-medoids and AHC successfully formed three clusters. Semarang City consistently formed a separate cluster due to its significantly high number of health workers, while other regions were grouped based on moderate to low availability. These findings offer a clear picture of health workforce distribution and can support the formulation of more targeted and equitable health resource policies.*

Keywords: Clustering, K-medoids, Agglomerative hierarchical clustering, PCA, Health Workers.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

PENDAHULUAN

Tenaga kesehatan yang tidak merata antar wilayah menjadi permasalahan penting dalam sistem pelayanan kesehatan di Indonesia. Jawa Tengah sebagai salah satu provinsi berpenduduk besar yang memerlukan ketersediaan tenaga kesehatan yang memadai dan merata untuk menjamin akses pelayanan yang setara. Data diperoleh dari Badan Pusat Statistika (BPS) tahun 2023 menunjukkan adanya ketimpangan jumlah tenaga kesehatan seperti Kota Semarang yang sangat tinggi, sementara beberapa kabupaten lainnya jauh tertinggal. Ketidakseimbangan ini dapat berdampak pada kualitas layanan kesehatan, terkhususnya di daerah terpencil. Dalam upaya menanggapi isu tersebut, pendekatan berbasis analisis data menjadi strategi penting, salah satunya melalui teknik *clustering* untuk mengelompokkan wilayah berdasarkan karakteristik jumlah tenaga kesehatan.

Penelitian sebelumnya telah membuktikan efektivitas metode *clustering* dalam mengelompokkan wilayah berdasarkan karakteristik tertentu, termasuk sektor kesehatan. Misalnya, metode *K-medoids* digunakan untuk menganalisis indikator kemiskinan (Alfiah *et al*., 2020) dan fasilitas rumah sakit (Romadon & Fitrianah, 2019), sedangkan *Agglomerative hierarchical clustering* (AHC) telah diaplikasikan untuk pemetaan wilayah banjir (Pratikto & Damastuti, 2021) maupun isu kesehatan nasional seperti stunting (Suraya & Wijayanto, 2022). Penelitian-penelitian tersebut menekankan keunggulan visualisasi struktur data dan ketahanan terhadap *outlier* yang dimiliki masing-masing metode yang relevan untuk mengatasi permasalahan jumlah tenaga kesehatan di Jawa Tengah.

Penelitian ini bertujuan menerapkan metode *K-medoids* dan AHC untuk mengeksplorasi pola jumlah tenaga kesehatan pada 35 kabupaten/kota di Provinsi Jawa Tengah. Data yang digunakan data tahun 2023 dari Badan Pusat Statistika (BPS), penelitian ini menerapkan strategi analisis data yang mencakup *preprocessing*, reduksi *multikolinearitas* dengan PCA, serta pemilihan jumlah *cluster* optimal menggunakan *Silhouette* dan menggunakan *median linkage*. Hasil *clustering* diharapkan dapat menggambarkan kelompok wilayah dengan karakteristik jumlah tenaga kesehatan yang serupa, sehingga memberikan dasar yang kuat bagi pemerintah daerah dalam menyusun kebijakan pemerataan tenaga kesehatan secara lebih strategis.

METODE

Jumlah tenaga kesehatan antar wilayah di Jawa Tengah, serta mencapai tujuan pengelompokkan kabupaten/kota berdasarkan jumlah tenaga kesehatan, penelitian ini menawarkan pendekatan berbasis data dengan menerapkan dua metode *clustering* yang memiliki ketahanan terhadap data dan mampu menggambarkan struktur hubungan antarwilayah, yaitu *K-medoids* dan *Agglomerative hierarchical clustering* (AHC).

Metode *K-medoids* digunakan karena keunggulannya dalam mengelompokkan data dengan lebih robust terhadap outlier dibandingkan metode lainnya seperti K-Means. Teknik metode *K-medoids* dimulai dengan pemilihan *medoid* secara acak, diikuti dengan proses perhitungan jarak antar data, pembaruan *medoid* berdasarkan total simpangan minimum, hingga selesai saat komposisi *cluster* stabil. Metode ini memastikan bahwa setiap *cluster* merupakan representasi aktual dari data asli.

Metode *Agglomerative hierarchical clustering* (AHC) dengan teknik *median linkage* dipilih karena mampu membentuk struktur hierarki pengelompokan yang mudah diinterpretasikan melalui dendogram. *Median linkage* dinilai lebih tahan terhadap pengaruh nilai ekstrem, karena menggunakan nilai tengah (median) dari suatu *cluster*. Proses AHC

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

dimulai dengan setiap objek sebagai *cluster* tunggal, kemudian menggambungkan dua *cluster* terdekat berdasarkan ukuran jarak *median linkage* hingga membentuk satu struktur hierarkis.

Sebelum proses *clustering* dilakukan, dilakukan tahap preprocessing data yaitu, pemeriksaan dan pembersihan data dari missing values dan outlier, uji kecukupan data dengan KMO dan Bartlett untuk memastikan kelayakan analisis, deteksi dan penanganan *multikolinearitas* antar variabel menggunakan *Principal Component Analysis* (PCA) sebagai metode reduksi dimensi, dan penentuan jumlah *cluster* optimal menggunakan metode *Silhouette*, yang mengevaluasi kesesuaian objek terhadap *cluster*-nya dibandingkan *cluster* terdekat lainnya.

Seluruh prosedur dilakukan dengan bantuan perangkat lunak RStudio, yang digunakan untuk menghitung nilai PCA, membentuk *cluster*, hingga memvisualisasikan hasilnya dalam bentuk plot dan dendrogram. Proses ini memungkinkan terbentuknya kelompok wilayah dengan kemiripan karakteristik jumlah tenaga kesehatan, yang secara langsung mendukung pembuatan kebijakan berbasis data oleh pemangku kepentingan di bidang kesehatan.

Dengan penerapan dua pendekatan ini, diharapkan solusi yang ditawarkan dapat membantu pemerintah dalam menyusun strategi distribusi tenaga kesehatan secara efisien, adil, dan tepat sasaran berdasarkan profil wilayah masing-masing.

HASIL DAN PEMBAHASAN

Dalam penelitian ini digunakan dua metode *clustering*, yaitu *K-medoids* dan *Agglomerative hierarchical clustering* (AHC) untuk mengelompokkan 35 kabupaten/kota di Jawa Barat berdasarkan data jumlah tenaga kesehatan tahun 2023 yang mencakup 10 jenis tenaga yaitu, tenaga medis, kebidanan, keperawatan, kefarmasian, kesehatan masyarakat, kesehatan lingkungan, tenaga gizi, keterapian fisik, teknik biomedika, dan teknik medis.

1. Tahap Pemrosesan Data dan Uji Asumsi

Pada penelitian ini, data tidak memiliki missing values namun terdapat outlier pada seluruh variable. Oleh karena itu, dipilih metode *clustering* yang robust terhadap outlier yaitu,*K-medoids* dan *Agglomerative hierarchical clustering* (AHC) *Median linkage.*

Uji KMO menunjukkan nilai 0,89 (>0,5) yang menandakan data layak untuk dilakukan análisis faktor. Uji barlett menghasilkan *p-value* 2,22e-16 (<0,05) artinya antar variable berkorelasi. Pada penelitian ini, data terdeteksi adanya *multikolinearitas* karena nilai VIF beberapa variable melebihi 10 (misal: VIF Teknik Biomedika = 102,81). Oleh karena itu, dilakukan reduksi dimensi dengan PCA. Nilai eigen menunjukkan dua komponen utama yang dipilih.

2. Hasil *K-medoids Clustering*

Penentuan jumlah *cluster* optimal menggunakan método *Silhouette* yang menunjukkan nilai optimal pada k = 3. Dari hasil PCA, setiap kabupaten/kota memiliki skor dua klomponen utama (PC1 dan PC2). Berdasarkan hasil algoritma PAM, diperoleh tiga *cluster* yaitu :

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

Cluster	Kabupaten	Jumlah Angka
1	Cilacap, Banyumas, Kebumen, Klaten, Sukoharjo, Grobogan, Pati, Tegal, Brebes, Kota Surakarta	10
2	Purbalingga, Banjarnegara, Purworejo, Wonosobo, Magelang, Boyolali, Wonogiri, Karanganyar, Sragen, Blora, Rembang, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Kota Magelang, Kota Salatiga, Kota Pekalongan, dan Kota Tegal	24
3	Kota Semarang	1

Karakteristik tiap *cluster* berbeda-beda. *Cluster* 1 memiliki jumlah tenaga medis, keperawatan, dan kefarmasian yang relatif tinggi. Kondisi ini dipengaruhi oleh faktor kepadatan penduduk dan keberadaan institusi pendidikan atau industri. Contohnya, Kota Surakarta memiliki banyak rumah sakit besar yang menjadi rujukan untuk wilayah daerah sekitarnya. Selain itu, institusi pendidikan seperti Universitas Sebelas Maret (UNS) dan sekolah tinggi kesehatan lainnya juga menghasilkan tenaga kesehatan sehingga ketersediaan tenaga kesehatan di wilayah ini lebih baik. Daerah industri seperti Cilacap dan Pati juga memerlukan fasilitas dan tenaga kesehatan lebih banyak untuk mendukung pekerja industri, sehingga karakteristik *cluster* ini cenderung menunjukkan kesiapan pelayanan kesehatan yang lebih baik.

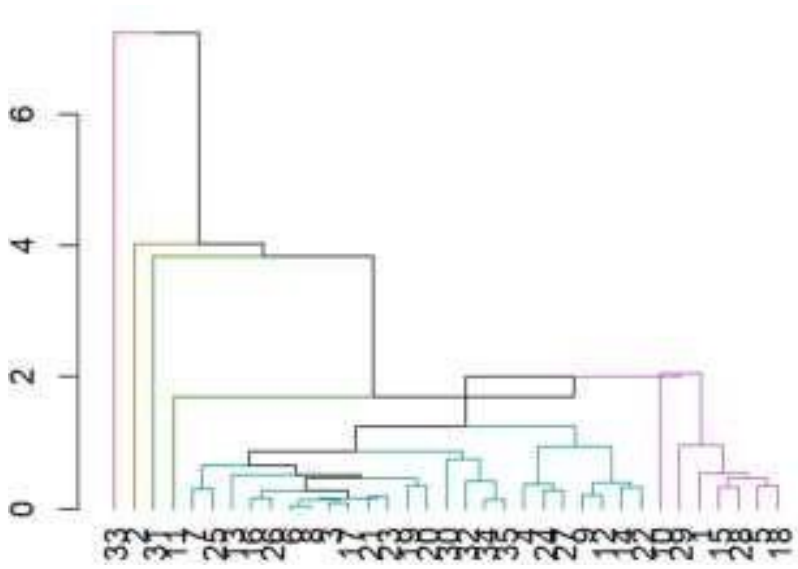
Cluster 2 memiliki jumlah tenaga kesehatan di tingkat menengah atau sedang, baik tenaga medis, bidan, perawat, maupun farmasi. Kabupaten/kota di *cluster* ini merupakan daerah industri skala menengah seperti Pekalongan (industri batik) dan Kudus (industri rokok), sehingga membutuhkan tenaga kesehatan dalam jumlah yang memadai. Sementara itu, wilayah seperti Wonosobo dan Temanggung yang berada di daerah pegunungan cenderung memiliki distribusi tenaga kesehatan yang lebih merata di puskesmas dan rumah sakit daerah. Secara umum, *cluster* ini menggambarkan wilayah dengan pelayanan kesehatan yang cukup baik, meskipun tidak sebanyak di kota-kota besar.

Cluster 3 memiliki tenaga kesehatan yang sangat banyak dan lengkap, mulai dari dokter, perawat, farmasi, hingga tenaga gizi. Hal ini karena Semarang adalah ibu kota provinsi Jawa Tengah dan menjadi pusat rujukan medis dengan rumah sakit besar seperti RSUP Dr. Kariadi. Selain itu, adanya kampus kesehatan besar seperti Universitas Diponegoro (UNDIP) dan Poltekkes Kemenkes untuk meningkatkan jumlah tenaga kesehatan di kota ini. Dengan jumlah penduduk yang besar serta fasilitas kesehatan yang lengkap maka Semarang berada di *cluster* tersendiri karena jauh lebih unggul dibandingkan daerah lain.

3. Hasil *Agglomerative hierarchical clustering* (AHC)

Metode AHC menggunakan dua komponen utama dari PCA untuk menghindari *multikolinearitas*. Proses *clustering* menggunakan *Median linkage*, yang lebih tahan terhadap outlier. Hasil *clustering* divisualisasikan dalam dendogram sebagai berikut:

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649



Gambar 1. Hasil *clusterin*

Dari hasil dendogram diatas menghasilkan tiga *cluster* yaitu, *cluster* pertama terdiri dari 33 kabupaten/kota antara lain, Cilacap, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan, Kota Tegal. Karakteristik pada *cluster* ini sebagian besar kabupaten/kota di Jawa Tengah yang memiliki jumlah tenaga kesehatan relatif rendah hingga sedang. Fasilitas pelayanan kesehatan seperti rumah sakit besar atau pusat tidak banyak tersedia di kabupaten/kota ini dan sebagian masyarakat mengandalkan puskesmas sebagai pusat layanan seperti Blora, Grobogan dan Kebumen. Beberapa kota dalam *cluster* ini seperti Kota Tegal atau Kota Salatiga meskipun berstatus kota tetapi secara jumlah tenaga kesehatan masih jauh di bawah Kota Semarang.

Cluster kedua hanya terdiri dari 1 kabupaten, antara lain Banyumas. Kabupaten Bayumas yang memiliki jumlah tenaga kesehatan relatif tinggi sehingga menjadi pusat layanan bagi wilayah barat Jawa Tengah di karenakan memiliki fasilitas rumah sakit yang cukup lengkap termasuk RSUD Margono Soekarjo yang menjadi rumah sakit rujukan. Selain itu, ketersediaan tenaga gizi, keterapian fisik, dan teknik biomedika di Banyumas juga cukup baik. Keberadaan institusi pendidikan tinggi kesehatan di kabupaten ini turut mendukung ketersediaan SDM kesehatan secara berkelanjutan.

Cluster ketiga terdiri dari 1 kota, antara lain, Kota Semarang. *Cluster* ini memiliki jumlah tenaga kesehatan tertinggi di Jawa Tengah pada jumlah tenaga kesehatan. Kota Semarang merupakan ibu kota provinsi dan pusat pendidikan, pelatihan, dan layanan kesehatan rujukan tingkat provinsi. Pusat rumah sakit besar di Kota Semarang seperti RSUP Dr. Kariadi, rumah sakit pendidikan, sert memiliki banyak institusi pendidikan di bidang kesehatan seperti UNDIP dan Poltekkes Semarang. Hal ini menjadikan jumlah tenaga kesehatan di Kota Semarang tertinggi pada *cluster* ini.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649

SIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa metode *K-medoids* dan *Agglomerative hierarchical clustering* (AHC) berhasil mengelompokkan 35 kabupaten/kota di Provinsi Jawa Tengah tahun 2023 berdasarkan jumlah tenaga kesehatan. Kedua metode menunjukkan bahwa Kota Semarang secara konsisten membentuk *cluster* tersendiri, mengindikasikan bahwa wilayah ini memiliki jumlah tenaga kesehatan yang sangat tinggi dibandingkan dengan daerah lain. Metode *K-medoids* menghasilkan tiga *cluster* dengan *medoid* pada Pati, Purworejo, dan Kota Semarang. Sementara itu, metode AHC menggunakan pendekatan *median linkage* yang juga membentuk tiga *cluster*, dengan hasil yang hampir serupa, namun lebih menekankan struktur hierarki hubungan antarwilayah. Proses analisis pada kedua metode menunjukkan bahwa penggunaan Principal Component Analysis (PCA) efektif dalam mengatasi permasalahan *multikolinearitas* pada data, sedangkan pemilihan *K-medoids* dan *median linkage* terbukti tepat dalam menghadapi data yang mengandung outlier. Secara keseluruhan, hasil *clustering* memberikan gambaran yang jelas mengenai ketimpangan distribusi tenaga kesehatan di Jawa Tengah, sehingga dapat menjadi dasar yang kuat bagi pemerintah dalam menyusun strategi pemerataan tenaga kesehatan secara lebih merata dan tepat sasaran.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta, khususnya Program Studi Statistika, yang telah memberikan fasilitas dan dukungan selama proses pelaksanaan kegiatan penelitian ini. Ucapan terima kasih juga disampaikan kepada Ibu Indira Ihnu Brilliant, M.Stat., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan yang sangat berharga dalam setiap tahapan penelitian dan penyusunan artikel ini. Tak lupa, penulis menyampaikan apresiasi kepada Badan Pusat Statistik (BPS) Provinsi Jawa Tengah atas tersedianya data yang menjadi dasar utama dalam analisis *clustering* ini. Dukungan dari pihak-pihak tersebut sangat berperan dalam kelancaran kegiatan penelitian dan penyusunan artikel ini hingga selesai.

DAFTAR PUSTAKA

Alfiah, N., Ismail, M. I., & Gunardi, I. (2020). Analisis pengelompokan wilayah berdasarkan indikator kemiskinan menggunakan metode K-medoids. Jurnal Statistika, 10(2), 134–142.

Romadon, A. & Fitrianah, R. (2019). K-Means dan K-medoids untuk pengelompokan rumah sakit berdasarkan fasilitas layanan. Jurnal Teknik Informatika dan Sistem Informasi, 5(1), 25–34.

Pratikto, H. & Damastuti, E. (2021). Pemetaan wilayah rawan banjir menggunakan metode Agglomerative hierarchical clustering. Jurnal Geografi Indonesia, 36(3), 215–222.

Suraya, N. & Wijayanto, H. (2022). Penerapan AHC untuk pengelompokan kabupaten/kota berdasarkan angka stunting. Jurnal Kesehatan Masyarakat Digital, 4(2), 45–53.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis (5th ed.). John Wiley & Sons.

Copyright © 2026, Jurnal Statistika dan Sains Data, Online ISSN: 3025-9649