Jurnal Kajian dan Terapan Matematika
Volume 11, Edisi 1, Bulan Maret 2025, 14-26
http://journal.student.uny.ac.id/ojs/index.php/jktm
https://doi.org/10.21831/jktm.v11i1.23088

PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS CLUSTERING* UNTUK PEMETAAN KECAMATAN DI KABUPATEN BANTUL BERDASARKAN STATUS GIZI BALITA

COMPARISON OF K-MEANS AND K-MEDOIDS ALGORITHMS FOR CLUSTERING NUTRITIONAL STATUS OF TODDLERS IN BANTUL DISTRICT

Guntur Andreansyah, Program Studi Matematika Universitas Negeri Yogyakarta, Indonesia
Bambang Sumarno Hadi Marwoto*, Program Studi Matematika Universitas Negeri Yogyakarta, Indonesia
*e-mail korespondensi: bambang@uny.ac.id

Abstrak
Berdasarkan survei status gizi, keadaan gizi balita di Kabupaten Bantul memiliki masalah gizi ganda (*double burden*). Selain melalui survei status gizi, pemahaman tentang keadaan status gizi balita dapat dilakukan melalui pemetaan. Penelitian bertujuan untuk memetakan status gizi balita di Kabupaten Bantul berdasarkan hasil perbandingan terbaik antara algoritma *K-Means* dan *K-Medoids*. Data status gizi yang digunakan mencakup 15 variabel penyebab tidak langsung yang mempengaruhi status gizi dari kecamatan di Kabupaten Bantul tahun 2022 dan 2023. *Clustering* menggunakan algoritma *K-Means* dan *K-Medoids* yang hasilnya divalidasi dengan metode *Davies-Bouldin Index* dan *Silhouette Coefficient*. Hasil analisis menunjukan bahwa algoritma *K-Means* dengan tiga *cluster* merupakan algoritma teroptimal dengan nilai validasi *Davies-Bouldin Index* 1,3829301131 dan *Silhouette Coefficient* 0,2088311419. Tiga *cluster* yang terbentuk selanjutnya dikategorikan sebagai: tingkat status gizi balita sedang (14 data), buruk (8 data), dan baik (12 data). Berdasarkan hasil tersebut, diharapkan dapat menjadi dasar untuk evaluasi dan peningkatan status gizi balita di Kabupaten Bantul.
Kata kunci: Kabupaten Bantul, *K-Means*, *K-Medoids*, Pemetaan, Status Gizi Balita.

Abstract
Based on nutritional status surveys, the nutritional status of toddlers in Bantul Regency faces a double burden of malnutrition. In addition to surveys, understanding the nutritional status of toddlers can also be done through mapping. The research aims to map the nutritional status of toddlers in Bantul Regency by comparing the performance of the K-Means and K-Medoids algorithms. The nutritional condition of toddlers data used includes 15 indirect variables that affect the nutritional status across sub-districts in Bantul Regency for the years 2022 and 2023. Clustering was performed using the K-Means and K-Medoids algorithms, with the results validated by the Davies-Bouldin Index and Silhouette Coefficient methods. The analysis showed that the K-Means algorithm with three clusters was the optimal algorithm, with a Davies-Bouldin Index value of 1,3829301131 *and a Silhouette Coefficient of* 0,2088311419. *The three clusters formed were categorized as follows: moderate nutritional status (14 data), poor nutritional status (8 data), and good nutritional status (12 data). Based on these results, it is hoped that this can serve as a basis for the evaluation and improvement of toddler nutritional status in Bantul Regency.*
***Keywords**: Bantul Regency, K-Means, K-Medoids, Mapping, Nutritional Status of Toddlers.*

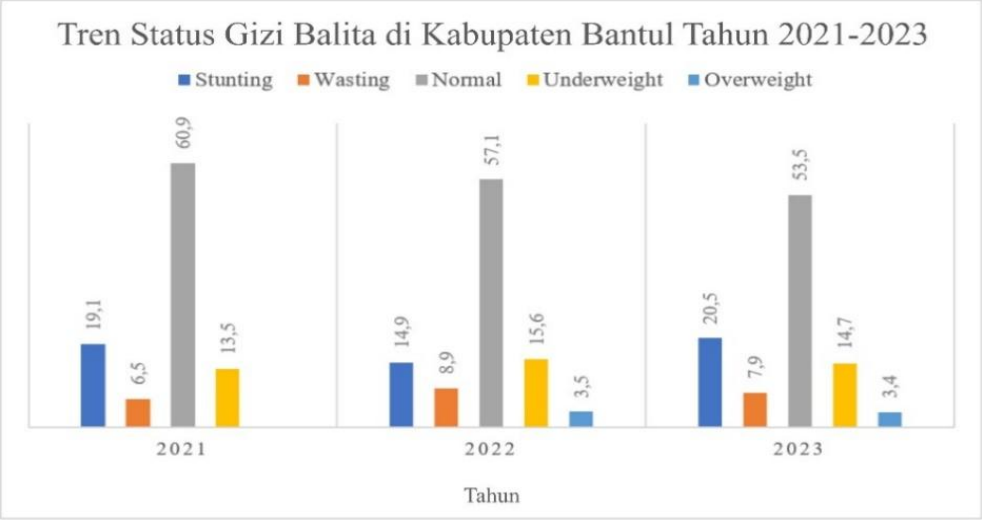
PENDAHULUAN

Pembangunan nasional merupakan upaya yang dirancang secara sistematis dan berkelanjutan dan mencakup seluruh aspek kehidupan bermasyarakat, berbangsa, dan bernegara untuk meningkatkan kualitas hidup manusia baik lahir maupun batin (Prasetyo, 2013). Keberhasilannya tentu sangat bergantung pada kualitas sumber daya manusia (SDM) yang tangguh secara fisik, mental, memiliki kesehatan prima, serta menguasai dan mampu berkembang terhadap kemajuan IPTEK. Demi tersedianya SDM yang berkualitas, diperlukan pembangunan kesehatan. Pembangunan kesehatan harus diprioritaskan dan diperhatikan secara serius demi mencapai keberhasilan pembangunan nasional.

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Prioritas pembangunan kesehatan dapat dimulai dengan fokus pada proses tumbuh kembang anak sejak dini (balita) melalui pemenuhan kebutuhan dasar anak demi terbentuknya individu yang sehat, cerdas, dan produktif (Fauziah & Apsari, 2010). Untuk menilai kondisi tubuh seseorang akibat dari makanan yang dikonsumsi dan cara tubuh memanfaatkan zat gizi tersebut, dapat diketahui dari status gizinya (Supardi *et al*., 2023). Antropometri, penilaian berdasarkan dimensi dan komposisi tubuh menggunakan alat ukur (seperti timbangan dan meteran) merupakan metode yang kerap digunakan untuk menilai status gizi balita. Dari hasil pengukuran dengan metode antropometri dapat diketahui apakah balita memiliki status gizi baik, kekurangan gizi, atau kelebihan gizi.

Demi memahami keadaan status gizi secara menyeluruh, Indonesia rutin melaksanakan survei status gizi melalui program nasional seperti Riset Kesehatan Dasar (Riskesdas), Survei Status Gizi Indonesia (SSGI), dan Survei Kesehatan Indonesia (SKI). Berdasarkan hasil survei tersebut, diketahui bahwa balita Indonesia mengalami masalah gizi ganda (*double burden*), dimana balita mengalami obesitas di satu sisi, sedangkan di sisi lainnya balita mengalami anemia, kurus, *stunting*, hingga gizi buruk (Supardi *et al*., 2023). Menurut UNICEF (1998) status gizi disebabkan oleh tiga faktor utama, yaitu penyebab langsung (asupan makanan dan penyakit infeksi), penyebab tidak langsung (akses terhadap makanan, perawatan anak dan ibu hamil, serta keadaan sanitasi dan layanan kesehatan), dan penyebab utama (adalah krisis ekonomi langsung yang berimbas pada pokok masalah status gizi, terutama yang berkaitan dengan kemiskinan, pengetahuan, keterampilan, dan perilaku).



Sumber: Kementerian Kesehatan Indonesia

Gambar 1. Tren Status Gizi Balita di Kabupaten Bantul Tahun 2021 Hingga 2023

Berdasarkan Gambar 1, persentase balita dengan status gizi normal di Kabupaten Bantul mengalami penurunan. Pada tahun 2021, persentase balita dengan status gizi normal mencapai 60,9%, kemudian turun sebesar 4,8% pada tahun 2022 menjadi 56,1%, dan terus turun sebesar hingga 53,3% pada tahun 2023. Selain itu, *stunting* masih mendominasi masalah gizi balita, sehingga pemerintah Kabupaten Bantul menyikapi keadaan tersebut dengan menetapkan lokasi fokus prioritas dalam upaya penurunan *stunting* melalui Keputusan Bupati Bantul Nomor 127 Tahun 2022. Salah satu cara untuk mendukung penetapan wilayah tersebut adalah dengan melakukan pemetaan wilayah melalui teknik *clustering*.

Seiring dengan pentingnya pemetaan status gizi balita, teknik *clustering* telah digunakan dalam berbagai penelitian untuk mengelompokkan data berdasarkan karakteristik yang serupa. Penelitian yang dilakukan oleh Efendi *et al*. pada tahun 2019 menjelaskan bahwa menggunakan

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

algoritma *K-Means* dapat membagi derajat kesehatan 20 puskesmas di Kota Bengkulu menjadi 3 kelompok. Selanjutnya pada tahun 2021, Julyantari *et al*. menggunakan algoritma *K-Means*, didapatkan tiga *cluster* status gizi, yaitu 30,43% sampel berada di kelompok gizi buruk, 47,38% sampel berada di kelompok gizi normal, dan 21,78% sampel berada di kelompok gizi lebih. Penelitian serupa juga dilakukan oleh Dona & Rifqi (2022), yaitu melakukan *cluster*ing untuk penentuan status gizi baik dan gizi buruk pada balita di Kabupaten Rokan Hulu menggunakan algoritma *K-Means*. Penelitian terbaru oleh Hidayat *et al*. pada tahun 2024, bertujuan membandingkan algoritma *K-Means* dan *K-Medoids* dalam pengelompokan data balita dengan gizi buruk di Provinsi Jawa Barat pada tahun 2019 hingga 2022 yang melibatkan 324 sampel data dengan 9 atribut.

Algoritma *K-Means* dan *K-Medoids* dinilai efektif dalam melakukan *clustering* status gizi balita, sebagaimana ditunjukan dalam sejumlah penelitian terdahulu. Hal tersebut menjadi dasar penulis untuk melakukan penelitian tentang pemetaan status gizi balita di Kabupaten Bantul dengan membandingkan kedua algoritma tersebut. Penelitian ini bertujuan untuk memetakan hasil *clustering* status gizi balita setiap kecamatan yang ada di Kabupaten Bantul sebagai evaluasi kebijakan yang telah terlaksana dan informasi penyusunan kebijakan di masa depan. Proses analisis data dan *clustering* akan dilakukan menggunakan bahasa pemrograman Python, sementara pemetaan dilakukan dengan *software* Quantum GIS. Penulis berharap dari penelitian ini dapat memberikan manfaat yang berarti bagi seluruh lapisan masyarakat, khususnya di Kabupaten Bantul pada bidang matematika dan kesehatan.

METODE

Deskripsi Data

Dalam penelitian ini, digunakan data sekunder dari Dinas Kesehatan Kabupaten Bantul yang mencakup 17 kecamatan, yaitu Bambanglipuro, Banguntapan, Bantul, Dlingo, Imogiri, Jetis, Kasihan, Kretek, Pajangan, Pandak, Piyungan, Pleret, Pundong, Sanden, Sewon, Sedayu, dan Srandakan. Terdapat total 34 data dengan 15 variabel dalam satuan persentase yang akan dianalisis, yaitu balita dengan status gizi *stunting* ($X1$), balita dengan status gizi *wasting* ($X2$), balita dengan status gizi baik (normal) ($X3$), balita dengan status gizi *underweight* ($X4$), balita dengan status gizi *overweight* ($X5$), pengelolaan air minum dan makanan ($X6$), imunisasi dasar lengkap ($X7$), pemberian Air Susu Ibu (ASI) Eksklusif ($X8$), cakupan pelayanan kesehatan bayi ($X9$), cakupan kunjungan neonatal (KN) lengkap ($X10$), layanan ANC ($X11$), persalinan di Fasyankes ($X12$), cakupan kunjungan nifas lengkap ($X13$), akses rumah sehat ($X14$), dan sanitasi yang layak ($X15$). Adapun tahapan penelitian yang akan dilakukan adalah sebagai berikut.

Uji Multikolinearitas

Tujuan uji multikolinearitas adalah mengidentifikasi korelasi antar variabel. Salah satu metode yang dapat digunakan untuk mengetahui korelasi yang terjadi adalah koefisien korelasi (korelasi *pearson*) dengan persamaan sebagai berikut (Miftahuddin *et al*., 2021).

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

dengan r_{xy} adalah korelasi *pearson*, x adalah variabel terkait, y adalah variabel bebas, dan n adalah banyak objek. Adapun interpretasi nilai korelasi pearson secara lengkap dapat ditemukan pada Tabel 1 (Mundir, 2012, p.114).

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Tabel 1. Interpretasi Nilai r Korelasi Pearson

Nilai r	Interpretasi
$0{,}00 \leq \lvert r \rvert \leq 0{,}30$ $(0{,}00 \leq \lvert r \rvert \leq -0{,}30)$	Korelasi Tidak Berarti
$0{,}30 \leq \lvert r \rvert \leq 0{,}50$ $(-0{,}30 \leq \lvert r \rvert \leq -0{,}50)$	Korelasi Rendah
$0{,}50 \leq \lvert r \rvert \leq 0{,}70$ $(-0{,}50 \leq \lvert r \rvert \leq -0{,}70)$	Korelasi Sedang
$0{,}70 \leq \lvert r \rvert \leq 0{,}90$ $(-0{,}70 \leq \lvert r \rvert \leq -0{,}90)$	Korelasi Tinggi
$0{,}90 \leq \lvert r \rvert \leq 1{,}00$ $(-0{,}90 \leq \lvert r \rvert \leq -1{,}00)$	Korelasi Sangat tinggi

Ketika didapatkan interpretasi nilai r dengan korelasi tinggi atau sangat tinggi, maka perlu dilakukan pemilihan variabel lebih lanjut. Pemilihan variabel tersebut dapat dilakukan dengan menentukan nilai *Variance Inflation Factor* (VIF), dimana ketika nilai VIF melebihi 10 menandakan bahwa adanya gejala multikolinearitas. Adapun menurut Anniza Andyani *et al*. (2024), persamaan untuk menentukan VIF adalah sebagai berikut dengan R_k^2 adalah koefisien korelasi antara variabel dengan variabel lainnya.

$$VIF = \frac{1}{1 - R_k^2} \tag{2}$$

K-Means Clustering

Salah satu metode *non-hierarchical clustering* berbasis partisi adalah *K-Means*, di mana jumlah *cluster* (k) harus ditentukan sebelum melakukan *clustering* (Park *et al*., 2020). Pada iterasi pertama, *centroid* setiap *cluster* ditentukan secara acak, kemudian untuk iterasi selanjutnya *centroid* setiap *cluster* ditentukan dari hasil perhitungan sebelumnya. Adapun proses perhitungan *K-Means* adalah sebagai berikut (Tanjung *et al*., 2021):

1. Penentuan jumlah *cluster* k dan *centroid* secara acak.
2. Menentukan jarak terpendek setiap data ke setiap *centroid* menggunakan persamaan *Euclidean Distance*.

$$d(x_i, c_j) = \|c_j - x_i\| = \sqrt{\sum_{k=1}^{p} (x_{ik} - c_{jk})^2} \tag{3}$$

Persamaan *Euclidean Distance* $(d(x_i, c_j))$ mengukur jarak antara data ke-i pada variabel data ke-k (x_{ik}) dengan *centroid cluster* ke-j pada variabel data ke-k (c_{jk}) untuk variabel data sejumlah p.

3. Melakukan pengelompokan setiap data ke *cluster* yang memiliki jarak terpendek dari *centroid*, hasil dari perhitungan *Euclidean Distance*.
4. Memperbarui *centroid* dengan perhitungan rata-rata setiap *cluster* setelah data dikelompokan dalam setiap *cluster*.
5. Mengulangi langkah 2–4 hingga anggota dari setiap *cluster* sama pada iterasi ke-i dan pada iterasi ke-$(i - 1)$.

K-Medoids Clustering

Partitioning Around Medoids (PAM) atau *K-Medoids* adalah metode pengelompokan yang berbasis partisi dan mirip dengan *K-Means.* Perbedaan utamanya adalah pada *K-Medoids* titik pusat *cluster* didapatkan dengan memilih titik data aktual yang meminimalkan jarak (*medoids*) (Sadeghi *et al*., 2024). Adapun proses perhitungan *K-Medoids* adalah sebagai berikut (Ningrum *et al*., 2021) (Han *et al*., 2011) (Bahri *et al*., 2023).

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

1. Penentuan jumlah *cluster* dan *medoid* awal sebanyak k secara acak.
2. Menentukan jarak terdekat setiap data ke setiap *medoid* awal menggunakan persamaan *Euclidean Distance* pada persamaan (1).
3. Memilih *medoid* baru yang dipilih secara acak dengan jumlah tetap seperti sebelumnya, yaitu sebanyak *k cluster*.
4. Menentukan jarak setiap data di *cluster* ke setiap *medoids* baru menggunakan persamaan *Euclidean Distance* pada persamaan (1).
5. Menghitung total simpangan (S), yaitu dengan menghitung total jarak terdekat data ke *medoid* baru − total jarak terdekat data ke *medoid* awal. Dengan keputusan ketika hasil perhitungan diperoleh $S < 0$, maka tukar *medoid* awal dengan data untuk membentuk sekumpulan k baru sebagai *medoid*
6. Mengulangi langkah 2–5, langkah ini diulangi hingga tidak terjadi perubahan pada *medoids,* sehingga didapatkan k *cluster* beserta anggotanya.

Davies-Bouldin Index

Pada penelitian ini digunakan dua metode validasi, pertama adalah *Davies-Bouldin Index* (DBI), yaitu metode validasi yang melakukan pengukuran berdasar nilai kohesi dan separasi (Sholeh & Aeni, 2023). Nilai DBI berkisar antara 0 sampai 1 dan jumlah *cluster* terbaik adalah *cluster* yang memiliki nilai DBI paling kecil. Adapun berikut ini adalah langkah-langkah terurut untuk menentukan nilai *Davies-Bouldin Index* (DBI) (Butsianto & Saepudin, 2020).

1. Menghitung *Sum of Square Within-Cluster* (SSW) untuk mengetahui tingkat kemiripan setiap data yang berada pada *cluster* yang sama. Nilai SSW yang semakin kecil menunjukan data dalam *cluster* tersebut semakin mirip.
$$SSW_i = \frac{1}{m_i}\sum_{j=1}^{m_i} d(x_j, c_i) \quad (4)$$
Nilai SSW dihitung berdasarkan rata-rata jarak antara data x_j dengan *centroid* c_i $(d(x_j, c_i))$ pada semua data dalam *cluster* ke-i (m_i).
2. Menghitung *Sum of Square Between-Cluster* (SSB) untuk mengukur variabilitas atau sebesar apa perbedaan antar *cluster*. Nilai SSB yang semakin besar, semakin bagus karena semakin besar perbedaan antar *cluster*. Adapun SSB dapat dihitung dengan menentukan jarak antara *cluster* ke-i (c_i) dan ke-j (c_i), yaitu menghitung $(d(c_i, c_j))$ di *cluster* yang berbeda sebagai berikut.
$$SSB_{ij} = d(c_i, c_j) \quad (5)$$
3. Menghitung rasio, dimaksudkan untuk mengetahui nilai perbandingan antara variabilitas dalam *cluster* (SSW) dan antar *cluster* (SSB). Adapun persamaannya adalah sebagai berikut.
$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (6)$$
4. Menghitung DBI untuk mengetahui kualitas hasil dari *clustering* yang telah didapatkan. Jika didapatkan nilai DBI yang semakin kecil dan *non-negatif*, maka *cluster* tersebut semakin baik. Adapun persamaannya adalah sebagai berikut.
$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}(R_{i,j}) \quad (7)$$

Nilai k menunjukan jumlah *cluster* yang terbentuk dalam pengelompokan. Dalam konteks ini, nilai k merujuk pada jumlah *cluster* yang dipilih sebelum melakukan proses *cluster*ing menggunakan algoritma tertentu.

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Silhouette Coefficient

Metode validasi kedua yang digunakan dalam penelitian ini adalah *Silhouette Coefficient*. *Silhouette Coefficient* menghitung nilai rerata, yaitu selisih antara nilai *separation* dan nilai *compactness* kemudian dibagi dengan nilai maksimum antara keduanya, dari setiap titik dalam *cluster* (Harani *et al*., 2020). Adapun berikut ini adalah langkah-langkah terurut untuk menentukan nilai *Silhouette Coefficient* (Nahdliyah *et al*., 2019).

1. Menentukan *compactness* (a_i) untuk mengukur kedekatan setiap data yang berada pada *cluster* yang sama. Jika didapatkan nilai a_i yang semakin kecil maka semakin bagus karena setiap data dalam *cluster* tersebut semakin mirip. Adapun persamaannya adalah sebagai berikut.

$$a_i = \frac{1}{N_k - 1} \sum_{x_j \in N_k, x_i \neq x_j} d(x_i, x_j) \tag{8}$$

Nilai (a_i) dihitung berdasarkan rata-rata jarak antara data x_i dengan data x_j pada semua data dalam *cluster* ke-k (N_k).

2. Menentukan *Separation* (b_i) untuk mengukur kedekatan data dengan data di *cluster* lain dengan menentukan nilai minimum dari rata-rata jarak antara data dengan data di *cluster* lain. Jika didapatkan nilai b_i yang besar maka semakin bagus karena semakin besar pula perbedaan antar *cluster* Adapun persamaannya adalah sebagai berikut.

$$b_i = \min_{k' \neq k} \left(\frac{1}{N_{k'}} \sum_{x_j \in N_{k'}} d(x_i, c_j) \right) \tag{9}$$

Nilai (b_i) dihitung berdasarkan rata-rata jarak antara data x_i dengan data x_j pada semua data dalam *cluster* yang berbeda ($N_{k'}$).

3. *Silhouette Index* (S_i), yaitu menghitung perbandingan dari nilai *Compactness* (a_i) dan *Separation* (b_i). Adapun persamaannya adalah sebagai berikut.

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{10}$$

4. *Silhouette Coefficient* (SC) bertujuan bertujuan untuk menghitung rata-rata dari *Silhouette Index* (S_i) untuk semua data. Jika didapatkan nilai SC mendekati 1 maka *cluster* semakin baik. Jika didapatkan nilai SC mendekati -1 maka *cluster* semakin buruk. Adapun persamaannya adalah sebagai berikut.

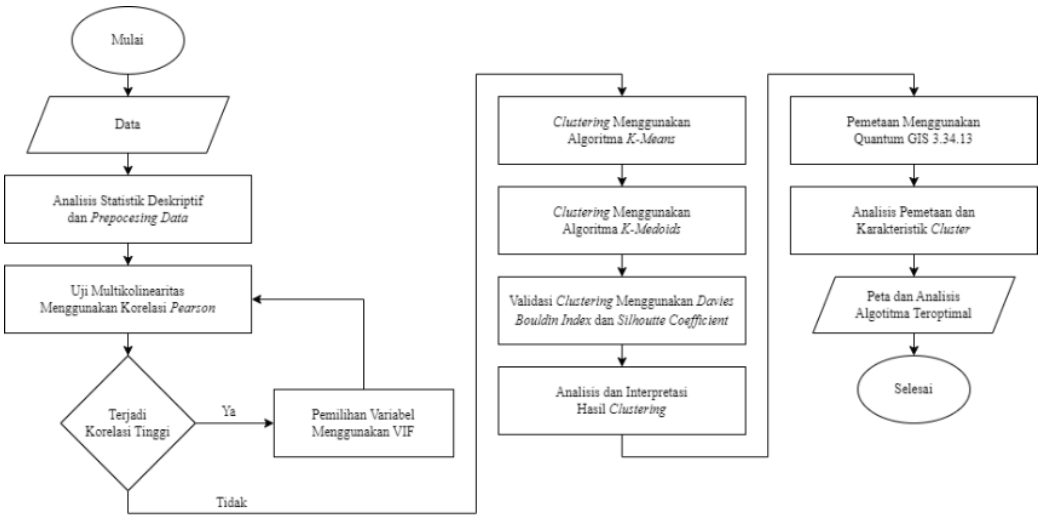
$$SC = \frac{1}{k} \sum_{i=1}^{k} s_i \tag{11}$$

Sama halnya dengan DBI, nilai k menunjukan jumlah *cluster* yang terbentuk dalam pengelompokan. Dalam konteks ini, nilai k merujuk pada jumlah *cluster* yang dipilih sebelum melakukan proses *cluster*ing menggunakan algoritma tertentu.

Langkah Penelitian

Gambar 2 adalah langkah-langkah penelitian yang akan dilakukan dalam penelitian ini. Secara garis besar, penelitian ini akan melakukan perbandingan hasil *clustering* antara algoritma *K-Means* dan *K-Medoids* menggunakan dua metode validasi, yaitu *Davies Bouldin Index* dan *Silhouette Coefficient*. Hasil *clustering* terbaik selanjutnya menjadi dasaran untuk membuat peta persebaran status gizi balita di Kabupaten Bantul pada tahun 2022 dan 2023.

Copyright © 2025, Jurnal Kajian dan Terapan Matematika



Gambar 2. *Flowchart* Penelitian

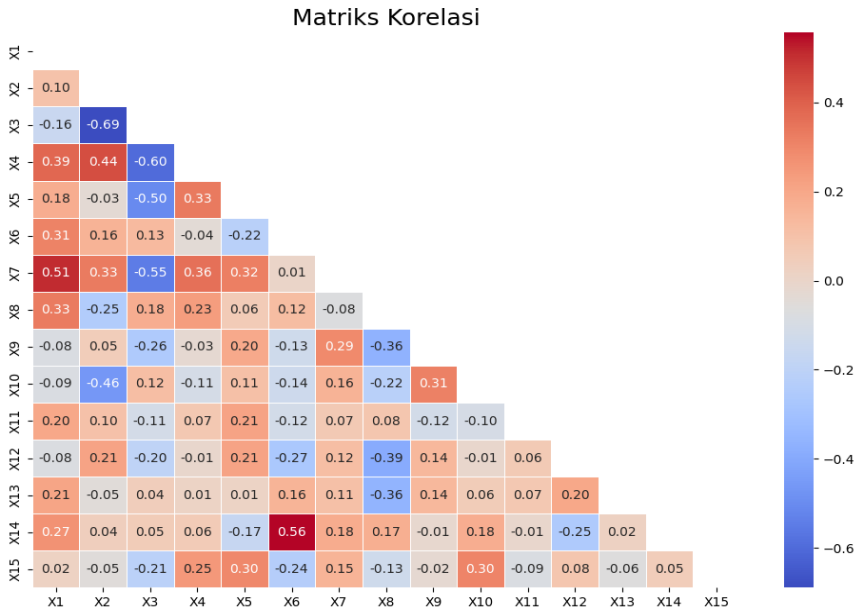
HASIL DAN PEMBAHASAN

Uji Multikolinearitas

Pada penelitian ini, sebelum melakukan proses *clustering* pada data, akan dilakukan uji multikolinearitas menggunakan koefisien korelasi (korelasi *pearson*). Sebagai contoh akan dihitung nilai korelasi antara dua variabel, X_1 dan X_2 pada data pertama, Kecamatan Bambanglipuro 2023. Telah diketahui dari data bahwa jumlah data $(n) = 34$; $X1_1 = 5{,}9$; dan $X2_1 = 8{,}2$, sehingga diperoleh $x1_1^2 = 34{,}81$; $x2_1^2 = 67{,}24$; dan $X1_1.X2_1 = 48{,}38$. Perhitungan serupa kemudian dilakukan untuk semua data pada semua variabel $X1$ dan $X2$. Selanjutnya menggunakan persamaan (1) ditentukan nilai korelasi antara variabel $X1$ dan $X2$ sebagai berikut.

$$r_{X1,X2} = \frac{34 \cdot 1280{,}455278 - 197{,}967 \cdot 216{,}717}{\sqrt{(34 \cdot 1420{,}631 - (197{,}967)^2) \cdot (34 \cdot 1512{,}677 - (216{,}717)^2)}} \approx 0{,}10$$

Dengan cara yang sama dapat ditentukan nilai korelasi untuk semua variabel. Adapun untuk visualisasi hasil dari uji multikolinearitas menggunakan koefisien korelasi disajikan menggunakan *heatmap* yang disajikan pada Gambar 3.



Gambar 3. Heatmap Hasil Uji Multikolinearitas

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Berdasarkan Gambar 3 diketahui bahwa koefisien korelasi antar variabel menunjukan korelasi yang rendah dengan nilai korelasi tertinggi terdapat pada korelasi antara variabel X_2 dan X_3, yaitu sebesar $-0{,}69$. Dengan demikian, dataset dapat langsung digunakan dalam proses *clustering* tanpa harus dilakukan modifikasi data lebih lanjut karena tidak menunjukan adanya korelasi yang tinggi antar variabel.

Clustering Menggunakan Algoritma *K-Means*

Pertama, akan dilakukan *clustering* dengan algoritma *K-Means*. *Clustering* dilakukan dengan jumlah *cluster* dua sampai lima *cluster*. Akan tetapi, pada penelitian ini akan ditnjukkan proses *clustering* dengan jumlah *cluster* sebanyak tiga *cluster*. Adapun proses *clustering* adalah sebagai berikut.

1. Inisialisasi *centroid* sebanyak tiga *centroid* yang dipilih secara acak dari data. Secara acak didapatkan *centroid* awal sebagai berikut.

Tabel 2. *Centroid* Awal

Cluster	**Kecamatan**	***X*1**	…	***X*15**
1	Jetis 2023	5,5	…	96,07
2	Pajangan 2023	6,6	…	99,9
3	Piyungan 2022	6,7	…	98,93

2. Penentuan jarak menggunakan persamaan (2). Sebagai contoh, dibawah ini dilakukan untuk jarak antara baris pertama yaitu Bambanglipuro 2022 terhadap ketiga *centroid* yang telah ditentukan.

$$d(d_1, c_1) = \sqrt{(5{,}90 - 5{,}50)^2 + \cdots + (96{,}70 - 96{,}07)^2} = 44{,}83$$
$$d(d_1, c_2) = \sqrt{(5{,}90 - 6{,}60)^2 + \cdots + (96{,}70 - 99{,}90)^2} = 23{,}98$$
$$d(d_1, c_3) = \sqrt{(5{,}90 - 6{,}70)^2 + \cdots + (96{,}70 - 98{,}93)^2} = 35{,}34$$

   Perhitungan jarak di atas selanjutnya dilakukan untuk semua data.
3. Dari perhitungan sebelumnya dapat diketahui data tersebut termasuk kedalam *cluster* mana berdasarkan jarak terpendek yang diperoleh. Sebagai contoh, pada perhitungan data Bambanglipuro 2022, karena jarak $d(d_1, c_2) = 23{,}98$ merupakan jarak terpendek, maka data tersebut termasuk kedalam *cluster* kedua.
4. Pembaruan nilai *centroid* selanjutnya dilakukan berdasarkan rata-rata dari anggota setiap *cluster*. Sebagai contoh, pada hasil perhitungan sebelumnya, terdapat 8 data pada *cluster* pertama, sehingga perhitungan nilai *centroid* baru untuk *cluster* pertama adalah sebagai berikut

$$X_1 = \frac{4{,}8 + 5{,}5 + 6{,}1 + 4{,}5 + 3{,}55 + 6{,}5 + 6{,}3 + 5{,}2}{8} \approx 5{,}306$$

5. Langkah 2 sampai 4 selanjutnya diulangi hingga anggota dari setiap *cluster* sama pada iterasi ke-i dan pada iterasi ke-$(i - 1)$. Pada perhitungan ini, didapatkan hasil akhir *clustering* bahwa 14 data kecamatan yang termasuk anggota *cluster* 1, 8 data kecamatan yang termasuk anggota *cluster* 2, dan 12 data kecamatan yang termasuk anggota *cluster* 3.

   Perhitungan serupa kemudian juga dilakukan dengan jumlah *cluster* 3, 4, dan 5.

Clustering Menggunakan Algoritma *K-Medoids*

Langkah selanjutnya adalah *clustering* menggunakan algoritma *K-Medoids*. *Clustering* ini dilakukan dengan jumlah *cluster* dua sampai lima *cluster*. Akan tetapi, pada penelitian ini

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

akan ditunjukkan proses *clustering* dengan jumlah *cluster* sebanyak tiga *cluster*. Adapun proses *clustering* adalah sebagai berikut.

1. Inisialisasi *medoids* awal sebanyak tiga *medoids* yang dipilih secara acak dari data. Secara acak didapatkan *centroid* awal sebagai berikut.

Tabel 3. *Medoids* Awal

Medoids	**Kecamatan**	$X1$	…	$X15$
1	Pleret 2023	7,1	…	100
2	Dlingo 2022	11,19	…	98,17
3	Pajangan 2022	7,2	…	99,72

2. Penentuan jarak dengan cara yang sama seperti pada penentuan jarak algoritma *K-Means*, yaitu menggunakan persamaan (2). Sebagai contoh, dari perhitungan jarak antara baris pertama yaitu Bambanglipuro 2022 terhadap ketiga *medoids* yang telah ditentukan, didapatkan $d(d_1, m_1) \approx 40{,}175$, $d(d_1, m_2) \approx 44{,}439$, dan $d(d_1, m_3) \approx 35{,}91$, sehingga data tersebut termasuk kedalam *cluster* ketiga. Perhitungan yang sama juga dilakukan untuk semua data.
3. Pilih secara acak kembali objek pada tiap-tiap *cluster* yang terbentuk sebagai *centroid non medoids*. dalam penelitian ini secara acak didapatkan *centroid non medoids* sebagai berikut.

Tabel 4. *Centroid Non Medoids*

Centroid* Non *Medoids	**Kecamatan**	$X1$	…	$X15$
1	Sanden 2022	3,1	…	100
2	Dlingo 2023	6,75	…	99,13
3	Sewon 2023	3,55	…	99,37

4. Lakukan kembali perhitungan jarak menggunakan persamaan (2), yaitu jarak antara data terhadap tiga *centroid non medoids*.
5. Perhitungan simpangan (S) selanjutnya dilakukan. Pada perhitungan sebelumnya didapatkan nilai total jarak terdekat data ke *medoid* baru (*centroid non medoids*) sebesar 846,67321869 dan total jarak terdekat data ke *medoid* awal sebesar 831,4251813. Berdasarkan nilai total jarak tersebut, didapatkan nilai simpangan sebesar 15,2480373.
6. Langkah 2 sampai 5 selanjutnya diulangi hingga nilai $S > 0$. Pada perhitungan ini, karena telah didapat nilai $S = 15{,}2480373 > 0$, maka iterasi dihentikan. Didapatkan hasil akhir *clustering* 13 data kecamatan yang termasuk anggota *cluster* 1, 13 data kecamatan yang termasuk anggota *cluster* 2, dan 8 data kecamatan yang termasuk anggota *cluster* 3.

Perhitungan serupa kemudian juga dilakukan dengan jumlah *cluster* 3, 4, dan 5.

Validasi *Clustering*

Tahapan setelah *cluster*ing pada penelitian ini adalah melakukan validasi *cluster*ing. Penelitian ini menggunakan metode *Davies Bouldin Index* dan *Silhouette Coefficient*. Adapun hasil validasi dapat ditemukan pada Tabel 5.

Tabel 5. Nilai Validasi *Cluster*ing

Jumlah *Cluster*	**Nilai Validasi *K-Means***		**Nilai Validasi *K-Medoids***	
	DBI	**SC**	**DBI**	**SC**
2	1,8013765237	0,1777996650	2,0778739079	0,1216096029
3	1,3829301131	0,2088311419	1,6433729552	0,1525151568
4	1,3997681963	0,1958030685	1,6600276398	0,0930613857
5	1,4166393517	0,1879300591	1,7566752926	0,0595499178

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Berdasarkan Tabel 5, algoritma *K-Means* dengan jumlah *cluster* tiga adalah algoritma paling optimal (ditunjukan dengan warna kuning). Hal tersebut karena diperoleh nilai validasi DBI yang paling kecil dari perhitungan lainnya, yaitu sebesar 1,3829301131. Selain itu, diperoleh juga nilai validasi SC paling mendekati 1 dari perhitungan lainnya, yaitu sebesar 0,2088311419.

Analisis dan Interpretasi Hasil *Clustering*

Variabel-variabel yang digunakan dalam penelitian ini dinyatakan dalam satuan persentase serta masing-masing memiliki karakteristik yang berbeda. Pertama, terdapat variabel yang memiliki karakteristik semakin kecil nilai persentasenya, maka semakin baik data atau daerah tersebut, yaitu terdapat pada variabel $X1$, $X2$, $X4$, dan $X5$. Sedangkan variabel lainnya, merupakan variabel yang memiliki karakteristik semakin besar nilai persentasenya, maka semakin baik data atau daerah tersebut. Berdasarkan karakteristik tersebut, pada Tabel 6, algoritma teroptimal berdasarkan perhitungan yang memiliki nilai karakteristik kurang baik diberi warna oranye.

Tabel 6. Karakteristik *Centroid*

Variabel	*Cluster* 1	*Cluster* 2	*Cluster* 3
$X1$	5,327679	6,5484375	5,628472
$X2$	5,925	7,95625	5,837326
$X3$	82,47097	82,17536948	82,99118
$X4$	5,392301	5,535752961	5,251328
$X5$	2,694539	2,234008764	2,437616
$X6$	79,9793	96,56089065	92,10913
$X7$	97,54643	97,40625	97,66528
$X8$	83,42857	84,66875	83,24549
$X9$	90,41429	86,01875	93,20139
$X10$	80,14286	66,69375	91,10556
$X11$	87,49979	86,94152162	80,31283
$X12$	99,96429	99,93125	99,92639
$X13$	96,63527	97,58125	97,0467
$X14$	67,32295	79,63596611	84,10214
$X15$	99,30286	97,83453125	99,10469

Tabel 7. Tingkat Status Gizi Balita Hasil *Cluster*ing Teroptimal

Cluster	Tingkat Status Gizi Balita	Keterangan
1	Sedang	Tingkat status gizi balita pada *cluster* pertama dapat dikatakan sedang karena terdapat 4 variabel yang kurang baik, yaitu memiliki rata-rata tertinggi dibandingkan dengan dua *cluster* lainnya pada variabel balita dengan status gizi *overweight* (X5) serta memiliki rata-rata terendah diantara ketiga *cluster* pada pengelolaan air minum dan makanan (X6), cakupan kunjungan nifas lengkap (X13), dan akses rumah sehat (X14).
2	Buruk	Tingkat status gizi balita pada *cluster* kedua dapat dikatakan buruk karena terdapat 8 variabel yang kurang baik, yaitu memiliki rata-rata tertinggi dibandingkan dengan dua *cluster* lainnya pada variabel status gizi *stunting* (X1), *wasting* (X2), dan *underweight* (X4) serta memiliki rata-rata terendah diantara ketiga *cluster* pada balita dengan status gizi baik (normal) (X3), imunisasi dasar lengkap (X7), cakupan pelayanan kesehatan bayi (X9), cakupan Kunjungan Neonatal (KN) lengkap (X10), dan sanitasi yang layak (X15).

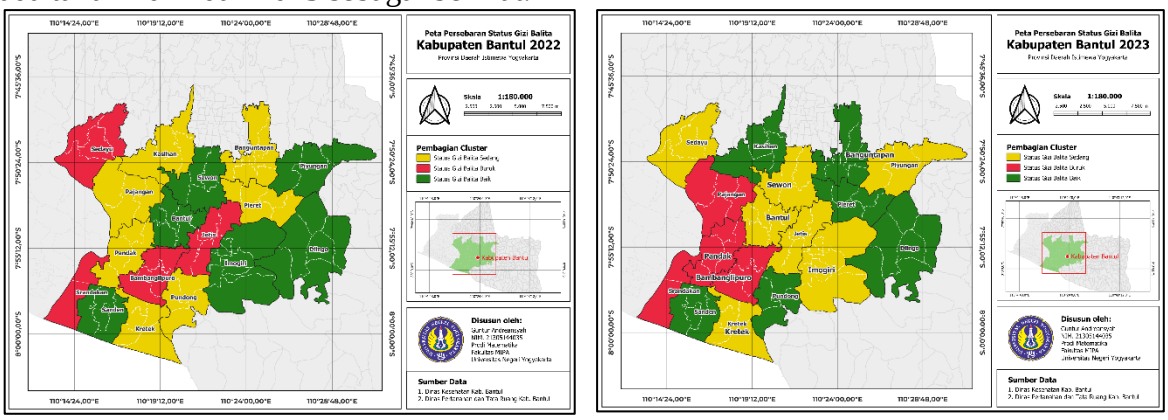
Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Cluster	**Tingkat Status Gizi Balita**	**Keterangan**
3	Baik	Tingkat status gizi balita pada *cluster* pertama dapat dikatakan sedang karena terdapat 3 variabel yang kurang baik, yaitu memiliki rata-rata terendah dibandingkan dengan dua *cluster* lainnya pada variabel pemberian Air Susu Ibu (ASI) Eksklusif (X8), layanan ANC (X11), dan persalinan di Fasyankes (X12).

Berdasarkan pada Tabel 6, terlihat bahwa *Cluster* 1 memiliki 4 variabel yang kurang baik, sehingga kategori karakteristik *cluster* berada pada kategori sedang. *Cluster* 2 memiliki 8 variabel yang kurang baik, menyebabkan *cluster* ini berada dalam kategori buruk. Sedangkan, *cluster* 3 yang memiliki 3 variabel yang kurang baik, menyebabkan *cluster* ini berada dalam kategori baik. Adapun rincian dari tingkat status gizi balita berdasarkan hasil *clustering* tersaji dalam Tabel 7 di atas.

Pemetaan dan Interpretasi

Tiga *cluster* yang terbentuk hasil dari *cluster*ing menggunakan algoritma K-Means selanjutnya divisualisasikan menjadi peta persebaran status gizi balita di Kabupaten Bantul pada tahun 2022 dan 2023 sebagai berikut.



(a) (b)

Gambar 4. Peta Status Gizi Balita Kabupaten Bantul (a) Tahun 2022 dan (b) Tahun 2023

Berdasarkan Gambar 3, dapat diketahui bahwa selama dua tahun terakhir hanya terdapat enam kecamatan yang berada pada *cluster* ketiga (hijau) dengan tingkat status gizi balita baik. Selain itu, diketahui bahwa status gizi balita di Kabupaten Bantul masih menunjukkan adanya tantangan, terlihat dari masih sedikit kecamatan yang dikategorikan sebagai kelompok kecamatan dengan tingkat status gizi balita baik. Hal ini menunjukkan bahwa meskipun sudah ada berbagai upaya intervensi untuk perbaikan status gizi, masih banyak kecamatan yang belum sepenuhnya berhasil mencapai status gizi yang baik. Namun, disisi lain terdapat kecamatan yang mengalami perubahan *cluster* ke lebih baik, yaitu Kecamatan Banguntapan, Jetis, Kasihan, Pleret, Pundong, dan Sedayu.

SIMPULAN

Hasil penelitian mengenai *clustering* untuk pemetaan kecamatan di Kabupaten Bantul berdasarkan status gizi balita tahun 2022 dan 2023, menunjukan bahwa kedua algoritma memiliki *cluster* teroptimal sebanyak tiga *cluster*. Namun, algoritma *K-Means* lebih unggul berdasarkan nilai validasi *Davies-Bouldin Index* dan *Silhouette Coefficient*, yaitu secara berturut-turut sebesar 1,3829301131 dan 0,2088311419 sedangkan untuk *K-Medoids* sebesar

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

1,6433729552 dan 0,1525151568. Hasil pengelompokan menggunakan algoritma *K-Means* didapatkan tiga tingkatan status gizi balita, yaitu *cluster* pertama (14 data kecamatan) dikategorikan menjadi status gizi balita sedang ditandai oleh rata-rata persentase *overweight* tertinggi serta rendahnya pengelolaan air minum dan makanan, cakupan KF lengkap, dan akses rumah sehat. *Cluster* kedua (8 data kecamatan) dikategorikan menjadi status gizi balita buruk dengan rata-rata persentase balita *stunting*, *wasting*, dan *underweight* tertinggi serta rendahnya rata-rata persentase balita baik (normal), imunisasi dasar lengkap, cakupan pelayanan kesehatan bayi, cakupan Kunjungan Neonatal (KN) lengkap, persalinan di Fasyankes, dan sanitasi yang layak. Terakhir, *Cluster* ketiga (12 data kecamatan) dikategorikan menjadi status gizi balita baik menunjukan rata-rata persentase pemberian Air Susu Ibu (ASI) Eksklusif dan layanan ANC terendah. Pemetaan tahun 2022 dan 2023 menunjukkan 7 kecamatan berada di *cluster* pertama, 4 kecamatan berada di *cluster* kedua, dan 6 kecamatan berada di *cluster* ketiga. Berdasarkan hasil tersebut, peneliti menyarankan untuk melakukan upaya yang lebih intensif guna meningkatkan kualitas gizi balita terutama di kecamatan *cluster* pertama dan kedua agar tercapai status gizi yang lebih baik di seluruh kecamatan. Selain itu, untuk meningkatkan keakuratan hasil, disarankan untuk mempertimbangkan penggunaan algoritma lain dalam penelitian serupa guna memperoleh hasil yang terbaik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing skripsi, koordinator program studi, dan seluruh dosen Departemen Pendidikan Matematika FMIPA UNY atas ilmu dan doa yang telah diberikan, yang telah mendukung penyelesaian artikel ini.

DAFTAR PUSTAKA

Anniza Andyani, R., Qolbi Shobri, M., Adzib Baihaqi, M., Balqis Al-Kubro, P., Syahriandi Adhantoro, M., & Penulis Rhavida Anniza Andyani, K. (2024). Aplikasi Metode Ward dengan Berbagai Pengukuran Jarak (Studi Kasus: Klasifikasi Tingkat Perekonomian di Indonesia). *Jurnal Ilmiah Kampus Mengajar*, *4*(4), 177–190.

Bahri, S., Marisa Midyanti, D., & Korespondensi, P. (2023). *Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out*.

Butsianto, S., & Saepudin, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. J. Nas. Komputasi Dan Teknol. Inf, 3(1), 51–59.

Dona, D., & Rifqi, M. (2022). Penerapan Metode K-Means Clustering Untuk Menentukan Status Gizi Baik dan Gizi Buruk Pada Balita (Studi Kasus Kabupaten Rokan Hulu). *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *7*(2), 179–191.

Efendi, R., Coastera, F. F., & Tanjung, F. R. (2019). Pengelompokan dan Pemetaan Derajat Kesehatan Kota Bengkulu Dengan Metode K-Means Clustering. *Rekursif: Jurnal Informatika*, *7*(1).

Fauziah, N. A., & Apsari, P. (2010). Gambaran Peran Serta Kader Posyandu dalam Upaya Peningkatan Status Gizi Balita di Kalurahan Laweyan Surakarta Tahun 2009. *Jurnal Kebidanan Indonesia, 1(1)*.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.

Harani, N. H., Prianto, C., & Nugraha, F. A. (2020). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA), 10(2)*, 133–146.

Hidayat, F. M., Rohana, T., Nurlaelasari, E., & Masruriyah, A. F. N. (2024). Klasterisasi Kabupaten dan Kota di Jawa Barat Dalam Kasus Gizi Buruk Menggunakan Algoritma K-

Copyright © 2025, Jurnal Kajian dan Terapan Matematika

Means dan K-Medoids. *Jurnal Tekinkom (Teknik Informasi Dan Komputer)*, *7*(1), 251–261.

Julyantari, N. K. S., Budiarta, I. K., & Putri, N. M. D. K. (2021). Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih). *Jurnal Janitra Informatika Dan Sistem Informasi*, *1*(2), 92–101.

Miftahuddin, Pratama, A., & Setiawan, I. (2021). Analisis Hubungan Antara Kelembaban Relatif Dengan Beberapa Variabel Iklim. *Jurnal Siger Matematika*, *02*(01), 25–33.

Mundir. (2012). *Statistik Penddidikan Pengatar Analisis Data Untuk Penulisan Skripsi dan Tesis* (Muhibbin & H. Hasanah, Eds.; 1st ed.). STAIN Jember Press.

Nahdliyah, M. A., Widiharih, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian, 8(2)*, 161–170.

Ningrum, H., Irawan, E., & Lubis, M. R. (2021). Implementasi K-Medoids Dalam Pengelompokan Data Penyakit Alergi Pada Anak. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, *6*(1), 130–139.

Park, J., Park, K. V., Yoo, S., Choi, S. O., & Han, S. W. (2020). Development of the WEEE grouping system in South Korea using the hierarchical and non-hierarchical clustering algorithms. *Resources, Conservation and Recycling*, *161*.

Prasetyo, Y. (2013). Kesadaran Masyarakat Berolahraga untuk Peningkatan Kesehatan dan Pembangunan Nasional. *Jurnal Ilmiah Kesehatan Olahraga*, *11(2)*.

Sadeghi, M., Casey, P., Carranza, E. J. M., & Lynch, E. P. (2024). Principal components analysis and K-means clustering of till geochemical data: Mapping and targeting of prospective areas for lithium exploration in Västernorrland Region, Sweden. *Ore Geology Reviews*, *167*.

Sholeh, M., & Aeni, K. (2023). Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means. *STRING (Satuan Tulisan Ris. Dan Inov. Teknol.*, *8*, 56–65.

Supardi, N., Sinaga, T. R., Hasanah, F. L. N., Fajriana, H., Puspareni, P. L. D., Atjo, N. M., Maghfiroh, K., & Humaira, W. (2023). *Gizi pada Bayi dan Balita*. Yayasan Kita Menulis.

Tanjung, F. A., Windarto, A. P., Fauzan, M., Studi, M. P., Informasi, S., & Tunas Bangsa, S. (2021). *Penerapan Metode K-Means Pada Pengelompokkan Pengangguran Di Indonesia*. *6*, 61–74.

Copyright © 2025, Jurnal Kajian dan Terapan Matematika