

# Perbandingan Metode K-Prototypes dan K-Medoids dengan Gower Distance dalam Klasterisasi Prioritas Penanganan Stunting di Jawa Timur

Deanita Nur Fauzizah<sup>1</sup>, Dwi Arman Prasetya<sup>2</sup>, Muhammad Nasrudin<sup>3</sup>

<sup>1</sup>Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur

<sup>2</sup>Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur

<sup>3</sup>Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur

<sup>1</sup>[22083010005@student.upnjatim.ac.id](mailto:22083010005@student.upnjatim.ac.id)

<sup>2</sup>[arman.prasetya.sada@upnjatim.ac.id](mailto:arman.prasetya.sada@upnjatim.ac.id)

<sup>3</sup>[nasrudin.fasilkom@upnjatim.ac.id](mailto:nasrudin.fasilkom@upnjatim.ac.id)

## ABSTRAK

Stunting masih menjadi permasalahan serius di Indonesia, termasuk di Provinsi Jawa Timur yang memiliki karakteristik sosial, ekonomi, dan kesehatan yang beragam. Upaya penanganan stunting perlu dilakukan berdasarkan prioritas wilayah dengan kondisi serupa agar intervensi dapat lebih tepat sasaran. Penelitian ini bertujuan untuk membandingkan kinerja metode K-Prototypes dan K-Medoids dengan Gower Distance dalam mengelompokkan kabupaten/kota di Jawa Timur berdasarkan indikator penyebab stunting. Variabel yang digunakan meliputi cakupan ASI eksklusif, akses sanitasi aman, ketidakcukupan pangan, imunisasi lengkap, angka kemiskinan, gini ratio, dan prevalensi stunting. Proses klasterisasi dilakukan menggunakan Python dengan tahapan pembersihan data, normalisasi, penentuan jumlah kluster optimal menggunakan metode Elbow, serta evaluasi kualitas kluster menggunakan Silhouette Score dan Davies Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa metode K-Prototypes menghasilkan struktur kluster yang lebih baik dengan nilai Silhouette Score sebesar 0.4317 dan DBI sebesar 0.8080, dibandingkan K-Medoids dengan nilai Silhouette Score 0.2898 dan DBI 1.3569. Klasterisasi menghasilkan tiga kelompok wilayah prioritas, yaitu kluster prioritas tinggi, sedang, dan rendah, berdasarkan kondisi sosial-ekonomi dan kesehatan masyarakat. Dengan demikian, metode K-Prototypes dinilai lebih efektif untuk mengelompokkan data campuran dalam analisis prioritas penanganan stunting di Jawa Timur.

Keywords K-Prototypes, K-Medoids, Gower Distance, Klasterisasi, Stunting, Jawa Timur

## I. PENDAHULUAN

Stunting merupakan salah satu masalah kesehatan yang masih menjadi fokus utama di Indonesia. Stunting di tandai dengan gangguan pertumbuhan yang terjadi pada anak-anak yang disebabkan oleh kekurangan gizi yang berlangsung lama yang berdampak pada perkembangan fisik, kognitif, dan produktivitas [1]. Berdasarkan data Survei Status Gizi Indonesia (SSGI) tahun 2024 menunjukkan bahwa jumlah stunting di seluruh Indonesia mencapai 19,8%, atau setara dengan sekitar 4,6 juta balita dengan kondisi stunting di Indonesia, angka ini masih sangat tinggi jika dibandingkan dengan standar kesehatan *World Health Organization* (WHO), yang menetapkan ambang batas stunting maksimal 20% dari total populasi balita[2][3]. Kondisi ini menunjukkan bahwa penurunan stunting masih menjadi masalah besar bagi Indonesia. Oleh karena itu, diperlukan berbagai pendekatan intervensi yang dapat disesuaikan dengan kondisi wilayah dan perbedaan sosial-ekonomi di antara kabupaten dan kota dengan tingkat stunting yang berbeda.

Di Provinsi Jawa Timur, prevalensi stunting menunjukkan tren penurunan, turun dari 19,2% pada tahun 2022 menjadi 14,7% pada tahun 2024, yang setara dengan sekitar 430.780 balita. Hasil Survei Status Gizi Indonesia (SSGI) dan publikasi resmi Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan (Kemenko PMK) menunjukkan bahwa beberapa kabupaten, seperti Jember (34,9%), Bondowoso (32%), dan Situbondo (30,9%), masih menunjukkan prevalensi diatas rata-rata. sehingga hal ini masih menjadi fokus utama pemerintah daerah dalam upaya penanganan stunting[2]. Perbedaan ini menunjukkan bahwa ada perbedaan di antara daerah tertentu, olah karena itu masalah ini menjadi dasar untuk menentukan seberapa penting penanganan stunting di Jawa Timur. Selain itu, Menurut Badan Pusat Statistik 2024, menunjukkan bahwa pendapatan keluarga memiliki hubungan signifikan dengan kejadian stunting, anak-anak yang berasal dari keluarga berpenghasilan rendah memiliki risiko lebih tinggi untuk menderita stunting [4].

Perbedaan karakteristik antarwilayah tersebut menunjukkan bahwa penyebab stunting tidak dapat disamaratakan. Setiap kabupaten/kota memiliki kondisi sosial-ekonomi, akses sanitasi, cakupan imunisasi, tingkat kemiskinan, dan ketahanan pangan yang berbeda-beda [2]. Oleh karena itu, diperlukan pendekatan analisis yang mampu mengidentifikasi kesamaan karakteristik antarwilayah sehingga wilayah yang memiliki kondisi serupa dapat dikelompokkan dalam satu kluster. Pendekatan clustering menjadi solusi yang tepat karena mampu mengelompokkan wilayah berdasarkan faktor-faktor penyebab stunting dan memberikan gambaran pola keterkaitan antarvariabel [5]. Melalui metode ini, pemerintah dapat menentukan prioritas intervensi penanganan stunting secara lebih spesifik dan berbasis kebutuhan wilayah.

Beberapa penelitian sebelumnya telah membuktikan efektivitas algoritma clustering dalam bidang kesehatan masyarakat dan pembangunan sosial. Penelitian oleh [5] menggunakan algoritma K-Means untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan faktor penyebab stunting. Penelitian ini menghasilkan dua kluster optimal melalui elbow method, yaitu kluster dengan tingkat faktor penyebab stunting tinggi (324 kabupaten/kota) dan rendah (49 kabupaten/kota), dengan nilai selisih SSE sebesar 1401,5156. Hasil ini membuktikan bahwa metode K-Means efektif dalam mengidentifikasi wilayah dengan karakteristik serupa terkait stunting. Selanjutnya, penelitian [6] membandingkan pengukuran jarak Euclidean dan Gower pada algoritma K-Medoids. Hasil penelitian menunjukkan bahwa Euclidean distance unggul dalam metrik Silhouette dan Connectivity, sedangkan Gower distance unggul dalam Dunn Index. Kesimpulannya, Gower distance lebih baik digunakan untuk data campuran (numerik dan kategorikal), yang banyak ditemui dalam analisis sosial dan kesehatan. Temuan ini mendukung penggunaan K-Medoids dengan Gower distance untuk data stunting yang bersifat heterogen.

Penelitian [7] menerapkan K-Prototypes untuk mengelompokkan kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan indikator kesejahteraan rakyat (pendapatan, pendidikan, kesehatan, dan fasilitas). Hasilnya, metode ini menghasilkan dua kluster optimal, kluster pertama dengan kesejahteraan tinggi, dan kluster kedua dengan kesejahteraan rendah. Hal ini menunjukkan bahwa K-Prototypes efektif untuk data campuran dan relevan digunakan untuk pengelompokan wilayah berdasarkan indikator sosial-ekonomi. Berdasarkan penelitian-penelitian terdahulu menunjukkan masih memiliki beberapa keterbatasan. Sebagian besar penelitian terdahulu hanya menggunakan satu metode klusterisasi, sehingga belum memberikan gambaran mengenai metode mana yang paling tepat untuk mengelompokkan wilayah dengan karakteristik data campuran. Penelitian yang membandingkan kinerja metode klusterisasi seperti K-Prototypes dan K-Medoids dengan Gower Distance masih jarang dilakukan pada konteks penentuan prioritas penanganan stunting di tingkat kabupaten/kota, khususnya di Provinsi Jawa Timur yang memiliki karakteristik wilayah, sosial, dan budaya yang beragam. K-Prototypes memiliki keunggulan dalam mengelompokkan data campuran, namun metode ini sensitif terhadap outlier dan variasi skala data. Sementara

itu, K-Medoids dengan Gower Distance lebih mampu menjelaskan kemiripan antarwilayah berdasarkan variabel yang berbeda skala, serta lebih stabil dalam kondisi data yang heterogen. Oleh karena itu, diperlukan analisis perbandingan untuk mengetahui metode mana yang memberikan struktur klaster yang stabil untuk penentuan prioritas kebijakan penanganan stunting di Jawa Timur.

Selain pemilihan metode, pemilihan variabel yang digunakan dalam proses klasterisasi juga dibuktikan dengan penelitian sebelumnya. Berdasarkan penelitian [8] status ketahanan pangan rumah tangga memiliki hubungan signifikan dengan kejadian stunting, di mana rumah tangga yang berada dalam kondisi rawan pangan dengan kelaparan memiliki risiko hingga 153 kali lebih tinggi memiliki balita stunting dibandingkan rumah tangga dengan ketahanan pangan baik ( $p < 0,001$ ). Hal ini menunjukkan bahwa akses dan kualitas pangan merupakan penentu utama dalam pertumbuhan anak. Sementara itu, penelitian [9] anak yang tidak mendapatkan ASI eksklusif tercatat memiliki risiko 24 kali lebih besar mengalami stunting dibandingkan anak yang diberikan ASI eksklusif secara penuh ( $p < 0,05$ ). Selain pemberian ASI eksklusif, menurut penelitian [9] riwayat berat badan lahir rendah (BBLR) juga memiliki pengaruh yang signifikan terhadap stunting dengan Odds Ratio sebesar 49,88 (95% CI = 1,842–96,078), balita dengan riwayat BBLR memiliki kemungkinan hampir lima puluh kali lebih besar mengalami stunting dibandingkan balita dengan berat lahir normal.

Selain itu, Penelitian [10] menyatakan bahwa anak yang tinggal di lingkungan dengan sanitasi buruk memiliki prevalensi stunting yang lebih tinggi dibandingkan anak yang tinggal di lingkungan dengan sanitasi layak. Dengan demikian, persentase akses sanitasi layak menjadi variabel yang relevan dalam melihat kerentanan wilayah terhadap stunting. Dengan demikian, persentase akses sanitasi layak menjadi variabel yang relevan dalam melihat kerentanan wilayah terhadap stunting. Berdasarkan kajian variabel-variabel tersebut, penelitian ini akan menggunakan indikator ketahanan pangan, pemberian ASI eksklusif, riwayat BBLR, akses sanitasi layak, kemiskinan, ketidakcukupan pangan, serta gini ratio sebagai dasar pengelompokan wilayah. Pemilihan variabel ini diharapkan mampu memberikan gambaran yang lebih menyeluruh mengenai kondisi penyebab stunting di setiap kabupaten/kota. Dengan menerapkan perbandingan metode K-Prototypes dan K-Medoids berbasis Gower Distance, penelitian ini bertujuan untuk menentukan struktur klaster yang paling stabil dan representatif, sehingga dapat mendukung penyusunan strategi penanganan stunting yang lebih tepat sasaran di Provinsi Jawa Timur.

## II. METODOLOGI PENELITIAN

### II.1 Variabel Penelitian

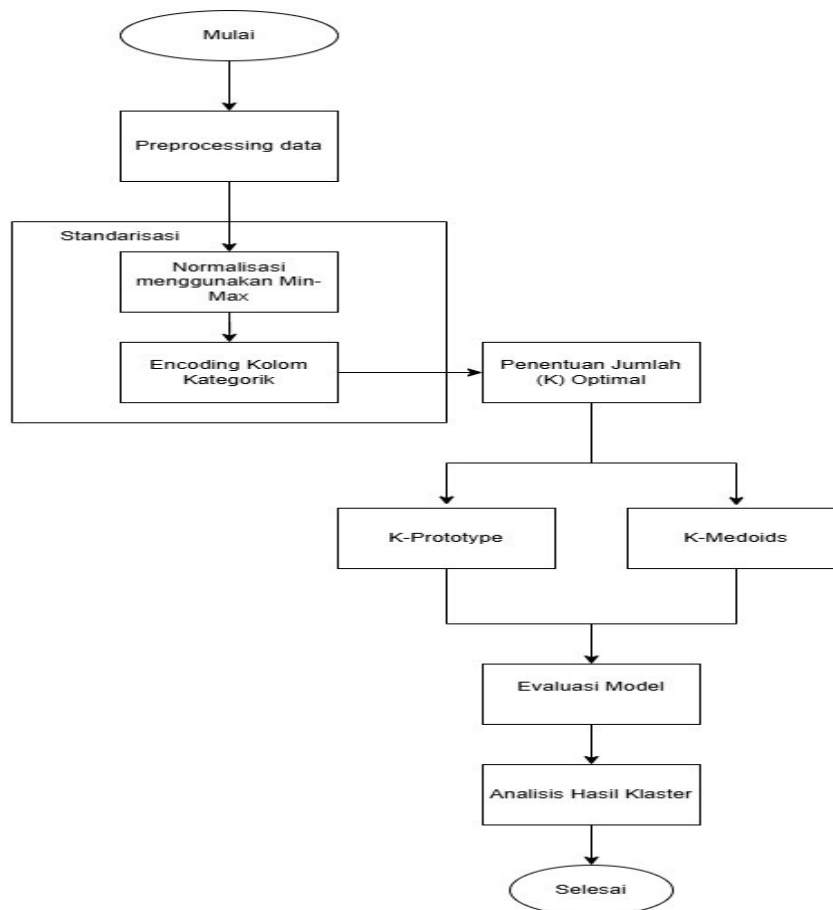
Dalam penelitian ini menggunakan data dari Dinas Kesehatan Provinsi Jawa Timur Tahun 2024 dan Badan Pusat Statistika Tahun 2024. Untuk data kategori menggunakan pengkategorian seperti penelitian [11] sebagai referensi dalam menentukan rentang kategori. Analisis klaster dilakukan berdasarkan sejumlah variabel yang mencakup indikator kesehatan, ekonomi, dan gizi anak. Tabel berikut menyajikan daftar variabel yang digunakan beserta tipe datanya:

**Tabel 2.1 Variabel Dataset**

| Variabel | Keterangan            | Tipe Data   |
|----------|-----------------------|-------------|
| $X_1$    | Cakupan ASI Eksklusif | Numerik     |
| $X_2$    | Akses Sanitasi Aman   | Numerik     |
| $X_3$    | Ketidakcukupan        | Kategorikal |

|     |                                 |             |
|-----|---------------------------------|-------------|
|     | Pangan                          |             |
| 224 | Imunisasi lengkap               | Numerik     |
| 225 | Angka Kemiskinan                | Numerik     |
| 226 | Gini Ratio                      | Kategorikal |
| 227 | Prevalensi Stunting             | Kategorikal |
| 228 | Berat Badan Rendah Lahir (BBLR) | Numerik     |

Gambar berikut menyajikan alur kerja penelitian yang menggambarkan proses pengolahan data hingga pembentukan kluster. Workflow ini menunjukkan tahapan mulai dari *preprocessing* data, standarisasi, penentuan jumlah kluster optimal, penerapan metode K-Prototypes dan K-Medoids, hingga evaluasi serta analisis hasil klusterisasi.



**Gambar 2.1** Workflow

Workflow pada gambar 2.1 penelitian ini disusun untuk menggambarkan secara sistematis tahapan pengolahan data hingga analisis hasil klusterisasi. Proses dimulai dengan melakukan *preprocessing* data sebagai langkah awal untuk memastikan kualitas dan konsistensi dataset. Tahap ini dilanjutkan dengan standarisasi yang mencakup normalisasi menggunakan metode Min-Max serta *encoding* pada atribut kategorik agar seluruh variabel berada dalam skala yang seragam. Setelah data distandarisasi, dilakukan penentuan jumlah kluster (K) optimal sebagai dasar dalam penerapan dua metode klusterisasi, yaitu K-Prototypes dan K-Medoids. Kedua metode tersebut

kemudian dievaluasi untuk mengukur performanya, sebelum akhirnya dianalisis lebih lanjut untuk memperoleh interpretasi dan pemahaman terhadap hasil kluster yang terbentuk.

## II.2 Analisis Kluster

Analisis kluster merupakan metode statistik yang digunakan untuk mengelompokkan objek ke dalam kelompok berdasarkan tingkat kemiripan karakteristik antar objek sehingga objek dalam satu kluster memiliki pola yang sama [11]. Metode kluster banyak digunakan dalam analisis data kesehatan dan sosial untuk mengidentifikasi wilayah dengan karakteristik permasalahan yang sejenis, seperti pada penentuan kluster stunting[12]. Pada penelitian ini analisis kluster menggunakan metode perbandingan metode k-prototype dan k-medoids dengan gower distance. Penelitian ini menggunakan software python untuk melakukan klusterisasi. Berikut adalah tahapan-tahapan yang digunakan untuk melakukan analisis kluster sebagai berikut:

1. Melakukan pembersihan data dengan menghapus outlier.
2. Melakukan pemisahan variabel pada data numerik dan kategori, selanjutnya data kategori di encoding dengan skala 0-2.
3. Normalisasi data dengan menggunakan Min-Max Normalisasi diterapkan untuk menyamakan skala antar variabel numerik agar tidak terjadi dominasi variabel yang memiliki rentang nilai besar. Untuk menentukan nilai min-max normalisasi menggunakan rumus yang ditulis pada persamaan (1) sebagai berikut:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Dimana:

$X$  = Nilai asli sebelum dinormalisasi

$X'$  = Nilai setelah dinormalisasi (rentang 0-1)

$X_{\min}$  = Nilai minimum dalam variabel

$X_{\max}$  = Nilai maksimum dalam variabel

4. Menentukan Jumlah Kluster Optimal pada penelitian ini menggunakan *Metode Elbow*. Proses dilakukan dengan menjalankan algoritma K-Prototypes untuk rentang nilai k 2-6. Sementara itu, untuk nilai cost dihitung berdasarkan total jarak campuran dalam kluster, dengan jarak numerik menggunakan squared Euclidean distance dan jarak kategorikal dihitung dengan fungsi kesamaan *Hamming Dissimilarity* (2). Nilai cost yang dihasilkan kemudian diplot ke dalam grafik untuk mengidentifikasi titik siku (elbow point), yaitu titik dimana penurunan nilai cost mulai melambat. Titik tersebut digunakan sebagai dasar pemilihan jumlah kluster optimal. Rumus persamaan squared Euclidean distance untuk menghitung variabel numerik ditulis dalam persamaan (2) sebagai berikut:

$$d_{\text{num}}(x_i, x_j) = \sum_{k=1}^n (x_{ik} - x_{jk})^2 \quad (2)$$

Dimana:

$x_i$  = nilai data pada variabel numerik ke- $i$

$x_j$  = nilai centroid (pusat kluster) pada variabel numerik ke- $j$

Selanjutnya, untuk mengukur nilai variabel kategori menggunakan persamaan *Hamming Dissimilarity* yang ditulis dalam persamaan (3) sebagai berikut:

$$d_{\text{cat}}(x_i, x_j) = 0, \text{ jika } x_i = x_j, \text{ dan } 1 \text{ jika } x_i \neq x_j \quad (3)$$

Tahapan selanjutnya untuk mengukur jarak variabel campuran yaitu numerik dan

kategorikal menggunakan rumus persamaan K-Prototype untuk rumus persamaanya ditulis dalam persamaan (4) sebagai berikut:

$$W(\mathbf{z}_i, \mathbf{z}_j) = \sum_{k \in \text{numerical attributes}} (z_{ik} - z_{jk})^2 + \sum_{k \in \text{categorical attributes}} W_k(z_{ik}, z_{jk}) \quad (4)$$

Dimana:

$W$  = bobot untuk menyeimbangkan variabel numerik dan kategorikal

Selanjutnya, evaluasi kluster campuran dengan jarak gower ditulis dengan rumus persamaan (5) sebagai berikut:

$$W_{ij} = \frac{\sum_{k=1}^p \delta_{ijk}}{\sum_{k=1}^p W_k} \quad (5)$$

Dimana:

$p$  = jumlah atribut

$W_k$  = bobot atribut

$\delta_{ijk}$  = jarak atribut ke-  
k

$\delta_{ijk} = 1$  jika nilai tersedia pada kedua objek

Matriks jarak ini digunakan sebagai input untuk visualisasi *Multidimensional Scaling* (MDS) 2D dan evaluasi kualitas kluster. Rumus persamaan MSD ditulis dalam persamaan (6) sebagai berikut:

$$D_{ij} = \sqrt{\sum_{k=1}^p (z_{ik} - z_{jk})^2} \quad (6)$$

Dimana:

$D_{ij}$  = jarak asli antara objek  $i$  dan  $j$  (misal dari matriks Gower)

$\hat{D}_{ij}$  = jarak Euclidean antara objek  $i$  dan  $j$  di ruang baru hasil MDS

$z_{ik}$  = koordinat objek  $i$  pada dimensi  $k$  di ruang baru

- Setelah menentukan jumlah kluster optimal menggunakan Metode Elbow pada K-Prototypes, penelitian ini juga membandingkan metode K-Medoids untuk mengevaluasi stabilitas kluster pada data campuran. Untuk rumus persamaan K-Medoids dituliskan dalam persamaan (7) sebagai berikut:

$$W(\mathbf{z}_i, \mathbf{z}_j) = \sum_{k=1}^p W_k \delta_{ijk}, \quad \mathbf{z}_k = \arg \min_{\mathbf{z}_k \in \mathbf{Z}} \sum_{i \in \mathbf{Z}_k} W(\mathbf{z}_i, \mathbf{z}_k) \quad (7)$$

Dimana:

$\delta_{ijk}$  = jarak pada atribut ke- $k$  (numerik, kategorikal, atau campuran)

Selain K-Prototypes, K-Medoids juga menggunakan matriks jarak Gower sebagai ukuran dissimilarity antar objek, sehingga dapat menangani data campuran dengan atribut numerik dan kategorikal secara bersamaan. Matriks jarak Gower ini kemudian digunakan sebagai input untuk Multidimensional Scaling (MDS) 2D, sehingga hasil kluster dapat divisualisasikan dalam ruang berdimensi rendah dan pola pemisahan antar kluster. Untuk rumus persamaannya seperti yang telah dituliskan di persamaan (5) dan (6).

- Evaluasi model kualitas kluster

Evaluasi kualitas kluster pada penelitian ini menggunakan *Silhouette Score* dan *Davies–Bouldin Index* (DBI). Untuk rumus persamaan *Silhouette Score* ditulis dalam persamaan (8) sebagai berikut:

$$S_i = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (8)$$

Dimana:

$\bar{d}(i)$  = jarak rata-rata objek ke anggota klasternya sendiri

$\bar{d}(i)$  = jarak rata-rata objek ke klaster terdekat

Nilai mendekati 1 menunjukkan objek berada di klaster yang sesuai

Selanjutnya, rumus persamaan *Davies-Bouldin Index (DBI)* ditulis dalam Persamaan (9) sebagai berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{j \neq k} \frac{S_k + S_j}{M_{kj}} \quad (9)$$

Dimana:

$K$  = jumlah klaster

$S_k$  = jarak rata-rata anggota ke centroid/medoid klaster  $k$  (mengukur compactness)

$M_{kj}$  = jarak antara centroid/medoid klaster  $k$  dan  $j$  (mengukur separation)

Berdasarkan perhitungan menggunakan data penelitian, evaluasi kualitas klaster untuk masing-masing metode dapat dilihat pada tabel 2.1 sebagai berikut:

**Tabel 2.2 Evaluasi Kualitas Klaster**

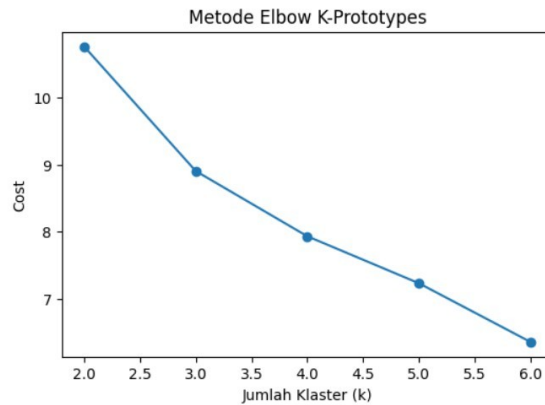
| Evaluasi Kualitas Klaster | K-Prototypes | K-Medoids |
|---------------------------|--------------|-----------|
| Silhouette Score          | 0.4317       | 0.2898    |
| Davies-Bouldin Index      | 0.8080       | 1.3569    |

### III. HASIL DAN ANALISIS

Pada tahap hasil dan analisi ini menyajikan hasil pengolahan data dan interpretasi yang diperoleh dari proses analisis. Data stunting yang telah dikumpulkan diolah menggunakan metode klasterisasi untuk mengelompokkan daerah berdasarkan kesamaan karakteristik indikator cakupan ASI eksklusif, akses sanitasi aman, ketidacukupan pangan, imunisasi lengkap, angka kemiskinan, gini ratio, prevalensi stunting, dan prevalensi berat badan lahir rendah. Tahapan analisis meliputi pemilihan variabel, standarisasi data, penentuan jumlah klaster, hingga interpretasi hasil pengelompokan. Hasil akhir dari klasterisasi ini kemudian dianalisis untuk memahami perbedaan kondisi antarwilayah terhadap upaya penanganan stunting.

#### III.1 Menentukan Jumlah Klaster Optimal

Penentuan jumlah klaster yang optimal pada algoritma K-Prototypes dilakukan melalui analisis metode Elbow. Metode ini berfungsi menampilkan hubungan antara jumlah klaster ( $k$ ) dan nilai *cost* yang dihasilkan. Nilai *cost* mencerminkan tingkat keseragaman data dalam setiap klaster; semakin kecil nilai *cost*, maka semakin baik hasil pengelompokan data. Gambar 3.1 berikut memperlihatkan hasil penerapan metode Elbow pada proses klasterisasi menggunakan algoritma K-Prototypes. Berikut di tampilkan gambar 3.1 metode *Elbow*.

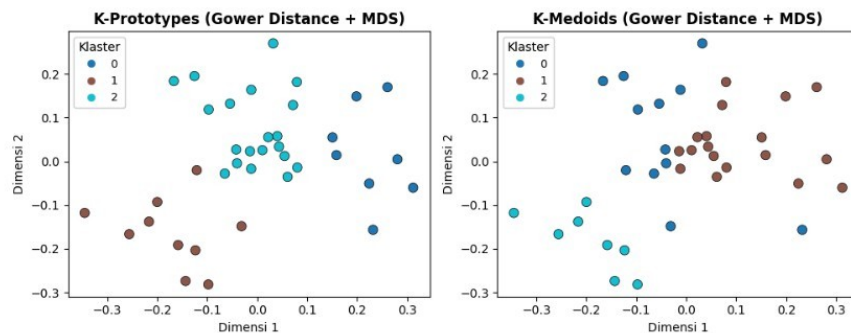


**Gambar 3.1 Metode Elbow**

Berdasarkan gambar 3.1 hasil analisis menggunakan metode *Elbow* pada algoritma K-Prototypes, diperoleh bahwa nilai cost mengalami penurunan yang cukup signifikan hingga jumlah kluster  $k = 3$ , kemudian penurunan tersebut mulai melandai pada nilai  $k$  berikutnya. Oleh karena itu, dapat disimpulkan bahwa jumlah kluster yang optimal dalam penelitian ini adalah sebanyak 3 kluster, karena pada titik tersebut terjadi keseimbangan antara efisiensi dan keakuratan klasterisasi.

### III.2 Visualisasi Kluster

Visualisasi hasil klasterisasi menggambarkan persebaran data berdasarkan hasil pengelompokan menggunakan algoritma K-Prototypes dan K-Medoids dengan pendekatan *Gower Distance* serta reduksi dimensi melalui *Multidimensional Scaling* (MDS). Proses MDS menghasilkan dua dimensi utama, yaitu Dimensi 1 dan Dimensi 2, yang merupakan representasi dari kombinasi seluruh variabel numerik dan kategorik dalam data. Kedua dimensi tersebut tidak menunjukkan variabel tertentu, akan tetapi menggambarkan posisi relatif setiap data berdasarkan tingkat kemiripan karakteristiknya. Titik-titik yang berdekatan menunjukkan kemiripan tinggi, sedangkan titik yang berjauhan mencerminkan perbedaan yang signifikan antar data. Berikut di tampilkan gambar 3.2 visualisasi hasil kluster.



**Gambar 3.2 Visualisasi Hasil Kluster**

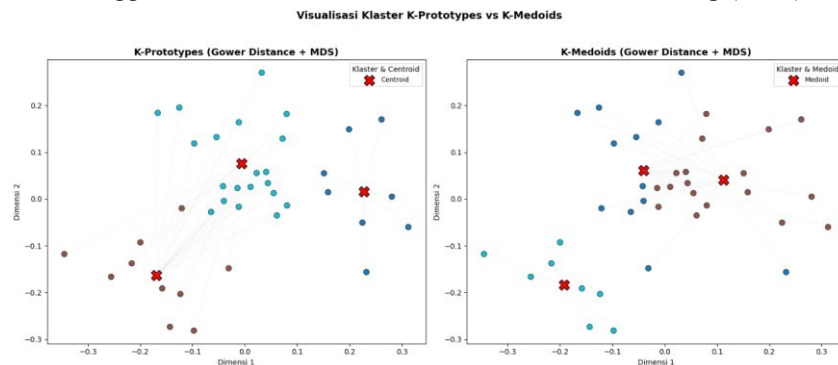
Gambar 3.2 memperlihatkan distribusi hasil klasterisasi antara K-Prototypes dan K-Medoids menggunakan pendekatan *Gower Distance* dan *Multidimensional Scaling* (MDS). Visualisasi ini memberikan gambaran mengenai pola keterpisahan antar-kluster serta tingkat kedekatan data di dalam masing-masing kluster. Berdasarkan hasil tersebut, dapat dijelaskan beberapa poin sebagai berikut:

1. Pada metode K-Prototypes, titik-titik data dari setiap kluster terlihat lebih terpisah dan membentuk kelompok yang lebih jelas. Hal ini menunjukkan bahwa pembentukan kluster oleh K-Prototypes menghasilkan pemisahan yang lebih baik antar kelompok data.

2. Sebaliknya, hasil visualisasi K-Medoids menampilkan penyebaran data yang cenderung lebih tumpang tindih antar kluster. Kondisi ini menandakan bahwa batas antar-kluster pada K-Medoids kurang tegas dibandingkan K-Prototypes.

Berdasarkan hasil visualisasi gambar 3.2, metode K-Prototypes memberikan struktur kluster yang lebih teratur dan mudah dibedakan, sedangkan K-Medoids menunjukkan pemisahan yang kurang optimal pada data campuran.

Selanjutnya, ditampilkan visualisasi pada Gambar 3.3 menampilkan hasil pembentukan kluster oleh K-Prototypes dan K-Medoids berdasarkan hasil reduksi dimensi menggunakan *Gower Distance* dan *Multidimensional Scaling* (MDS).



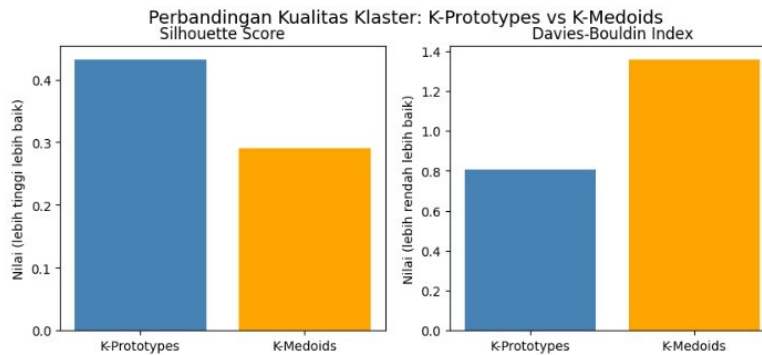
**Gambar 3.3 Visualisasi Hasil Kluster dengan Centroid**

Gambar 3.3 menampilkan hasil visualisasi kluster yang dihasilkan oleh metode K-Prototypes dan K-Medoids menggunakan pendekatan *Gower Distance* dan *Multidimensional Scaling* (MDS). Setiap titik menggambarkan data yang dikelompokkan berdasarkan kesamaan karakteristik, sedangkan tanda “X” merah menunjukkan posisi pusat kluster (centroid pada K-Prototypes dan medoid pada K-Medoids).

1. Pada metode K-Prototypes, posisi centroid terlihat berada di tengah kelompok data yang mewakili tiap kluster dengan baik. Penyebaran titik di sekitar centroid tampak lebih seimbang, menandakan pembentukan kluster yang lebih representatif terhadap karakteristik data.
2. Pada metode K-Medoids, posisi medoid cenderung lebih dekat dengan beberapa titik tertentu dan tidak sepenuhnya berada di tengah distribusi kluster. Hal ini memperlihatkan bahwa hasil klusterisasi K-Medoids memiliki penyebaran data yang lebih tumpang tindih dan batas antar-kluster yang kurang tegas dibandingkan K-Prototypes.

### III.3 Perbandingan Kualitas Kluster

Perbandingan kualitas kluster hasil pengelompokan data, digunakan untuk melihat perbandingan antara dua metode klusterisasi yaitu K-Prototypes dan K-Medoids. Evaluasi dilakukan menggunakan dua metrik utama, yaitu *Silhouette Score* dan *Davies-Bouldin Index*. Nilai *Silhouette Score* yang lebih tinggi menunjukkan kluster yang lebih kompak dan terpisah dengan baik, sedangkan nilai *Davies-Bouldin Index* yang lebih rendah menandakan kluster yang memiliki tingkat kesamaan internal lebih tinggi dan jarak antar-kluster yang lebih besar. Berikut di tampilkan gambar 3.4 perbandingan kualitas kluster.



**Gambar 3.4 Perbandingan Kualitas Kluster**

Berdasarkan gambar grafik 3.4, metode K-Prototypes menunjukkan hasil yang lebih baik dibandingkan dengan K-Medoids. Berikut nilai perbandingan berdasarkan nilai *Silhouette Score* dan *Davies-Bouldin Index* sebagai berikut:

1. *Silhouette Score* pada K-Prototypes lebih tinggi (~0.4) dibandingkan dengan K-Medoids (~0.3), yang berarti pembentukan kluster oleh K-Prototypes lebih optimal dalam memisahkan data antar-kluster.
2. *Davies-Bouldin Index* pada K-Prototypes lebih rendah (~0.8) dibandingkan K-Medoids (~1.3), yang menunjukkan bahwa kluster yang dihasilkan oleh K-Prototypes lebih kompak dan memiliki jarak antar-kluster yang lebih baik.

Maka dari nilai *Silhouette Score* dan *Davies-Bouldin Index* dapat disimpulkan bahwa K-Prototypes memberikan performa klusterisasi yang lebih baik untuk data campuran (numerik dan kategorikal) dibandingkan metode K-Medoids.

### IV.3 Hasil Kluster

Pada tahap ini dilakukan proses pengelompokan wilayah berdasarkan seluruh variabel yang digunakan dalam penelitian menggunakan metode K-Prototypes. Metode ini mampu mengelompokkan data yang terdiri dari variabel numerik dan kategorikal, sehingga karakteristik setiap daerah dapat terlihat lebih menyeluruh. Hasil klusterisasi menunjukkan bahwa wilayah di Jawa Timur terbagi ke dalam tiga kluster dengan kondisi yang berbeda. Setiap kluster menggambarkan tingkat kesejahteraan dan risiko stunting yang tidak sama, mulai dari kluster dengan kondisi paling rentan hingga kluster dengan kondisi relatif lebih baik. Rincian hasil klusterisasi K-Prototypes ditunjukkan pada Tabel 4.1 berikut.

**Tabel 4.1 Hasil Kluster K-Prototype**

| Kluster | K-Prototype                                                                       | Rata-rata                                                                                                                                                                                                                                      | Kategori                 |
|---------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| 0       | Pacitan, Jember, Bondowoso, Situbondo, Probolinggo, Bangkalan, Pamekasan, Sumenep | <p>ASI Eksklusif : 0.514</p> <p>Akses Sanitasi : 0.128</p> <p>Ketidakcukupan Pangan : 0.681</p> <p>Imunisasi : 0.573</p> <p>Kemiskinan : 0.637</p> <p>BBLR : 0.525</p> <p>Kategori Stunting : Tinggi</p> <p>Ketidakcukupan Pangan : Sedang</p> | Kluster Prioritas Tinggi |

|   |                                                                                                                                                                                                                                              | Gini Ratio : Sedang                                                                                                                                                                                                                             |                             |
|---|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------|
| 1 | Sidoarjo, Gresik,<br>Kota Blitar, Kota<br>Malang, Kota<br>Probolinggo, Kota<br>Pasuruhan, Kota<br>Mojokerto, Kota<br>Madiun, Kota<br>Surabaya, Kota<br>Batu                                                                                  | ASI Eksklusif :<br>0.780<br>Akses Sanitasi :<br>0.892<br>Ketidacukupan<br>Pangan : 0.113<br>Imunisasi : 0.736<br>Kemiskinan : 0.141<br>BBLR : 0.361<br>Kategori Stunting :<br>Tinggi<br>Ketidacukupan<br>Pangan : Rendah<br>Gini Ratio : Sedang | Klaster Prioritas<br>Rendah |
| 2 | Ponorogo,<br>Trenggalek,<br>Tulungagung,<br>Blitar, Kediri,<br>Malang, Lumajang,<br>Banyuwangi,<br>Pasuruhan,<br>Mojokerto,<br>Jombang, Nganjuk,<br>Madiun, Magetan,<br>Ngawi,<br>Bojonegoro, Tuban,<br>Lamongan,<br>Sampang, Kota<br>Kediri | ASI Eksklusif :<br>0.790<br>Akses Sanitasi :<br>0.727<br>Ketidacukupan<br>Pangan : 0.450<br>Imunisasi : 0.475<br>Kemiskinan : 0.413<br>BBLR : 0.434<br>Kategori Stunting :<br>Tinggi<br>Ketidacukupan<br>Pangan : Sedang<br>Gini Ratio : Sedang | Klaster Prioritas<br>Sedang |

Berdasarkan tabel 4.1 hasil klasterisasi menggunakan metode K-Prototypes menghasilkan tiga klaster dengan karakteristik yang berbeda berdasarkan indikator kesehatan, sosial, dan ekonomi. Penentuan prioritas penanganan dilakukan dengan mempertimbangkan variabel yang berhubungan langsung dengan risiko stunting, yaitu ketidacukupan pangan, tingkat kemiskinan, jumlah bayi dengan berat badan lahir rendah (BBLR), kategori stunting, ketidacukupan pangan (kategori), serta tingkat ketimpangan (gini ratio). Semakin tinggi nilai variabel tersebut dalam suatu klaster, maka semakin tinggi tingkat risiko prioritas penanganan.

Klaster pertama (Klaster 0) terdiri dari Pacitan, Jember, Bondowoso, Situbondo, Probolinggo, Bangkalan, Pamekasan, dan Sumenep. Klaster ini dikategorikan sebagai Klaster Prioritas Tinggi. Hal ini ditunjukkan oleh tingginya rata-rata ketidacukupan pangan dan angka kemiskinan, diikuti dengan nilai BBLR yang relatif tinggi. Selain itu, kondisi akses sanitasi pada wilayah ini berada pada level sangat rendah. Kombinasi faktor tersebut memperlihatkan adanya kerentanan yang signifikan terhadap masalah gizi kronis, sehingga risiko stunting lebih tinggi dan memerlukan intervensi pemerintah yang bersifat komprehensif. Fokus penanganan pada klaster ini diperlukan pada aspek

pemenuhan kebutuhan gizi keluarga, peningkatan sanitasi lingkungan, dan pemberdayaan sosial-ekonomi masyarakat.

Klaster kedua (Klaster 1) terdiri dari wilayah seperti Sidoarjo, Gresik, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, dan Kota Batu. Klaster ini termasuk Klaster Prioritas Rendah karena sebagian besar indikator kesehatan dan sosial ekonomi berada dalam kondisi baik. Akses sanitasi dan cakupan imunisasi berada pada rata-rata tinggi, serta ketidakcukupan pangan dan angka kemiskinan relatif rendah. Meskipun kategori stunting pada klaster ini masih berada dalam kategori tinggi, faktor risiko stunting yang bersifat struktural seperti gizi dan sanitasi telah lebih terkendali. Oleh karena itu, penanganan pada klaster ini lebih difokuskan pada penyuluhan mengenai pola asuh, peningkatan konsumsi pangan bergizi, serta pemantauan rutin pertumbuhan balita. Pendekatan ini dinilai lebih sesuai dibandingkan dengan program berskala besar yang memerlukan perubahan kebijakan.

Klaster ketiga (Klaster 2) mencakup wilayah Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Banyuwangi, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Sampang, dan Kota Kediri. Klaster ini dikategorikan sebagai Klaster Prioritas Sedang. Kondisi ketidakcukupan pangan dan angka kemiskinan berada pada tingkat menengah, namun cakupan imunisasi balita cenderung rendah. Kondisi tersebut menunjukkan bahwa faktor penyebab stunting lebih dipengaruhi oleh pola konsumsi gizi dan rendahnya cakupan layanan kesehatan. Oleh karena itu, fokus penanganan pada klaster ini ditujukan pada layanan kesehatan, peningkatan cakupan imunisasi, serta penyuluhan gizi keluarga. Berdasarkan hasil klasterisasi menunjukkan bahwa klaster dengan tingkat kemiskinan tinggi, akses sanitasi rendah, ketidakcukupan pangan tinggi, dan tingginya kasus BBLR merupakan klaster yang membutuhkan perhatian prioritas dalam penanganan stunting. Dengan berbagai karakteristik pada setiap klaster, pemerintah daerah dapat merancang program penanganan yang lebih sesuai dengan kondisi serta kebutuhan di masing-masing wilayah.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai perbandingan metode K-Prototypes dan K-Medoids dengan Gower Distance dalam mengelompokkan wilayah prioritas penanganan stunting di Provinsi Jawa Timur, menunjukkan bahwa hasil dari perbandingan K-Prototypes lebih unggul dengan *Silhouette Score* 0.4317 dan *Davies–Bouldin Index* 0.8080 dibandingkan K-Medoids. Metode K-Prototypes membentuk tiga klaster utama yang mewakili wilayah prioritas tinggi, sedang, dan rendah. Dengan demikian, K-Prototypes lebih efektif digunakan untuk mengelompokkan data campuran dalam analisis stunting.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim DISCOVER atas dukungannya dalam penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing, Program Studi Sains Data, serta Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan bimbingan, fasilitas, dan arahan selama proses penelitian ini berlangsung.

#### REFERENSI

- [1] Kemenkes RI, *Profil Kesehatan Indonesia 2021*. 2022.
- [2] Kemenkes RI, “SSGI 2024 DALAM ANGKA”.
- [3] WHO, “Global Nutrition Targets 2025 Stunting Policy Brief,” *Art Revising*

- Poet. 21 U.S. Poet. their Draft. Craft, Process*, no. 9, pp. 119–123, 2023, doi: 10.5040/9781350289291.ch-017.
- [4] B. P. Statistik, “Indikator Kesejahteraan Rakyat 2024,” *Badan Pus. Stat.*, vol. 53, p. xxii+197, 2024.
- [5] Alia Fadilah, Mohammad Nurfaizy and dan S. D. Supriyanto Lumbanbatu, “Pengelompokan kabupaten/kota di indonesia berdasarkan faktor penyebab stunting pada balita menggunakan algoritma k-means,” vol. 6, no. 2, pp. 223–230, 2022.
- [6] A. Aditya, B. N. Sari, T. N. Padilah, C. A. Aditya, B. N. Sari, and T. N. Padilah, “Perbandingan pengukuran jarak Euclidean dan Gower pada klaster k-medoids cluster,” vol. 9, no. October 2020, pp. 1–7, 2021, doi: 10.14710/jtsiskom.2021.13747.
- [7] Z. Rais, S. Annas, and M. Refaldy, “Algoritma K-Prototype dalam Pengelompokan Kabupaten / Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020,” vol. 6, no. 3, pp. 144–151, 2024, doi: 10.35580/variasiunm20.
- [8] A. Febriyanti and E. R. Isaura, “Hubungan antara Ketahanan Pangan Rumah Tangga , dengan Kejadian Stunting pada Balita Usia 24-59 Bulan Relationship between Household Food Security and Stunting in Children Age 24-59 Months,” pp. 7–12, 2022.
- [9] I. B. Zakiyah, “Prevalensi dan Determinan Kejadian Stunting pada Balita,” vol. 18, no. 2, pp. 113–120, 2023.
- [10] E. D. Purwanti, S. Masitoh, and S. Ronoatmodjo, “Association Between Basic Immunization Status and Stunting in Toddlers Aged 12-59 Months in Indonesia,” pp. 298–306, 2025.
- [11] A. F. H. Marsandy, M. N. Hayati, and M. Fauziyah, “Klasterisasi Prevalensi Stunting Menggunakan K-Prototype pada Data Campuran,” *MetikJ.*, vol. 8, no. 2, pp. 48–54, 2024, doi: 10.47002/metik.v8i2.824.
- [12] D. Satriawan and D. A. Styawan, “Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki,” vol. 5, no. 1, pp. 61–70, 2021.