ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 14, Nomor 2, Tahun 2025, Halaman 335 - 344
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PERBANDINGAN METODE OPTIMASI *SILHOUETTE, ELBOW,* DAN *GAP STATISTICS* DALAM MENENTUKAN BANYAKNYA *CLUSTER* TERBAIK PADA ANALISIS *K-MEANS CLUSTERING*

(Studi Kasus : Pengelompokan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita *Stunting* Tahun 2021)

Metalia Widya Diantika[1*], Agus Rusgiyono[2], Bagus Arya Saputra[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail: metaliawidya2@gmail.com**

DOI: 10.14710/j.gauss.14.2.335-344

Article Info:
Received: 2024-10-27
Accepted: 2025-10-13
Available Online: 2024-10-14

Keywords:
Stunting; Clusterin; K-Means; Silhouette; Elbow; Gap Statistics

Abstract: Stunting is a condition of malnutrition status that is chronic in growth and development from the beginning of life, malnutrition puts children at greater risk of death. One of the efforts to overcome stunting is to determine in advance the provinces that need to be prioritized in handling the factors that cause stunting by grouping 34 provinces in Indonesia. This study uses k-means clustering to partition data according to their respective characteristics into the form of two or more clusters, determining the optimal number of clusters through elbow optimization methods, gap statistics and silhouette. The method used to test the best cluster results is the Davies Bouldin Index (DBI) method. The results of the elbow method clustering test produce K = 3 with a DBI value of 0.6392677, the gap statistics method produces K = 1 without DBI testing because only 1 cluster is formed, while the silhouette method produces K = 2 with a DBI value of 0.2116945. This shows that the results of clustering k-means with the silhouette method produce better cluster quality because it has a lower DBI value than other methods.

1. PENDAHULUAN

Stunting adalah suatu kondisi status gizi kurang yang bersifat kronik pada masa pertumbuhan dan perkembangan sejak awal kehidupan (Margawati dan Astuti, 2018). UNICEF (2020) menjelaskan hampir separuh dari semua kematian pada anak balita disebabkan oleh kekurangan gizi. Hal ini dikarenakan kekurangan gizi menempatkan anak-anak pada risiko yang lebih besar untuk meninggal akibat infeksi umum, meningkatkan frekuensi dan keparahan infeksi, dan menunda pemulihan infeksi tersebut. Salah satu upaya dalam mengatasi *stunting* adalah menentukan terlebih dahulu provinsi yang perlu diprioritaskan penanganan faktor penyebab *stunting*nya dengan cara mengelompokan 34 provinsi di Indonesia tahun 2021.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Blake *et al* (2016) dengan judul *LBW and SGA Impact Longitudinal Growth and Nutritional Status of Filipino Infants*, diperoleh hasil bahwa terdapat relasi yang relevan antara bayi berat lahir rendah (BBLR) dengan berat di bawah 2,5 kg dan *stunting*. Arifin *et al* (2012), berpendapat bahwa BBLR, asupan gizi buruk, sanitasi tidak memadai, dan infeksi semasa pertumbuhan memicu tumbuh kembang terhambat dan melahirkan anak yang *stunting*. Faktor *stunting* meliputi pemberian ASI, pola pangan anak, infeksi, asupan dan suplai pangan, sanitasi serta kesehatan lingkungan.

Clustering adalah metode atau algoritma yang digunakan untuk menemukan pengelompokan secara alami sesuai variabel dalam kumpulan data yang telah ditentukan

sebelumnya (Malik dan Tuckfield, 2019). *K-Means clustering* merupakan salah satu metode pengelompokan dengan mempartisi data sesuai karakteristiknya masing-masing, ke dalam bentuk dua atau lebih kelompok. *K-Means clustering* merupakan salah satu jenis "unsupervised machine learning algorithms" yang paling sederhana.

Banyaknya *cluster* optimum pada pengelompokan data dapat diketahui melalui metode optimasi *elbow*, *gap* statistics, dan *silhouette*. Metode *elbow* menentukan banyaknya *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara banyaknya *cluster* yang akan membentuk siku pada suatu titik. *Gap Statistics* menentukan banyaknya *cluster* optimal secara lebih konstan dibandingkan pengukuran lainnya, sedangkan melalui metode s*ilhouette* dapat melihat kualitas dan kekuatan *cluster* dalam penentuan banyaknya *cluster* terbaik, serta seberapa baik atau buruknya suatu objek ditempatkan dalam suatu *cluster*.

Pada penelitian sebelumnya yang dilakukan oleh Helilintar dan Farida (2018) dengan judul Penerapan Algoritma *K-Means Clustering* untuk Prediksi Prestasi Nilai Akademik Mahasiswa, membahas mengenai pengelompokan predikat kelulusan mahasiswa menggunakan selisih *euclidean* dengan metode *k-means clustering*. Hasil penelitian tersebut, dipeoleh 4 kelompok predikat kelulusan mahasiswa yaitu sangat baik, baik, cukup, dan kurang.

Berdasarkan hal tersebut, penelitian ini menggunakan metode *k-means clustering* untuk mengelompokan 34 provinsi di Indonesia tahun 2021 berdasarkan faktor penyebab *stunting* di wilayah tersebut. Banyaknya *cluster* terbaik pada penelitian ini, ditentukan melalui perbandingan metode optimasi *elbow*, *gap statistics*, dan *silhouette*. Tujuan dari penelitian ini adalah memperoleh hasil *clustering* terbaik untuk menentukan provinsi yang perlu diprioritaskan penanganan faktor penyebab *stunting*nya, sebagai upaya mengatasi tingkat *stunting* di wilayah tersebut.

2. TINJAUAN PUSTAKA

Analisis *cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya (Awalluddin & Taufik, 2017). Analisis *cluster* memiliki asumsi yang perlu dipenuhi yakni :

a. Representatif

Pengujian (sampel mewakili populasi) representatif bisa melalui uji *Kaiser Meyer Olkin* (KMO). Menurut Widarjono (2010), rumus KMO yakni:

$$\text{KMO} = \frac{\sum_{j=1}^{p}\sum_{l=1,j\neq l}^{p} r_{jl}^2}{\sum_{j=1}^{p}\sum_{l=1,j\neq l}^{p} r_{jl}^2 + \sum_{j=1}^{p}\sum_{l=1,j\neq l}^{p} a_{jl,m}^2} \quad (1)$$

dengan p banyaknya variabel, n banyaknya objek, x_{ij} objek ke-i dengan variabel ke-j , x_{il} objek ke-i dengan variabel ke-l, r_{jl} koefisien *pearson correlation* variabel j dengan l, $a_{jl,m}$ koefisien *partial correlation* variabel j dengan l serta variabel m konstan

b. Multikolinieritas

Multikolinearitas adalah adanya hubungan linear atau korelasi yang tinggi antar variabel. Menurut Gujarati (2009), cara untuk mengetahui ada tidaknya multikolinearitas melalui perhitungan nilai (VIF) *Variance Inflation Factor* dengan rumus:

$$\text{VIF} = \frac{1}{(1-R^2)} \quad (2)$$

R^2 adalah koefisien determinasi variabel terikat dengan variabel bebasnya. Batas dari nilai VIF yaitu 10. Apabila VIF lebih rendah dari 10 maka tiada multikolinearitas (Gujarati, 2009).

Analisis *clustering* mengelompokan data sesuai dengan kemiripan atau ukuran selisih untuk mengetahui kemiripan antar objek. Salah satu ukuran selisih yang digunakan dalam

penelitian ini adalah selisih *euclidean*. Selisih *euclidean* antara dua titik adalah akar dari jumlah kuadrat selisih antar objek. Berikut adalah rumus untuk menghitung selisih *euclidean* (Johnson dan Winchern, 2007):

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{p}(x_{i,j} - c_{k,j})^2} \quad (3)$$

dengan,

k : 1,2, … , q

$d_{(\text{xi,Ck})}$: selisih *euclidean* objek ke-*i* variabel ke-*j,* dengan pusat *cluster* ke-*k* variabel ke-*j*

$x_{\text{i,j}}$: nilai objek ke-*i,* variabel ke-*j*

$c_{k,j}$: pusat *cluster(centroid)* ke-*k* variabel ke-*j*

p : jumlah variabel yang diamati

q : *cluster* terbesar

Metode *k-means* berusaha mengelompokkan data yang ada ke dalam satu kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain (Helilintar dan Farida, 2018). Dasar algoritma *k-means* sebagai berikut:

1. Menentukan jumlah *K-cluster* yang akan dibentuk
2. Menentukan pusat *cluster* (*centroid*) awal secara acak.
3. Menghitung selisih setiap objek dengan setiap *centroid* menggunakan selisih *Euclidean*

$$d(x_i, c_k) = \sqrt{\sum_{j=1}^{p}(x_{i,j} - c_{k,j})^2}$$

4. Mengelompokkan masing-masing objek ke dalam *centroid* yang paling dekat. Suatu objek akan menjadi anggota dari *cluster ke-k* apabila selisih objek tersebut ke *centroid ke-k* bernilai paling kecil jika dibandingkan dengan selisih ke *centroid* lainnya.
5. Menentukan *centroid* yang baru dengan cara menghitung rata-rata dari objek yang ada pada masing-masing *cluster* dengan persamaan berikut:

$$\text{C}_{\text{k,j}} = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{i,j} \quad (4)$$

dengan,

k : *1,2,3,…,q*

j : *1,2,3,…,p*

$C_{k,j}$: *centroid cluster* ke-*k* variabel ke-*j*

$\text{x}_{i,j}$: nilai pada objek ke-*i* pada variabel ke-*j*

n_k : banyaknya objek pada *cluster* ke-*k*

6. Mengulangi kembali langkah 3-5 hingga anggota tiap *cluster* tidak ada yang berubah.

Ada beberapa cara untuk menetapkan banyaknya *cluster* optimal, di bawah ini merupakan beberapa teknik optimasi dalam penentuan banyaknya *cluster*.

a. **Elbow**

Identifikasi *cluster* suatu data bertujuan untuk meminimalkan selisih antar titik *cluster*. Metode *elbow* adalah teknik yang digunakan dalam menentukan jumlah *cluster* optimal melalui memperhitungkan persentase perbandingan jumlah *cluster* dengan bentuk siku di sebuah titik (Madhulatha, 2012). Penentuan nilai *cluster* terbaik dengan metode *elbow* diperoleh melalui perhitungan nilai sum square error (SSE) terendah pada data, dengan rumus sebagai berikut:

$$\text{SSE} = \sum_{k=1}^{k}\sum_{x_i \in S_k}(x_i - c_k)^2 \quad (5)$$

SSE menunjukkan nilai *sum square error,* k adalah banyaknya *cluster*, x_i objek ke-i anggota *cluster* ke- k, S_k merupakan himpunan objek-objek *cluster* ke-*k,* dan c_k yaitu pusat (centroid) *cluster* ke-k

b. Gap Statistics

Gap Statistics adalah salah satu metode yang efektif untuk menemukan banyaknya *cluster* optimal dalam kumpulan data, dengan tujuan menentukan banyaknya *cluster* lebih konstan dibandingkan pengukuran lainnya. Metode g*ap statistics* membandingkan nilai kurva log W_k antara *cluster* yang terbentuk melalui data observasi dan data referensi dengan distribusi *uniform.* Distribusi *uniform* adalah distribusi yang peluang setiap peubah acaknya sama. Selisih objek berpasangan dalam *cluster* dihitung dengan rumus:

$$D_r = \sum_{i,j \in C_r} d_{i,j} \quad (6)$$

dengan,

$d_{i,j}$ = selisih euclidean kuadrat antara objek ke-i dan ke-j

D_r = total selisih *euclidean* data pada cluster ke-r

r = banyaknya cluster

kemudian menghitung jumlah kuadrat di dalam *cluster* (W_k) dengan n banyaknya objek, menggunakan rumus:

$$W_k = \sum_{r=1}^{k} \frac{1}{2n_r} D_r \quad (7)$$

Nilai gap didapatkan dengan menghitung selisih pendekatan standarisasi W_k data referensi berdistribusi *uniform* dan data observasi.

$$Gap_n(k) = E_n^*\{Log(W_k)\} - Log\ (W_k) \quad (8)$$

dengan $E_n^*\{Log(W_k)\}$ merupakan nilai ekspektasi banyaknya *resampling* (B) dari hasil Log (W_{kb}^*) pada data referensi, dengan rumus:

$$E_n^*\{Log(W_k)\} = \left[\frac{1}{B}\right] \sum_b Log(W_{kb}^*) \quad (9)$$

b = 1, 2, 3, … B

B = banyaknya *resampling* data distribusi *uniform* terbesar

k = 1, 2, 3, …, K

K = banyaknya cluster terbesar

Kriteria jumlah *cluster* terbaik ditetapkan sesuai nilai *gap statistics* tertinggi atau *gap statistics* pertama yang menunjukkan peningkatan *gap* yang minimum apabila *gap* semakin meningkat (Silvi, 2018).

c. Silhouette

Nilai perhitungan *silhouette* berkisar antara 1 hingga -1. Apabila nilainya antara 0 dan -1, berarti *cluster* tersebar atau selisih antar titik *cluster* besar. Jika nilai *silhouette* suatu *cluster* besar atau mendekati 1, berarti selisih titik *cluster* kecil dan selisih dengan titik *cluster* lainnya besar. Dengan demikian, nilai *silhouette* ideal sebuah *cluster* ketika mendekati satu (Malik & Tuckfield, 2019). Jumlah *cluster* terbaik dipilih grafik *silhouette* dengan nilai k terbesar. Perhitungan nilai k menggunakan metode *silhouette index* yaitu:

1. Menghitung rata-rata selisih objek ke-i dengan semua objek dalam satu *cluster*

$$a(i) = \frac{1}{n_k - 1} \sum_{h \in Cl_k, h \neq i} d(i,h) \quad (10)$$

dengan,

$a(i)$: rata-rata selisih objek ke-*i* dengan seluruh objek dalam satu *cluster*

h : objek lain dalam satu *cluster* ke-*k*

d(*i,h*) : selisih antara objek *i* dan *h*

n_k : banyaknya objek pada *cluster* ke-*k*

Cl_k : himpunan objek-objek *cluster* ke-*k*

2. Menghitung rata-rata selisih objek ke-*i* dengan semua objek yang berada pada *cluster* lain

$$d(i,v) = \frac{1}{n_v} \sum_{h \in Cl_v, h \neq i} d(i,h) \quad (11)$$

dengan d(i,v) adalah selisih rata-rata objek i dengan semua objek pada *cluster* lain ke-v, nv adalah banyaknya objek pada *cluster* ke-v, dan Cl_v yaitu himpunan objek-objek *cluster* ke-v dimana $k \neq v$

3. Menentukan nilai minimumnya yaitu b(i) yang menunjukkan perbedaan rata-rata objek i untuk *cluster* terdekat dengan tetangganya

$$b(i) = \min d(i,v) \quad (12)$$

dengan $b(i)$ adalah nilai minimum dari selisih rata-rata objek i dengan semua objek pada *cluster* lain ke-v

4. Menghitung nilai *silhouette*

$$S(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (13)$$

S(i) adalah nilai s*ilhouette, b(i)* nilai minimum dari selisih rata-rata objek i dengan semua objek antar cluster, dan a(i) rata-rata selisih objek ke-i dengan seluruh objek dalam satu *cluster*. Selanjutnya dapat menghitung koefisien silhouette yang didefinisikan sebagai rata-rata s(i), sebagai berikut:

$$SC = \frac{1}{n}\sum_{i=1}^{n} s(i) \quad (14)$$

dengan n adalah banyak pengamatan, dan SC adalah nilai *Silhouette Coefficient.*

Davies Bouldin Index (DBI) ialah metode pengukuran evaluasi *cluster* pada suatu pengelompokan. Perhitungan DBI diuji dari segi nilai *kohesi* dan *separasi*. Nilai k*ohesi* didefinisikan sebagai nilai kedekatan data suatu *cluster* dengan *centroid* pada *cluster*nya. Sedangkan nilai *separasi* diartikan sebagai selisih antar *centroid* dari *cluster* yang lainnya. Perhitungan validasi *Davies Bouldin Index* meliputi (Kartikasari, 2021):

1. Menghitung nilai Sv untuk mencari matrik *kohesi* pada masing-masing *cluster* dengan persamaan :

$$S_k = \frac{1}{n_k}\sum_{i=1}^{n_k} d(x_i, C_k) \quad (15)$$

dengan,
s_k : rata-rata dari selisih objek ke-i dengan *centroid cluster* ke-k
k : 1,2, … , q
q : *cluster* terbesar
n_k : banyaknya objek pada *cluster* ke-k
c_k : pusat *cluster* ke-k
$d_{(xi,Ck)}$: selisih setiap objek ke-i ke *centroid* yang berada dalam *cluster* ke-k

2. Menghitung nilai Mk,v untuk mengetahui *separasi* atau selisih antar *cluster*, dengan cara mengukur selisih *centroid* dalam suatu *cluster* dengan *centroid* yang berada dalam *cluster* lain. Berikut persamaannya:

$$M_{k,v} = d(C_{k,}C_v) \text{ , } k \neq v \quad (16)$$

dengan,
$M_{k,v}$: selisih pusat *cluster* ke-k dengan *pusat cluster* ke-v
C_k : pusat *cluster* ke-k
C_v : pusat *cluster* ke-v
$d_{(ck,Cv)}$: selisih *centroid* ke-k dan *centroid* ke-v

3. Menghitung rasio agar diperoleh perbandingan antara *cluster* ke-k dengan *cluster* ke-v

$$R_{k,v} = \frac{S_k + S_v}{M_{k,v}} \quad (17)$$

4. Menghitung ukuran kesamaan *cluster* maksimum

$$R_k = \max_{k \neq v}(R_{k,v}) \quad (18)$$

5. Menghitung nilai *DBI*

$$DBI = \frac{1}{K}\sum_{k=1}^{K} R_k \quad (19)$$

Persamaan di atas menunjukkan K sebagai banyaknya *cluster*. Apabila nilai DBI yang didapatkan semakin rendah, menunjukkan semakin optimal banyaknya *cluster* yang didapatkan.

Profilisasi hasil *clustering* optimal memaparkan karakteristik tiap-tiap, maka dapat diamati kecondongan masing-masing *cluster.* Karakteristik *cluster* dapat dilihat dengan cara menghitung rata-an dari anggota setiap variabel pada penelitian.

3. METODE PENELITIAN

Data dalam penelitian tugas akhir ini merupakan data sekunder faktor *stunting* pada balita di Indonesia berdasarkan 34 provinsi tahun 2021, data didapatkan dari Buku Profil Kesehatan Indonesia Tahun 2021. Data faktor *stunting* pada balita meliputi lima variabel, yaitu: sanitasi tidak layak, balita gizi buruk, kecukupan dokter, berat badan bayi lahir rendah, status gizi pendek. Analisis data pada penelitian ini menggunakan *software* R, dengan tahapan sebagai berikut:

1. *Input* data faktor penyebab balita *stunting*
2. Pengujian asumsi representatif menggunakan uji KMO
3. Pengujian asumsi multikolinieritas menggunakan nilai VIF. Jika terdapat multikolinieritas dapat dilakukan PCA, hasil perhitungan nilai PCA digunakan sebagai pengganti nilai data sebelumnya.
4. *Clustering* menggunakan metode *k-means* melalui perhitungan selisih *euclidean* serta mengelompokkan objek pada masing-masing *cluster*.
5. Menentukan jumlah *cluster* melalui metode *elbow, gap statistics,* dan *silhouette*
6. Menghitung nilai koefisien masing-masing *cluster* menggunakan validasi *davies bouldin index*
   a. Menghitung rata-rata selisih objek dengan *centroid cluster* yang diikuti
   b. Menghitung selisih *centroid* dalam suatu *cluster* dengan *centroid* yang berada dalam *cluster* lain
   c. Menghitung rasio sebagai perbandingan *cluster* ke-*k* dengan *cluster* ke-*v*
   d. Menentukan rasio antar *cluster* yang memiliki nilai maksimum
7. Interpretasi hasil pengelompokan data berdasarkan karakteristik masing-masing klaster yang terbentuk. Nilai *davies bouldin index* terendah akan dipilih sebagai jumlah *cluster* yang optimal.

4. HASIL DAN PEMBAHASAN

Uji asumsi sampel mewakili populasi menggunakan KMO dilakukan untuk mengetahui syarat kecukupan suatu sampel. Hasil uji KMO pada data penyebab balita *stunting* ditunjukkan pada tabel di bawah:

Tabel 1. Hasil Nilai KMO

Variabel	Hasil Nilai KMO
Sanitasi tidak layak	0,59
Balita gizi buruk (BB/TB)	0,74
Kecukupan dokter	0,66
Berat badan bayi lahir rendah	0,60
Status gizi pendek (TB/U)	0,68

Data penelitian yang telah mewakili populasi dapat dianalisis pada tahap selanjutnya dengan pengujian multikolinieritas, agar dapat mendeteksi hubungan linier antar variabel. Hasil pengolahan data pada penelitian ini, diperoleh nilai VIF lebih rendah dari 10. Maka dapat disimpulkan bahwa masing-masing variabel tidak ada hubungan multikolinieritas.

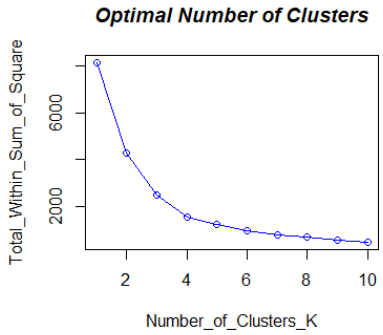
Tabel 2. Hasil Nilai VIF

Variabel	Hasil Nilai VIF
Sanitasi tidak layak	2,035156
Balita gizi buruk (BB/TB)	1,475299
Kecukupan dokter	2,193533
Berat badan bayi lahir rendah	1,564707
Status gizi pendek (TB/U)	2,096464

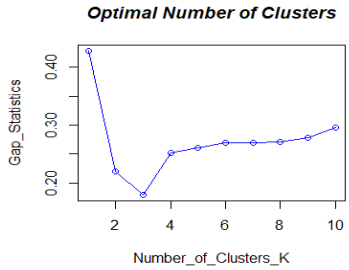
Penelitian ini menggunakan metode *k-means* dan ukuran selisih *euclidean* untuk mengelompokan data berdasarkan ukuran *cluster* yang telah ditentukan sebelumnya, yaitu K= 1, K=2, hingga K=10. Perhitungan manual dari perhitungan *k-means* yaitu:

1. Tentukan nilai K *cluster* yang terbentuk
2. Menetapkan pusat *cluster* awal secara acak.
3. Hitung selisih tiap objek dengan setiap pusat *cluster* menggunakan selisih *euclidean.*
4. Mengelompokkan setiap objek pada pusat *cluster* terdekat. Data dapat menjadi bagian dari *cluster* ke-k jika selisih data ke *centroid* ke-k lebih rendah daripada selisih dengan *centroid* yang lain
5. Menetapkan pusat *cluster* baru menggunakan perhitungan rata-rata dari objek pada masing-masing *cluster*
6. Hitung selisih masing-masing objek terhadap setiap *centroid* yang baru menggunakan selisih *euclidean* untuk iterasi ke-2
7. Mengelompokkan setiap objek pada *centroid* terdekat.
8. Karena masih terdapat objek yang berpindah *cluster* dengan selisih *euclidean* pada proses iterasi, maka selanjutnya mengulangi kembali langkah perhitungan nomor 5 sampai 7. Perhitungan berhenti pada saat tidak ada terdapat perubahan anggota tiap *cluster*.

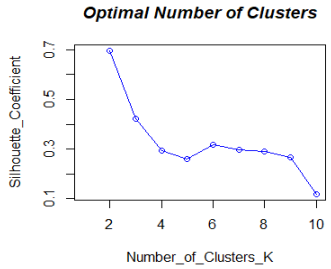
Tahap selanjutnya adalah penentuan nilai *cluster* terbaik menggunakan metode *elbow, gap statistics* dan *silhouette.* Setelah dilakukan pengolahan data, diperoleh nilai K terbaik menggunakan metode *elbow* pada K=3, metode *gap statistics* pada K=1, dan metode *silhouette* pada K=2. Hasil grafik penentuan nilai K menggunakan ketiga metode tersebut dapat diamati pada di bawah:



Gambar 1. *Output* Metode *Elbow*



Gambar 2. *Output* Metode *Gap Statistics*



Gambar 3. *Output* Metode *Silhouette*

Setelah diperoleh banyaknya *cluster* terbaik dari masing-masing metode optimasi, dilanjutkan proses validasi pada *cluster* dengan perhitungan DBI dalam menentukan banyaknya *cluster* terbaik. Hasil uji *clustering* dengan metode *elbow* diperoleh DBI= 0,6392677, dengan *gap statistics* tidak dilakukan uji DBI dikarenakan hanya terdapat satu *cluster*, sedangkan dengan metode *silhouette* didapatkan nilai DBI= 0,2116945.

Berdasarkan hasil pengujian di atas, dapat ditentukan bahwa *clustering* dengan metode *silhouette* memiliki kualitas *cluster* yang lebih baik dikarenakan nilai DBI *silhouette* lebih rendah dibandingkan metode lainnya. Jumlah *cluster* optimal dengan K=2, akan dilakukan tahap profilisasi untuk melihat karakteristik dari masing-masing *cluster*. Karakteristik dari setiap *cluster dapat* direpresentasikan dengan melihat rata-rata dari anggota masing-masing variabel. Berikut adalah rata-rata tiap variabel dalam *cluster* yang terbentuk:

Tabel 3. Rata-rata Variabel Faktor Penyebab Balita *Stunting*

Variabel	Cluster	
	1	2
Sanitasi tidak layak	17.82%	59.19%
Balita gizi buruk (BB/TB)	0.90%	1.70%
Kecukupan dokter	9.46%	49.50%
Berat badan bayi lahir rendah	3.61%	4.80%
Status gizi pendek (TB/U)	8.01%	10.40%

Tabel di atas menunjukkan rata-rata *cluster* faktor penyebab balita *stunting* tertinggi berada pada *cluster* 2. *Cluster* 2 mengartikan bahwa provinsi pada *cluster* 2 memiliki faktor penyebab balita *stunting* lebih tinggi dibandingkan *cluster* 1. *Cluster* 1 terlihat memiliki rata-rata *cluster* lebih kecil daripada *cluster* 2. Hal tersebut mengartikan bahwa anggota *cluster* 1 merupakan provinsi dengan faktor penyebab balita *stunting* yang rendah. Berdasarkan hal tersebut dapat diartikan bahwa Papua merupakan provinsi yang perlu mendapat perhatian oleh pemerintah Indonesia, karena memiliki faktor penyebab balita *stunting* tinggi, terutama pada faktor sanitasi tidak layak, balita dengan gizi buruk, kecukupan dokter, BBLR, serta status balita dengan gizi pendek.

5. KESIMPULAN

Hasil perhitungan dan analisis di atas, dapat disimpulkan sebagai berikut:

1. Pengelompokan metode *k-means* dari 34 provinsi yaitu:
   a. Pada K=1 diperoleh anggota *cluster* meliputi seluruh provinsi terdiri dari 34 provinsi
   b. Pada K=2 diperoleh anggota pada *cluster* ke-1 yaitu 33 provinsi dan *cluster* ke-2 yaitu 1 provinsi
   c. Pada K=3 diperoleh anggota *cluster* ke-1 sejumlah 1 provinsi, *cluster* ke-2 sejumlah 9 provinsi dan *cluster* ke-3 terdiri dari 24 provinsi
2. Hasil *clustering* faktor penyebab balita *stunting* di Indonesia dengan metode *k-means* diperoleh jumlah *cluster* optimal yaitu pada K=2. Hal tersebut dapat diamati dari hasil validasi DBI= 0,2116945 yang merupakan hasil DBI terendah. Pengelompokan tersebut memberikan hasil bahwa *cluster* ke-1 terdiri beranggotakan 33 provinsi sedangkan *cluster* ke-2 beranggotakan 1 provinsi.
3. Hasil profilisasi *cluster* menunjukkan pada jumlah 2 *cluster* diperoleh bahwa rata-rata *cluster* dengan faktor penyebab balita *stunting* tertinggi berada pada *cluster* 2, hal tersebut mengartikan bahwa anggota *cluster* 2 merupakan provinsi dengan faktor penyebab balita *stunting* yang tinggi. *Cluster* 1 memiliki rata-rata *cluster* lebih rendah daripada *cluster* 2, hal tersebut mengartikan bahwa anggota *cluster* 1 merupakan provinsi dengan faktor penyebab balita *stunting* yang rendah. Oleh karena itu, diharapkan kepada pemerintah di Indonesia untuk lebih memperhatikan provinsi pada *cluster* 2 yaitu Papua yang merupakan provinsi dengan rata-rata faktor penyebab balita *stunting* yang tingi dibandingkan *cluster* 1, sehingga pada provinsi tersebut dapat menekan faktor-faktor penyebab balita *stunting*.

DAFTAR PUSTAKA

Arifin, D. Z., Irdasari, S. Y., & Sukandar, H. 2012. *Analisis Sebaran dan Faktor Risiko Stunting pada Balita di Kabupaten Purwakarta*. Jurnal Pustaka Unpad, 1–9.

Awalluddin, A., & Taufik, I. 2017. *Analisis Cluster Data Longitudinal pada Pengelompokan Daerah Berdasarkan Indikator IPM di Jawa Barat*. Prosiding Seminar Nasional Metode Kuantitatif, 187–194

Blake, R. A., *et al.* 2016 . *LBW And SGA Impact Longitudinal Growth And Nutritional Status Of Filipino Infants*. *PLoS ONE* Vol.11, No. 7: Hal. 1–13.

Gujarati, D. 2009. *Dasar-dasar Ekonometrika Jilid 2.* Jakarta: Erlangga.

Helilintar, R., & Farida, I. N. 2018. *Penerapan Algoritma K-Means Clustering untuk Prediksi Prestasi Nilai Akademik Mahasiswa*. Jurnal Sains dan Informatika Vol. 4, No.2: Hal. 80–87.

Johnson, W.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th Edition. Pearson Prentice Hall: New Jersey.

Kartikasari, M. D. 2021. *Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan*. Jambura Journal of Mathematics Vol. 3, No. 2: Hal. 187–196.

Kemenkes RI. 2021. *Profil Kesehatan Indonesia*. Jakarta : Kementerian Kesehatan Republik Indonesia.

Madhulatha, T.S. 2012. *An Overview On Clustering Methods*. Journal of Engineering Vol. 2, No. 4, Hal 719-725.

Malik, A., & Tuckfield, B. 2019. *Applied Unsupervised Learning with R*. Birmingham: Packt Publishing.

Margawati, A., & Astuti, A. M. 2018. *Pengetahuan ibu, pola makan dan status gizi pada anak stunting usia 1-5 tahun di Kelurahan Bangetayu, Kecamatan Genuk, Semarang*. Jurnal Gizi Indonesia (The Indonesian Journal of Nutrition) Vol. 6, No.2: Hal. 82–89.

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 Vol. *8, No.*2: Hal. 161–170.

Sari, D. N. P., & Sukestiyarno, Y. L. 2021 . *Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia*. PRISMA, Prosiding Seminar Nasional Matematika, 4, 602–610.

Sari, R. K., Yugiana, E., & Ponco, S. H. 2021. *Profil Statistik Kesehatan 2021*. Jakarta : Badan Pusat Statistik.

Silvi, R. 2018. *Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokkan Indikator HIV/AIDS di Indonesia.* Jurnal Matematika "MANTIK," Vol. 4, No.1: Hal. 22–31.

UNICEF. 2020. *Child Malnutrition*. https://data-uniceforg.translate.goog/topic/nutrition/malnutrition/. Diakses : 2 Maret 2023

Widarjano, A. 2010. *Analisis Stastistika Multivariat Terapan*. UPP STIM YKPN. Yogyakarta

Wulandari Leksono, *et al*. 2021. *Risiko Penyebab Kejadian Stunting pada Anak*. Jurnal Pengabdian Kesehatan Masyarakat: Pengmaskesmas Vol. 1, No. 2: Hal. 34–38