



Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesehatan Lingkungan Menggunakan Metode *K-Harmonic Means* dengan Optimasi *Particle Swarm Optimization*

Nida Faoziatun Khusna¹, Rochdi Wasono², Ariska Fitriyana Ningrum³

^{1,2}Program Studi Statistika, Universitas Muhammadiyah Semarang

¹nidafaok@gmail.com

²rochdi@unimus.ac.id

³Program Studi Sains Data, Universitas Muhammadiyah Semarang

³ariskafitriyana@unimus.ac.id

Corresponding author email: nidafaok@gmail.com

Abstract: The environment plays a crucial role in achieving optimal public health, with environmental health indicators serving as key factors in disease prevention and quality of life improvement. This study aims to cluster provinces in Indonesia based on the 2023 Environmental Health Indicators using two methods: *K-Harmonic Means (KHM)* and *Particle Swarm Optimization K-Harmonic Means (PSOKHM)*. The clustering results were validated using the *Silhouette Index*, *Dunn Index*, and *Connectivity Index* to determine the optimal parameters for each cluster. The standard deviation ratio was then calculated to determine the optimal number of clusters. The KHM method produced 4 optimal clusters with a parameter value of 4, while the PSOKHM method also produced 4 optimal clusters with the same parameter value. A comparison of the standard deviation ratios showed that PSOKHM had the lowest ratio of 0.1829759, indicating it as the more optimal method compared to KHM. Based on the interpretation and profiling of the clusters, there is a variation in environmental health conditions across Indonesia. Cluster 1 shows good water and sanitation supervision but needs improvement in habitable housing; cluster 2 has strong healthy living policies but sanitation issues; cluster 3 requires improved facility supervision; and cluster 4 has low access to water and sanitation. This study can serve as a foundation for data-driven policy formulation by directing interventions according to the characteristics of each province to improve environmental health quality nationwide.

Keywords: Clustering, Environmental Health, *K-Harmonic Means*, *Particle Swarm Optimization*

Abstrak: Lingkungan memiliki peran penting dalam mencapai kesehatan masyarakat yang optimal, dengan indikator kesehatan lingkungan sebagai faktor penentu dalam pencegahan penyakit dan peningkatan kualitas hidup. Penelitian ini bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan Indikator Kesehatan Lingkungan tahun 2023 menggunakan dua metode, yaitu *K-Harmonic Means (KHM)* dan metode *Particle Swarm Optimization K-Harmonic Means (PSOKHM)*. Hasil *cluster* divalidasi menggunakan *Silhouette Index*, *Dunn Index*, dan *Connectivity Index* untuk menentukan nilai parameter yang optimal setiap *cluster*. Kemudian, menghitung rasio simpangan baku untuk menentukan banyaknya *cluster* yang optimal. Metode KHM menghasilkan *cluster* optimal yaitu 4 *cluster* dengan nilai parameter yaitu 4, sedangkan PSOKHM *cluster* optimal yaitu 4 *cluster* dengan nilai parameter yaitu 4. Hasil perbandingan rasio simpangan baku menunjukkan bahwa PSOKHM memiliki nilai rasio terkecil yaitu 0.1829759, menjadikannya metode yang lebih optimal dibandingkan KHM. Berdasarkan hasil interpretasi terhadap karakteristik dan profilisasi *cluster*, terdapat variasi kondisi kesehatan lingkungan di Indonesia, dengan *cluster* 1 baik pengawasan air dan sanitasi tetapi perlu perbaikan rumah layak huni, *cluster* 2 kuat kebijakan hidup sehat tetapi ada masalah sanitasi, *cluster* 3 perlu tingkatkan pengawasan fasilitas dan *cluster* 4 memiliki akses air dan sanitasi yang masih rendah. Penelitian ini dapat menjadi dasar dalam perumusan kebijakan berbasis data yang lebih tepat sasaran, dengan mengarahkan intervensi sesuai karakteristik masing-masing provinsi untuk meningkatkan kualitas kesehatan lingkungan secara nasional.

Kata kunci: Kesehatan Lingkungan, *K-Harmonic Means*, *Particle Swarm Optimization*, Pengelompokan

I. PENDAHULUAN

Lingkungan mempunyai peran penting guna mencapai kesehatan masyarakat yang baik, bersama dengan faktor lainnya, misalnya mutu layanan kesehatan dan sikap publik [1]. Syarat lingkungan yang sehat meliputi stabilnya iklim, pasokan air yang mencukupi, sanitasi dan kebersihan, pemakaian senyawa kimia yang memenuhi unsur keamanan, terlindungi dari radiasi, lokasi bekerja yang aman dan sehat, implementasi pertanian yang layak, kawasan perkotaan yang memperkuat kesehatan dan pembangunan lingkungan, serta konservasi alam (WHO) (2024). Kondisi dan pengelolaan kesehatan



lingkungan mempunyai keterkaitan kuat dengan kesehatan masyarakat, dikarenakan kondisi kesehatan lingkungan dapat menimbulkan penyakit menular berbasis lingkungan. Salah satu penyebab terbanyak kematian di Indonesia yaitu penyakit menular berbasis lingkungan.

Upaya kesehatan lingkungan di Indonesia masih belum berhasil mencapai hasil yang memuaskan. Hal ini tercermin dari sembilan Indikator Kesehatan Lingkungan pada publikasi Profil Kesehatan Indonesia 2023 oleh Kementerian Kesehatan Republik Indonesia yang masih belum mencapai target. Indikator-indikator tersebut meliputi air minum, akses sanitasi layak, Sanitasi Total Berbasis Masyarakat (STBM), Tempat dan Fasilitas Umum (TFU), Tempat Pengelolaan Pangan (TPP), kabupaten/kota sehat, pengelolaan limbah medis, Gerakan Masyarakat Hidup Sehat (GERMAS), serta perumahan. Salah satu indikator yang mempunyai standar yang telah ditetapkan dalam RPJMN (Rencana Pembangunan Jangka Menengah Nasional) dan Rencana Strategis yaitu indikator sarana air minum yang diawasi/diperiksa kualitas air minum yang mempunyai standar sebesar 72% di tahun 2023. Namun, hanya 14 yang memenuhi standar atau diatas 72%, sementara provinsi lainnya masih belum memenuhi standar atau dibawah 72%. Ini menunjukkan bahwa sebagian besar provinsi masih belum mampu mencapai standar kualitas air minum yang diinginkan [3].

Dilihat dari berbagai Indikator Kesehatan Lingkungan setiap provinsi di Indonesia memiliki kondisi kesehatan lingkungan yang beragam. Oleh sebab itu, diperlukan analisis pengelompokan (*clustering*) guna mengelompokkan provinsi-provinsi berdasarkan kesamaan kondisi kesehatan lingkungan sehingga program kesehatan lingkungan di tiap provinsi juga diprioritaskan sesuai dengan kondisi tersebut. Salah satu metode *clustering* yaitu metode nonhierarki dimana jumlah *cluster* yang akan dibentuk telah ditetapkan sebelumnya, seperti *K-Means*. *K-Means* sering digunakan karena memiliki konsep yang sederhana, mudah diimplementasikan, serta efisien dalam mengolah data dalam skala besar [4]. Sebagai metode nonhierarki yang sering digunakan, *K-Means* mengalami berbagai pengembangan, salah satunya yaitu *K-Harmonic Means* [5]. *K-Harmonic Means* adalah metode *optimizing partitioning* dalam metode nonhierarki dan berbasis pusat sama seperti *K-Means*. Metode *K-Harmonic Means* yaitu metode analisis *cluster* yang diperkenalkan pertama kali oleh Zhang et al., (1999) yang kemudian dikembangkan oleh Hamerly & Elkan (2002) untuk menangani masalah inisialisasi pusat cluster yang sensitif pada metode *K-Means*. Proses inisialisasi centroid awal sangat memengaruhi hasil pengelompokan sehingga *K-Means* cenderung konvergen pada *local optima* [8]. Perbedaan antara *K-Means* dan *K-Harmonic Means* ada pada fungsi tujuan yang digunakan. Fungsi tujuan *K-Harmonic Means* dengan pendekatan rata-rata harmonik dari jarak tiap titik data menuju pusat cluster [7].

Algoritma *K-Harmonic Means* efektif dalam menangani masalah inisialisasi, namun masih menghadapi kendala dalam mengatasi *local optima*. Salah satu pendekatan yang sering digunakan untuk menyelesaikan permasalahan tersebut yaitu melalui optimasi. Optimasi yang sering digunakan yaitu *Particle Swarm Optimization* (PSO). PSO yaitu algoritma stokastik yang dipakai guna menemukan solusi optimal dalam masalah numerik. Algoritma ini meniru perilaku kelompok burung, kawanan ikan, atau makhluk hidup lainnya yang menyesuaikan diri dengan lingkungan, mencari makanan, atau menghindari pemangsa dengan menerapkan pendekatan "berbagi informasi" [9]. Algoritma *Particle Swarm Optimization* (PSO) memiliki banyak kelebihan seperti memerlukan parameter yang sedikit, penerapan yang mudah, konvergensi yang cepat dan simpel [10]. Optimasi PSO lebih efisien dalam perhitungan jika dibandingkan dengan optimasi lainnya [11].

Penelitian terdahulu mengenai *K-Harmonic Means* dan *Particle Swarm Optimization K-Harmonic Means* telah dilakukan sebelumnya. Penelitian oleh Christienova et al. (2017) yang berfokus pada perbandingan KHM dan PSOKHM dalam mengelompokkan 143 desa di Kabupaten Lanny Jaya Provinsi Papua dengan banyaknya *cluster* (k) 10 sampai 15 dan nilai parameter (p) yaitu 2, 2.5, 3 dan 3.5. Hasil dari penelitian tersebut yaitu mendapatkan metode dan nilai parameter optimal berdasarkan *Silhouette Index*, *Dunn Index* dan *Connectivity Index* yaitu metode PSOKHM dengan nilai parameter yaitu 3, serta mendapatkan *cluster* optimal berdasarkan nilai rasio simpangan baku terkecil yaitu 10 *cluster*. Hal ini menunjukkan bahwa PSOKHM lebih baik dari KHM dalam mengelompokkan 143 desa di Kabupaten Lanny Jaya Provinsi Papua.

Mengacu pada latar belakang tersebut, penelitian ini bermaksud untuk membandingkan metode *K-Harmonic Means* (KHM) dengan metode *Particle Swarm Optimization K-Harmonic Means*

(PSOKHM) yang akan memvariasikan banyaknya *cluster* (k) dan nilai parameter (p) guna mengelompokkan 34 provinsi di Indonesia menurut Indikator Kesehatan Lingkungan tahun 2023. Pemilihan nilai parameter optimal setiap *cluster* menggunakan *Silhouette Index*, *Dunn Index* dan *Connectivity Index* serta pemilihan *cluster* optimal setiap metode dan menentukan metode terbaik menggunakan rasio simpangan baku. Hasil dari penelitian ini bertujuan untuk memberikan wawasan yang lebih jelas mengenai kondisi kesehatan lingkungan di berbagai kelompok provinsi. Selain itu, temuan dari penelitian ini diharapkan dapat menjadi landasan bagi pihak-pihak berwenang dan pemerintah untuk merancang program peningkatan kesehatan lingkungan yang lebih fokus dan efisien.

II. METODE PENELITIAN

2.1. Sumber Data

Penelitian ini menggunakan data sekunder terkait Indikator Kesehatan Lingkungan 34 provinsi di Indonesia tahun 2023. Data tersebut didapatkan dari publikasi Profil Kesehatan Indonesia 2023 yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia pada tahun 2024, serta data yang bersumber dari *website* Badan Pusat Statistik (BPS). Penelitian ini menggunakan 12 variabel dari 9 indikator kesehatan lingkungan yang disajikan pada Tabel 1 dibawah ini.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Sumber
X_1	Persentase Sarana Air Minum yang Diawasi/Diperiksa Kualitas Air Minumnya Sesuai Standar	Kemenkes RI
X_2	Persentase Rumah Tangga yang Memiliki Akses Terhadap Air Minum Layak	BPS
X_3	Persentase Keluarga Dengan Akses Terhadap Fasilitas Sanitasi Yang Layak (Jamban Sehat)	Kemenkes RI
X_4	Persentase Rumah Tangga yang Memiliki Akses Terhadap Sanitasi Layak	BPS
X_5	Persentase Desa/Kelurahan Melaksanakan STBM	Kemenkes RI
X_6	Persentase Desa/Kelurahan Stop Buang Air Besar Sembarangan	Kemenkes RI
X_7	Persentase Tempat Dan Fasilitas Umum (TFU) yang Dilakukan Pengawasan Sesuai Standar	Kemenkes RI
X_8	Persentase Tempat Pengelolaan Pangan (TPP) yang Memenuhi Syarat Sesuai Standar	Kemenkes RI
X_9	Persentase Kabupaten/Kota Sehat	Kemenkes RI
X_{10}	Persentase Fasyankes yang Melakukan Pengelolaan Limbah Medis Sesuai Standar	Kemenkes RI
X_{11}	Persentase Kabupaten/Kota Menerapkan Kebijakan Gerakan Masyarakat Hidup Sehat	Kemenkes RI
X_{12}	Persentase Rumah Tangga yang Menempati Rumah Layak Huni	BPS

2.2. Analisis Cluster

Analisis *cluster* yaitu metode analisis multivariat yang dirancang guna mengelompokkan kumpulan objek menurut persamaan atributnya. Objek dalam satu *cluster* yang sama harus relatif serupa (homogen) dan mempunyai perbedaan jauh (heterogen) dari objek pada *cluster* yang berbeda. Analisis *cluster* berfokus pada dua asumsi yaitu sampel yang representatif dan tidak adanya indikasi multikolinieritas pada variabel yang diteliti [13].

2.2.1. Sample Representatif

Sampel dianggap representatif apabila sampel yang dipilih mampu mewakili atau merepresentasikan populasi yang dimaksud. *Uji Kaiser-Mayer-Olkin* (KMO) dapat digunakan untuk menguji apakah sampel bersifat representatif. Sampel dapat dikatakan representatif atau

dapat mewakili populasi apabila nilai KMO berada dalam rentang 0.5 hingga 1. KMO yang dirumuskan sebagai berikut [14].

$$KMO = \frac{\sum_{j=1}^P \sum_{l=1, l \neq j}^P r_{x_j x_l}^2}{\sum_{j=1}^P \sum_{l=1, l \neq j}^P r_{x_j x_l}^2 + \sum_{j=1}^P \sum_{l=1, l \neq j}^P \rho_{x_j x_l, x_m}^2} \quad (1)$$

dengan:

$$r_{x_j x_l}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}} \sqrt{\frac{\sum_{i=1}^n (x_{il} - \bar{x}_l)^2}{n}}} \quad (2)$$

$$\rho_{x_j x_l, x_m}^2 = \frac{r_{x_j x_l} - r_{x_j x_m} r_{x_l x_m}}{\sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}}} \quad (3)$$

- P : banyaknya variabel
 n : banyaknya objek
 x_{ij} : nilai dari objek ke- i pada variabel ke- j
 x_{il} : nilai dari objek ke- i pada variabel ke- l
 $r_{x_j x_l}$: korelasi antara variabel x_{ij} dan x_{il}
 $\bar{x}_j$: rata-rata variabel x_j
 $\bar{x}_l$: rata-rata variabel x_l
 $\rho_{x_j x_l, x_m}^2$: korelasi parsial antara variabel x_j dan x_l dengan menjaga agar x_m konstan

2.2.2. Uji Multikolinearitas

Untuk mengetahui adanya multikolinearitas dapat menggunakan *Variance Inflation Factor* (VIF) yang dihitung melalui persamaan berikut ini [15].

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4)$$

dengan:

- VIF_j : nilai VIF untuk variabel ke- j , $j = 1, 2, \dots, p$
 R_j^2 : koefisien determinasi antara x_j dan variabel bebas lainnya

Multikolinearitas terjadi ketika *Variance Inflation Factor* (VIF) bernilai sama dengan atau lebih dari 10. Jika multikolinearitas terdeteksi, harus dilakukan perbaikan untuk mengatasi masalah tersebut [14].

Dalam analisis *cluster*, pengukuran jarak dilakukan guna menentukan tingkat kesamaan antara objek-objek yang akan dikelompokkan. Salah satu pendekatan yang dapat diterapkan untuk mengukur kesamaan objek yaitu dengan menggunakan ukuran jarak (*distance measure*). Menurut Widayat (2018), jarak *euclidean* yaitu ukuran yang paling sering dan luas digunakan dalam menghitung jarak. Berikut rumus perhitungan jarak *euclidean*.

$$d_{euc}(x_i, c_k) = \sqrt{\sum_{j=1}^P (x_{ij} - c_{kj})^2} \quad (5)$$

dengan:

- $d_{euc}(x_i, c_k)$: jarak *euclidean* objek ke- i , variabel ke- j dengan pusat *cluster* ke- k
 x_{ij} : objek ke- i pada variabel ke- j
 c_{kj} : pusat *cluster* ke- k pada variabel ke- j
 P : banyaknya variabel yang diamati
 j : variabel yang diamati; $j = 1, 2, 3, \dots, m$
 i : objek yang diamati; $i = 1, 2, 3, \dots, n$
 k : banyaknya *cluster*; $k = 1, 2, 3, \dots, K$

2.3. *K-Harmonic Means (KHM)*

Pada *K-Harmonic Means*, pembentukan *cluster* dilakukan melalui penyempurnaan fungsi tujuan secara iteratif menggunakan rata-rata harmonik yang didasarkan pada posisi pusat setiap *cluster*. Metode ini ditujukan guna mengurangi rata-rata harmonik jarak semua titik data ke pusat *cluster* [7]. Rata-rata harmonik ini dirumuskan sebagai berikut.

$$HA(\{a_1, \dots, a_k\}) = \frac{K}{\sum_{k=1}^K \left(\frac{1}{a_k}\right)} \quad (6)$$

Fungsi ini memiliki karakteristik bahwa jika salah satu data di antara $a_1, \dots, a_k$ memiliki nilai yang kecil, bisa dinyatakan rata-rata harmonik yang dihasilkan juga akan kecil. Bila tidak terdapat nilai kecil, rata-rata harmonik nantinya meningkat. Metode ini beroperasi serupa dengan fungsi minimum tetapi juga memberi suatu bobot terhadap nilai-nilai lainnya [17]. Tahapan-tahapan algoritma *K-Harmonic Means* sebagai berikut [18].

1. Menetapkan posisi titik pusat *cluster* awal secara acak.
2. Hitung fungsi tujuan menggunakan Persamaan (7), dengan p sebagai parameter. Biasanya nilai $p \geq 2$.

$$KHM(X, C) = \sum_{i=1}^n \frac{K}{\sum_{k=1}^K \frac{1}{\|x_i, c_k\|^p}} \quad (7)$$

dengan:

- $KHM(X, C)$: fungsi tujuan *K-Harmonic Means*
 $\|x_i, c_k\|^p$: jarak pengamatan ke- i dengan pusat *cluster* ke- k
 C : himpunan titik pusat *cluster*
 X : himpunan data observasi
 K : banyaknya *cluster* yang digunakan
 i : $1, 2, 3, \dots, n$
 j : $1, 2, 3, \dots, m$
 k : $1, 2, 3, \dots, k$
 p : nilai parameter

3. Untuk tiap data x_i , lakukan penghitungan nilai keanggotaan $m(c_k|x_i)$ pada setiap pusat *cluster* ke- k berdasarkan pada Persamaan (8).

$$m(c_k|x_i) = \frac{\|x_i, c_k\|^{-p-2}}{\sum_{k=1}^K \|x_i, c_k\|^{-p-2}} \quad (8)$$

4. Untuk setiap objek, hitung bobot $w(x_i)$ dengan rumus yang terdapat pada Persamaan (9).

$$w(x_i) = \frac{\sum_{k=1}^K \|x_i, c_k\|^{-p-2}}{(\sum_{k=1}^K \|x_i, c_k\|^{-p})^2} \quad (9)$$

5. Untuk setiap c_k , perbarui nilainya terhadap seluruh objek dengan mempertimbangkan keanggotaan dan bobotnya dengan rumus yang terdapat pada Persamaan (10).



$$c_k = \frac{\sum_{i=1}^n m(c_k|x_i) \cdot w(x_i) \cdot x_i}{\sum_{i=1}^n m(c_k|x_i) \cdot w(x_i)} \quad (10)$$

6. Ulang tahapan 2 hingga 5 sampai proses iterasi memenuhi konvergen saat $|KHM_t - KHM_{t-1}| < \varepsilon$.
7. Tetapkan keanggotaan setiap objek ke dalam *cluster* ke- k dengan nilai keanggotaan yang terbesar.

2.4. Particle Swarm Optimization

Tahapan-tahapan untuk algoritma *Particle Swarm Optimization* (PSO) di antaranya sebagai berikut [19].

1. Menginisialisasi jumlah partikel dengan acak dan nilai kecepatan awal partikel = 0.
2. Menghitung nilai *fitness* untuk tiap partikelnya.
3. Menentukan nilai *Pbest* menggunakan rumus yang terdapat pada Persamaan (11).

$$Pbest_i^t = \begin{cases} Pbest_i^{(t-1)}, & \text{jika } f(x_i^t) \geq f(x_i^{(t-1)}) \\ x_i^t, & \text{jika } f(x_i^t) < f(x_i^{(t-1)}) \end{cases} \quad (11)$$

4. Menentukan nilai *Gbest* menggunakan rumus yang terdapat pada Persamaan (12).

$$Gbest_i^t = MIN(f(Pbest_1^t), \dots, f(Pbest_n^t)) \quad (12)$$

5. Memperbarui nilai kecepatan menggunakan rumus yang terdapat pada Persamaan (13).

$$v_i^{(t+1)} = \omega v_i^t + c_1 r_1 [Pbest_i^t - x_i^t] + c_2 r_2 [Gbest^t - x_i^t] \quad (13)$$

dengan:

- ω : bobot inersia
- v_i^t : kecepatan partikel ke- i pada iterasi ke- t
- c_1 dan c_2 : nilai konstanta akselerasi
- r_1 dan r_2 : nilai random dari 0 sampai 1
- x_i^t : posisi partikel ke- i pada iterasi ke- t
- $Pbest_i^t$: partikel terbaik ke- i pada iterasi saat ini
- $Gbest_i^t$: global terbaik pada iterasi saat ini

Partikel pada PSO dapat keluar dari ruang pencarian karena kondisi kecepatan yang terlalu lebar, yang berakibat pada menurunnya keefisienan dan laju konvergensi algoritma. Agar bisa menyelesaikan permasalahan ini, batasan tertentu diberlakukan pada komponen kecepatan, dimana semua komponen dievaluasi dalam beberapa dimensi dengan membatasi kisaran $[-V_{Max}, V_{Max}]$, untuk mengurangi kemungkinan partikel keluar dari ruang pencarian [20].

6. Memperbarui nilai posisi menggunakan rumus yang terdapat pada Persamaan (14).

$$x_i^{(t+1)} = x_i^t + v_i^{(t+1)} \quad (14)$$

dengan:

- $x_i^{(t+1)}$: nilai posisi pada iterasi ke- $t + 1$ atau iterasi selanjutnya
- x_i^t : partikel ke- i pada iterasi ke- t

7. Ulangi langkah 2 hingga mencapai iterasi maksimum.

2.5. Particle Swarm Optimization K-Harmonic Means (PSOKHM)

Integrasi KHM dengan PSO membentuk algoritma klasterisasi hibrida yang disebut PSOKHM, yang mempertahankan keunggulan KHM dan PSO [18]. Tahapan-tahapan analisis *Particle Swarm Optimization K-Harmonic Means* sebagai berikut [21].

1. Atur parameter awal.
2. Atur hitungan iteratif $t = 0$.
3. Terapkan algoritma PSO untuk memperbarui partikel-partikel *swarm*.
4. Hitung posisi vektor pusat *cluster* ke- i dari partikel menggunakan KHM.
5. Jika $t < t_{max}$, kembali ke langkah 3, jika tidak lanjutkan.



6. Tetapkan titik objek x_i ke *cluster* k dengan anggota $m(c_k|x_i)$ terbesar sebagai *cluster* awal.

2.6. Validasi Hasil Analisis Cluster

Menurut Jain & Dubes (1988) terdapat tiga pendekatan dalam validasi hasil analisis *cluster* yaitu Validasi Eksternal, Validasi Internal dan Validasi Relatif. Pada penelitian ini menggunakan *Silhouette Index*, *Dunn Index* dan *Connectivity Index* serta menggunakan rasio simpangan baku.

2.6.1. Silhouette Index

Silhouette Index digunakan untuk menilai validitas hasil analisis *cluster* yang telah dilakukan. *Index* ini tidak dipengaruhi oleh pendekatan pengelompokan yang digunakan. Menghitung *Silhouette Index* dapat dirumuskan sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (15)$$

dengan:

$s(i)$: nilai *Silhouette*, nilainya berada pada kisaran -1 hingga 1

$a(i)$: rata-rata hitung jarak objek ke- i dengan seluruh objek yang berada di dalam satu *cluster*

$b(i)$: nilai minimum dari rata-rata hitung objek ke- i dari *cluster* ke- k dengan seluruh objek di *cluster* ke- v lainnya

2.6.2. Dunn Index

Dunn Index ditentukan dengan mengukur jarak antara dua *cluster* terkait dengan jarak maksimum dalam sebuah *cluster*. Nilai *Dunn Index* yang lebih tinggi menunjukkan pembentukan *cluster* yang lebih baik [23]. Perhitungan *Dunn Index* dapat dirumuskan sebagai berikut.

$$D(k) = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left(\frac{d(c_i, c_j)}{\max_{v=1} \text{diam}(c_v)} \right) \right\} \quad (16)$$

dengan:

$D(k)$: nilai *Dunn Index*

k : jumlah *cluster*

$d(c_i, c_j)$: jarak antara *cluster* ke- i dan *cluster* ke- j

$\text{diam}(c_v)$: diameter *cluster* maksimum

Dunn Index berada pada rentang antara 0 sampai tak hingga (∞). Semakin besar nilai *Dunn Index* pada suatu *cluster*, semakin baik hasilnya [24].

2.6.3. Connectivity Index

Connectivity Index dilambangkan dengan *Conn*. Semakin kecil *Connectivity Index* maka diindikasikan menunjukkan banyak gerombol yang terbentuk lebih baik atau optimal [25]. Perhitungan *Connectivity Index* dapat dirumuskan sebagai berikut.

$$\text{Conn}(C) = \sum_{i=1}^N \sum_{j=1}^K X_{i,nn_{i(j)}} \quad (17)$$

dengan:

$\text{Conn}(C)$: *Connectivity Index*

$X_{i,nn_{i(j)}}$: nol jika i dan j berada dalam *cluster* yang sama dan $\frac{1}{j}$ jika sebaliknya

$nn_{i(j)}$: tetangga terdekat objek ke- j dari objek ke- i

N : jumlah data observasi

K : jumlah *cluster*

2.6.4. Rasio Simpangan Baku



Suatu metode pengelompokan yang dapat digunakan untuk membentuk *cluster* dianggap memiliki kinerja baik jika memiliki nilai simpangan baku dalam *cluster* yang minimum terhadap simpangan baku antar *cluster* [26]. *Cluster* optimal ditandai dengan varians besar antar *cluster* dan varians kecil dalam *cluster* [27]. Simpangan baku di dalam *cluster* atau *within* (S_W) dirumuskan sebagai berikut [28].

$$S_W = K^{-1} \sum_{k=1}^K S_k \quad (18)$$

dengan:

S_k : simpangan baku dari *cluster* ke- k

K : banyaknya *cluster* yang terbentuk

Adapun perhitungan simpangan baku antar *cluster* atau *between* (S_B) dirumuskan sebagai berikut [28].

$$S_B = \left[(K - 1)^{-1} \sum_{k=1}^K (\bar{X}_k - \bar{X})^2 \right]^{\frac{1}{2}} \quad (19)$$

dengan:

$\bar{X}_k$: nilai rata-rata dari *cluster* ke- k

$\bar{X}$: nilai rata-rata dari seluruh *cluster*

K : banyaknya *cluster* yang terbentuk

Semakin kecil rasio antara S_W dan S_B maka semakin baik *cluster* yang terbentuk [29].

Rasio antara S_W dan S_B atau rasio simpangan baku dapat dirumuskan sebagai berikut [30].

$$S = \frac{S_W}{S_B} \quad (20)$$

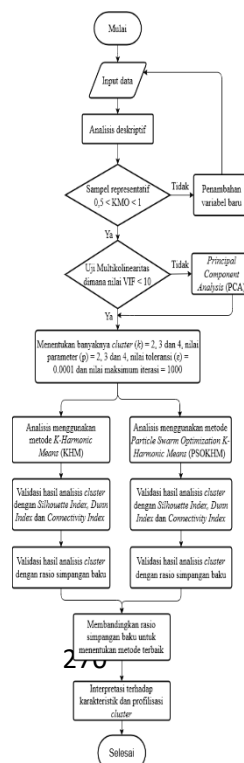
dengan:

S_W : simpangan baku di dalam *cluster*

S_B : simpangan baku antar *cluster*

2.7. Tahapan Penelitian

Tahapan penelitian yang digunakan pada penelitian ini disajikan pada gambar berikut.





III. HASIL DAN PEMBAHASAN

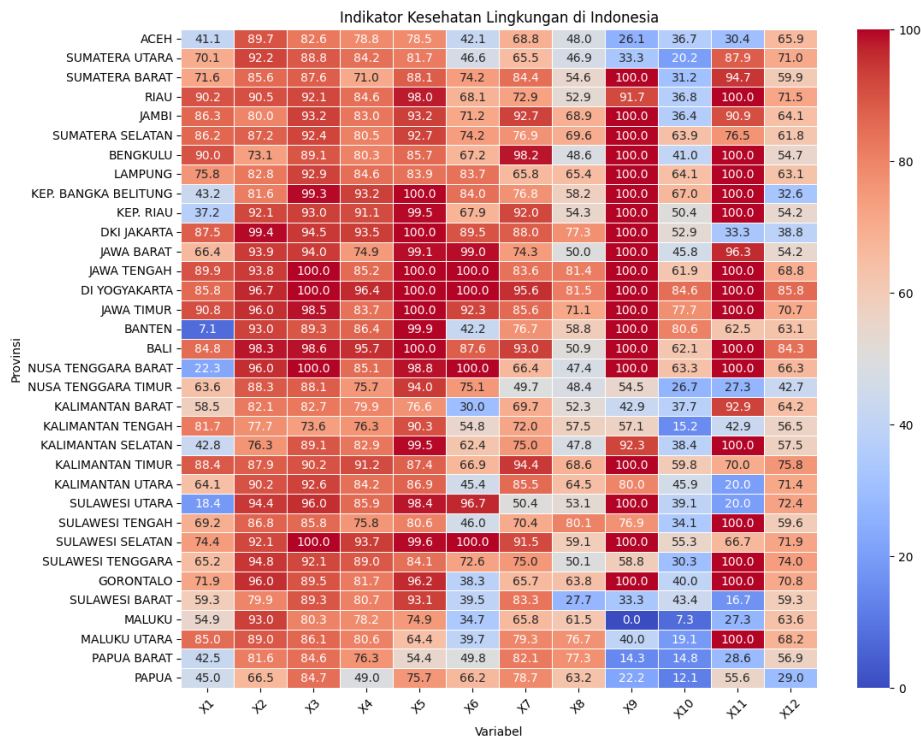
Analisis *cluster* menggunakan *K-Harmonic Means* dan *Particle Swarm Optimization K-Harmonic Means* perlu menentukan parameter awal. Pada penelitian ini menggunakan banyaknya *cluster* (k) yaitu 2 *cluster*, 3 *cluster* dan 4 *cluster*, nilai parameter (p) yaitu 2, 3 dan 4, nilai toleransi (ϵ) yaitu 0.0001 dan nilai maksimum iterasi yaitu 1000.

3.1. Analisis Deskriptif

Analisis deskriptif dilakukan guna mengetahui karakteristik dari data yang akan digunakan pada penelitian ini yaitu data Indikator Kesehatan Lingkungan dari 34 provinsi yang ada di Indonesia yang terdiri dari 12 variabel. Karakteristik masing-masing variabel pada Indikator Kesehatan Lingkungan dijelaskan dengan analisis deskriptif pada Tabel 2 dan Gambar 1.

Tabel 2. Nilai Rata-Rata Setiap Variabel

Variabel	Nilai Rata-Rata
X_1	65.04
X_2	88.19
X_3	90.90
X_4	82.75
X_5	89.86
X_6	67.88
X_7	77.81
X_8	59.93
X_9	77.16
X_{10}	43.99
X_{11}	74.72
X_{12}	62.49



Gambar 1. Gambaran Indikator Kesehatan Lingkungan Setiap Provinsi di Indonesia

Berdasarkan Tabel 2 dan Gambar 1, variabel X_1 memiliki nilai rata-rata sebesar 65.04%. Provinsi dengan persentase terendah yaitu Banten sebesar 7.1% dan provinsi dengan persentase tertinggi yaitu Jawa Timur sebesar 90.8%. Variabel X_2 memiliki nilai rata-rata sebesar 88.19%. Provinsi dengan persentase terendah yaitu Papua sebesar 66.49% dan provinsi dengan persentase tertinggi yaitu DKI Jakarta sebesar 99.42%. Variabel X_3 memiliki nilai rata-rata sebesar 90.90%. Provinsi dengan persentase terendah yaitu Kalimantan Tengah yaitu 73.6% dan provinsi dengan persentase tertinggi yaitu ada 4 provinsi sebesar 100%. Variabel X_4 memiliki nilai rata-rata sebesar 82.75%. Provinsi dengan persentase terendah yaitu Papua yaitu 43% dan provinsi dengan persentase tertinggi yaitu DI Yogyakarta sebesar 96.42%. Variabel X_5 memiliki nilai rata-rata sebesar 89.86%. Provinsi dengan persentase terendah yaitu Papua Barat yaitu 54.4% dan provinsi dengan persentase tertinggi yaitu ada 5 provinsi sebesar 100%. Variabel X_6 memiliki nilai rata-rata sebesar 67.88%. Provinsi dengan persentase terendah yaitu Kalimantan Barat sebesar 30% dan provinsi dengan persentase tertinggi yaitu Jawa Tengah, DI Yogyakarta, Nusa Tenggara Barat, dan Sulawesi Selatan sebesar 100%. Variabel X_7 memiliki nilai rata-rata sebesar 77.81%. Provinsi dengan persentase terendah yaitu Nusa Tenggara Timur sebesar 49.7% dan provinsi dengan persentase tertinggi yaitu Bengkulu sebesar 98.2%. Variabel X_8 memiliki nilai rata-rata sebesar 59.93%. Provinsi dengan persentase terendah yaitu Sulawesi Barat sebesar 27.7% dan provinsi dengan persentase tertinggi yaitu DI Yogyakarta sebesar 81.5%. Variabel X_9 memiliki nilai rata-rata sebesar 77.16%. Provinsi dengan persentase terendah yaitu Maluku sebesar 0% dan provinsi dengan persentase tertinggi yaitu ada 19 provinsi sebesar 100%. Variabel X_{10} memiliki nilai rata-rata sebesar 43.99%. Provinsi dengan persentase terendah yaitu Maluku sebesar 7.3% dan provinsi dengan persentase tertinggi yaitu DI Yogyakarta sebesar 84.6%. Variabel X_{11} memiliki nilai rata-rata sebesar 74.72%. Provinsi dengan persentase terendah yaitu Sulawesi Barat sebesar 16.7% dan provinsi dengan persentase tertinggi yaitu ada 15 provinsi yang memiliki persentase tertinggi sebesar 100%. Variabel X_{12} memiliki nilai rata-rata



sebesar 62.49%. Provinsi dengan persentase terendah yaitu Papua sebesar 29.01% dan provinsi dengan persentase tertinggi yaitu DI Yogyakarta sebesar 85.79%.

3.2. Uji Asumsi Analisis Cluster

Dalam analisis *cluster* ada 2 uji asumsi yang harus terpenuhi yaitu sampel mewakili populasi (*sample representative*) dan uji multikolinearitas.

3.2.1. Sampel Mewakili Populasi (*Sample Representative*)

Pengujian sampel yang mewakili populasi dapat dilakukan dengan uji KMO (*Kaiser Meyer Olkin*). Berdasarkan hasil pengujian KMO dalam penelitian ini dengan *RStudio* mendapatkan nilai KMO sebesar 0.72. Nilai tersebut berkisar antara 0.5 hingga 1, sehingga dapat dikatakan bahwa sampel mewakili populasi (*sample representative*).

3.2.2. Uji Multikolinearitas

Salah satu cara untuk mengidentifikasi adanya multikolinieritas yaitu dengan menggunakan nilai *Variance Inflation Factor* (VIF). Nilai $VIF \geq 10$ menunjukkan adanya masalah multikolinearitas pada variabel tersebut. Berikut ini disajikan nilai VIF dari masing-masing variabel pada Tabel 3 sebagai berikut.

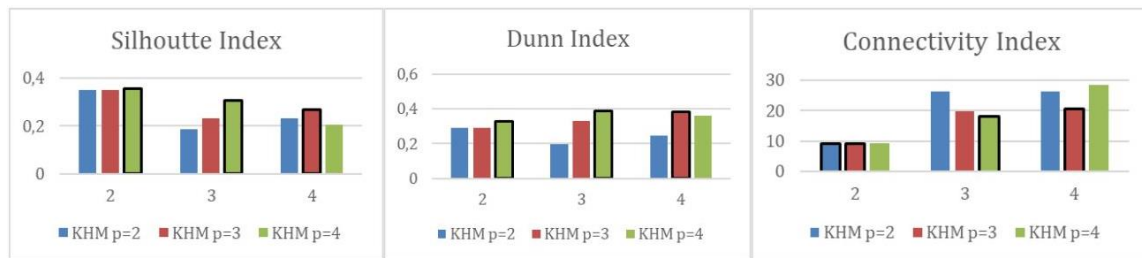
Tabel 3. Nilai VIF untuk Setiap Variabel

Variabel	Nilai VIF
X ₁	1.770123
X ₂	2.850527
X ₃	5.322414
X ₄	2.676846
X ₅	5.033400
X ₆	3.496080
X ₇	1.773011
X ₈	2.036761
X ₉	5.245479
X ₁₀	3.786357
X ₁₁	1.646908
X ₁₂	1.960472

Berdasarkan Tabel 3, dapat diketahui bahwa nilai VIF dari masing-masing variabel < 10 . Hal ini menunjukkan bahwa tidak terjadi multikolinearitas dari setiap variabel yang akan digunakan untuk analisis.

3.3. Analisis Metode K-Harmonic Means

Dalam analisis *K-Harmonic Means* dengan banyaknya *cluster* (*k*) yaitu 2 *cluster*, 3 *cluster* dan 4 *cluster* dan nilai parameter (*p*) yaitu 2, 3 dan 4 mendapatkan nilai *Silhouette Index*, *Dunn Index* dan *Connectivity Index* yang disajikan pada Gambar 2 sebagai berikut.



Gambar 2. Validasi Hasil Analisis *Cluster* Metode *K-Harmonic Means*

Berdasarkan Gambar 2, didapatkan nilai parameter optimal yaitu memiliki *Silhouette Index* paling besar, memiliki *Dunn Index* yang mendekati 1 dan memiliki *Connectivity Index* paling kecil setiap *cluster* yaitu nilai parameter (p) = 4 yang optimal pada 2 *cluster* dan 3 *cluster*.

Kemudian, menentukan banyaknya *cluster* yang optimal menggunakan rasio simpangan baku. Rasio yang lebih kecil menunjukkan kualitas *cluster* yang lebih baik. Hasil rasio simpangan baku masing-masing *cluster* dengan nilai parameter (p) = 4 yang disajikan pada Tabel 4 sebagai berikut.

Tabel 4.
Rasio Simpangan Baku Setiap *Cluster* Metode KHM

Banyaknya Cluster	Nilai Parameter (p)	S_w	S_B	Rasio
2	$p = 4$	13.39570	64.34433	0.2081877
3	$p = 4$	13.16600	60.89567	0.2162059
4	$p = 4$	12.26446	56.85146	0.2157281

Berdasarkan Tabel 4, didapatkan rasio simpangan baku dari masing-masing *cluster* pada metode *K-Harmonic Means* dengan nilai parameter (p) = 4. Rasio yang dihasilkan pada masing-masing *cluster* menunjukkan nilai terkecil pada saat banyaknya *cluster* yaitu 4 *cluster* sebesar 0.2157281. Oleh karena itu, pada metode *K-Harmonic Means* ini yaitu banyaknya *cluster* yang optimal yaitu 4 *cluster* dan nilai parameter (p) yang optimal yaitu 4.

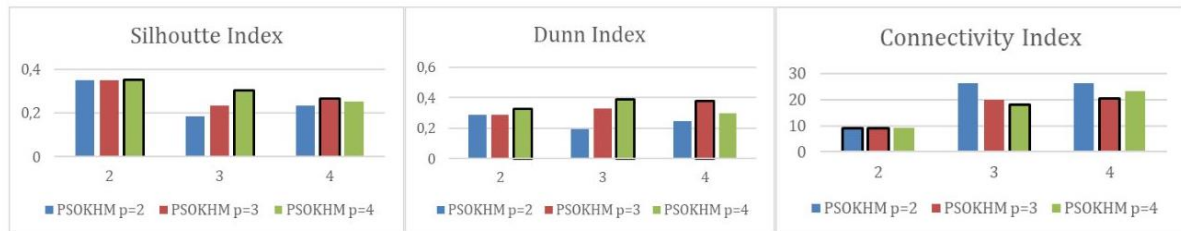
3.4. Particle Swarm Optimization K-Harmonic Means (PSOKHM)

Dalam analisis *Particle Swarm Optimization K-Harmonic Means* (PSOKHM) pada penelitian ini menggunakan parameter PSO yang didapatkan dari hasil pengujian berdasarkan rata-rata nilai *fitness* yang terendah. Parameter PSO yang digunakan sebagai berikut.

Tabel 5. Parameter PSO yang Digunakan

Parameter	Nilai
ω_{min}	0.2
ω_{max}	0.8
c_1	2
c_2	0.5
Jumlah iterasi	90
Jumlah partikel	50

Hasil analisis metode *Particle Swarm Optimization K-Harmonic Means* (PSOKHM) dengan banyaknya *cluster* (k) yaitu 2 *cluster*, 3 *cluster* dan 4 *cluster* dan nilai parameter (p) yaitu 2, 3 dan 4 mendapatkan nilai *Silhouette Index*, *Dunn Index* dan *Connectivity Index* yang disajikan pada Gambar 3 sebagai berikut.

**Gambar 3.** Validasi Hasil Analisis *Cluster Metode Particle Swarm Optimization K-Harmonic Means*

Berdasarkan Gambar 3, didapatkan nilai parameter optimal yaitu memiliki *Silhouette Index* paling besar, memiliki *Dunn Index* yang mendekati 1 dan memiliki *Connectivity Index* paling kecil setiap *cluster* yaitu nilai parameter (p) = 4 yang optimal pada 2 *cluster* dan 3 *cluster*.

Kemudian, menentukan banyaknya *cluster* yang optimal menggunakan rasio simpangan baku. Rasio yang lebih kecil menunjukkan kualitas *cluster* yang lebih baik. Hasil rasio simpangan baku masing-masing *cluster* dengan nilai parameter (p) = 4 yang disajikan pada Tabel 6 sebagai berikut.

Tabel 6. Rasio Simpangan Baku Setiap *Cluster Metode PSOKHM*

Banyaknya <i>Cluster</i>	Nilai Parameter (p)	S_W	S_B	Rasio
2	$p = 4$	13.39570	64.34433	0.2081877
3	$p = 4$	13.16600	60.89567	0.2162059
4	$p = 4$	11.22038	61.32161	0.1829759

Berdasarkan Tabel 6, didapatkan rasio simpangan baku dari masing-masing *cluster* pada metode PSOKHM. Rasio yang dihasilkan pada masing-masing *cluster* menunjukkan nilai terkecil pada saat banyaknya *cluster* yaitu 4 *cluster* sebesar 0.1829759. Oleh karena itu, pada PSOKHM yaitu banyaknya *cluster* yang optimal yaitu 4 *cluster* dan nilai parameter (p) yang optimal yaitu 4.

3.5. Perbandingan Metode KHM dan PSOKHM

Metode KHM dan PSOKHM akan dibandingkan nilai rasio simpangan bakunya untuk mendapatkan metode terbaik yang dapat digunakan dalam interpretasi terhadap karakteristik dan profilisasi pengelompokan provinsi di Indonesia berdasarkan Indikator Kesehatan Lingkungan tahun 2023. Hasil perbandingan nilai rasio simpangan baku yang disajikan pada Tabel 7 sebagai berikut.

Tabel 7. Perbandingan Rasio Simpangan Baku KHM dan PSOKHM

Metode	Banyaknya <i>Cluster</i>	Nilai Parameter (p)	S_W	S_B	Rasio
KHM	4	4	12.26446	56.85146	0.2157281
PSOKHM	4	4	11.22038	61.32161	0.1829759

Berdasarkan Tabel 7, dilihat dari nilai rasio simpangan baku yang memiliki nilai terkecil yaitu pada metode PSOKHM sebesar 0.1829759, sehingga untuk interpretasi terhadap karakteristik dan profilisasi *cluster* akan menggunakan hasil pengelompokan dari metode PSOKHM dengan banyaknya *cluster* (k) yaitu 4 *cluster* dan nilai parameter (p) yaitu 4. Hasil pengelompokan metode PSOKHM dengan banyaknya *cluster* (k) yaitu 4 *cluster* dan nilai parameter (p) yaitu 4 disajikan pada Tabel 8 sebagai berikut.

Tabel 8. Hasil Pengelompokan PSOKHM dengan $k=4$ dan $p=4$

Cluster	Jumlah Anggota	Provinsi
1	16	Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep.



		Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Kalimantan Timur dan Sulawesi Selatan
2	7	Sumatera Utara, Kalimantan Barat, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo dan Maluku Utara
3	3	Banten, Nusa Tenggara Barat dan Sulawesi Utara
4	8	Aceh, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Utara, Sulawesi Barat, Maluku, Papua Barat dan Papua

3.6. Interpretasi terhadap karakteristik dan Profilisasi cluster

Interpretasi terhadap karakteristik dan profilisasi hasil pengelompokan melibatkan pemeriksaan setiap *cluster* dan deskripsi karakteristik masing-masing *cluster* untuk menjelaskan perbedaannya dengan *cluster* lain. Karakteristik suatu *cluster* dapat diamati dari nilai rata-rata variabelnya. Rata-rata variabel pada PSOKHM dengan banyaknya *cluster* (k) yaitu 4 *cluster* dan nilai parameter (p) yaitu 4 yang disajikan pada Tabel 9 sebagai berikut.

Tabel 9.
Nilai Rata-Rata Variabel pada Setiap *Cluster*

Variabel	Cluster			
	1	2	3	4
X ₁	78.03	66.10	15.93	56.52
X ₂	89.43	88.17	94.45	83.36
X ₃	94.71	87.73	95.10	84.47
X ₄	86.41	82.02	85.81	74.91
X ₅	95.45	83.30	99.03	80.97
X ₆	82.86	47.94	79.63	50.95
X ₇	85.36	71.51	64.50	73.24
X ₈	63.27	59.67	53.10	56.01
X ₉	99.48	63.46	100.00	35.94
X ₁₀	55.68	31.40	61.00	25.26
X ₁₁	89.27	97.26	60.83	31.10
X ₁₂	63.27	66.45	67.25	55.67

Berikut merupakan interpretasi terhadap karakteristik dan profilisasi dari masing-masing *cluster*.

a. Cluster 1

Cluster 1 yang terdiri dari 16 provinsi memiliki lima variabel dengan nilai rata-rata tertinggi dibandingkan *cluster* lainnya yaitu X₁, X₄, X₆, X₇ dan X₈. *Cluster 1* memiliki variabel dengan nilai rata-rata terendah kedua yaitu X₁₂, sehingga *cluster 1* perlu melakukan perbaikan dalam hal hunian yang layak, yang berpengaruh langsung pada kesehatan dan kesejahteraan.

b. Cluster 2

Cluster 2 yang terdiri dari 7 provinsi memiliki variabel dengan nilai rata-rata tertinggi dibandingkan *cluster* lainnya yaitu X₁₁. Sementara itu, *cluster 2* memiliki variabel dengan nilai rata-rata terendah yaitu X₆, sehingga *cluster 2* perlu memberikan perhatian khusus dengan meningkatkan sosialisasi dan edukasi terkait pentingnya sanitasi yang baik.

c. Cluster 3

Cluster 3 yang terdiri dari 3 provinsi memiliki enam variabel dengan nilai rata-rata tertinggi dibandingkan *cluster* lainnya yaitu X₂, X₃, X₅, X₉, X₁₀ dan X₁₂. Sementara itu, *cluster 3* memiliki tiga variabel dengan nilai rata-rata terendah dibandingkan *cluster* lainnya yaitu X₁, X₇ dan X₈. Oleh karena itu, *cluster 3* perlu memberikan perhatian khusus pada peningkatan



pengawasan dan pengelolaan sarana air minum, tempat dan fasilitas umum, serta tempat pengelolaan pangan agar memenuhi standar kesehatan yang berlaku.

d. *Cluster 4*

Cluster 4 yang terdiri dari 8 provinsi memiliki variabel dengan nilai rata-rata tertinggi kedua yaitu X_4 . Sementara itu, *cluster 4* memiliki delapan variabel terendah dibandingkan *cluster* lainnya yaitu X_2 , X_3 , X_4 , X_5 , X_9 , X_{10} , X_{11} dan X_{12} . Oleh karena itu, *cluster 4* perlu meningkatkan akses terhadap air minum dan sanitasi layak, serta memperkuat pelaksanaan program STBM dan pengelolaan limbah medis sesuai standar serta perlu meningkatkan kesadaran masyarakat melalui kebijakan hidup sehat dan memperbaiki kualitas hunian untuk menciptakan lingkungan yang sehat dan layak huni.

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, didapatkan nilai parameter optimal pada metode KHM yang diperoleh berdasarkan *Silhouette Index*, *Dunn Index* dan *Connectivity Index* dengan nilai yaitu 4 yang optimal pada 2 *cluster* dan 3 *cluster*. *Cluster* optimal diperoleh berdasarkan rasio simpangan baku terkecil, yaitu 0.2157281 pada 4 *cluster*. Sedangkan, nilai parameter optimal pada metode PSOKHM yang diperoleh berdasarkan nilai *Silhouette Index*, *Dunn Index* dan *Connectivity Index* dengan nilai yaitu 4 yang optimal pada 2 *cluster* dan 3 *cluster*. *Cluster* optimal diperoleh berdasarkan rasio simpangan baku terkecil, yaitu 0.1829759 pada 4 *cluster*. Berdasarkan perbandingan rasio simpangan baku dari kedua metode diperoleh metode terbaik yaitu metode PSOKHM yang memiliki rasio simpangan baku yaitu 0.1829759 yang lebih kecil dibandingkan rasio simpangan baku metode KHM. Berdasarkan hasil interpretasi terhadap karakteristik dan profilisasi *cluster*, terdapat variasi kondisi kesehatan lingkungan di Indonesia. *cluster 1* terdiri dari 16 provinsi menunjukkan pengawasan air dan sanitasi yang baik tetapi perlu perbaikan rumah layak huni, *cluster 2* yang terdiri dari 7 provinsi memiliki kebijakan hidup sehat yang kuat tetapi ada masalah sanitasi, *cluster 3* yang terdiri dari 3 provinsi perlu meningkatkan pengawasan fasilitas dan *cluster 4* yang terdiri dari 8 provinsi memiliki akses air dan sanitasi yang masih rendah.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Kementerian Kesehatan Republik Indonesia (Kemenkes RI) dan Badan Pusat Statistik (BPS) atas dukungan sebagai penyedia data dan informasi statistik, serta kepada Program Studi Statistika dan Program Studi Sains Data Universitas Muhammadiyah Semarang yang telah memberikan kesempatan dan bantuan dalam penyelesaian artikel ilmiah ini.

REFERENSI

- [1] Y. D. Putri, I. R. HG, and H. Yozza, “Penerapan Analisis Cluster Ensemble Untuk Mengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Kesehatan Lingkungan,” *Jurnal Matematika UNAND*, vol. 8, no. 1, pp. 323–330, May 2019, doi: <https://doi.org/10.25077/jmu.8.1.323-330.2019>.
- [2] WHO, “Environmental Health,” WHO. Accessed: Nov. 06, 2024. [Online]. Available: https://www.who.int/health-topics/environmental-health#tab=tab_1
- [3] Kementerian Kesehatan Republik Indonesia, *Profil Kesehatan Indonesia 2023*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2024.
- [4] A. Zhu, Z. Hua, Y. Shi, Y. Tang, and L. Miao, “An improved k-means algorithm Based on Evidence Distance,” *Entropy*, vol. 23, no. 11, pp. 1–15, Nov. 2021, doi: 10.3390/e23111550.

- [5] D. I. Yunistya, R. Goejantoro, F. Deny, and T. Amijaya, “Penerapan Metode K-Harmonic Means Dalam Pengelompokan Kabupaten/Kota (Studi Kasus: Kemiskinan di Pulau Kalimantan Tahun 2020),” *Jurnal Matematika, Statistika dan Komputasi*, vol. 19, no. 1, pp. 51–64, Sep. 2022, doi: 10.20956/j.v19i1.21116.
- [6] B. Zhang, M. Hsu, and U. Dayal, “K-Harmonic Means-A Data Clustering Algorithm,” *Hewlett-Packard Labs Technical Report HPL-1999-124*, Oct. 1999.
- [7] G. Hamerly and C. Elkan, “Alternatives To The K-Means Algorithm That Find Better Clusterings,” *International Conference on Information and Knowledge Management*, pp. 600–607, Nov. 2002, doi: <https://doi.org/10.1145/584792.584890>.
- [8] J. M. Pena, J. A. Lozano, and P. Larranaga, “An Empirical Comparison Of Four Initialization Methods For The K-Means Algorithm,” *Pattern Recognit Lett*, vol. 20, pp. 1027–1040, 1999, doi: [https://doi.org/10.1016/S0167-8655\(99\)00069-0](https://doi.org/10.1016/S0167-8655(99)00069-0).
- [9] R. Hassan, B. Cohanin, O. De Weck, and V. Gerhard, “A Comparison Of Particle Swarm Optimization And The Genetic Algorithm,” *American Institute of Aeronautics and Astronautics*, vol. 2, pp. 1–13, 2004, doi: 10.2514/6.2005-1897.
- [10] I. L. Putra, “Implementasi Algoritma Particle Swarm Optimization (PSO) Dan K-Nearest Neighbor (K-NN) Dalam Memprediksi Keberhasilan Anak SMK Mendapatkan Kerja,” *Technologia*, vol. 13, no. 4, pp. 339–350, Oct. 2022, doi: <http://dx.doi.org/10.31602/tji.v13i4.8167>.
- [11] M. Tuegeh, A. Soeprijanto, and M. Hery, “Optimal Generator Scheduling Based on Particle Swarm Optimization,” *Seminar Nasional Informatika 2009*, pp. 25–32, May 2009.
- [12] S. I. Christienova, Y. Suparman, Zulhanif, and C. Afifanto, “Hybrid Analysis Using Particle Swarm Optimization K-Harmonic Means Clustering Method for Sister Village of Remote Areas in Survey,” *World Appl Sci J*, vol. 35, no. 10, pp. 2187–2193, 2017, doi: 10.5829/idosi.wasj.2017.2187.2193.
- [13] J. Hair, W. Black, B. Babin, and R. Anderson, *Multivariate Data Analysis*, 8th ed. Hampshire: Cengage Learning, 2019.
- [14] A. Widarjono, *Analisis Statistika Multivariat Terapan*, 1st ed. Yogyakarta: UPP STIM YKPN, 2010.
- [15] D. N. Gujarati, *Basic Econometrics Fourth Edition*, 4th ed. New York: MCGraw-Hill Higher Education, 2003. [Online]. Available: www.mhhe.com
- [16] Widayat, *Statistika Multivariat (Pada Bidang Manajemen dan Bisnis)*. Malang: UMM Press, 2018.
- [17] Z. Güngör and A. Ünler, “K-Harmonic Means Data Clustering With Simulated Annealing Heuristic,” *Appl Math Comput*, vol. 184, no. 2, pp. 199–209, Jan. 2007, doi: 10.1016/j.amc.2006.05.166.
- [18] F. Yang, T. Sun, and C. Zhang, “An Efficient Hybrid Data Clustering Method Based On K-Harmonic Means And Particle Swarm Optimization,” *Expert Syst Appl*, vol. 36, pp. 9847–9852, 2009, doi: 10.1016/j.eswa.2009.02.003.
- [19] B. V. Silpa, “An Optimal Power Flow To Improve Power System Security By Using Particle Swarm Optimization,” *Electrical and Electronics Engineering: An International Journal*, vol. 4, no. 2, pp. 79–92, May 2015, doi: 10.14810/elelij.2015.4207.
- [20] F. van den Bergh and A. P. Engelbrecht, “A Cooperative Approach To Particle Swarm Optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225–239, Jun. 2004, doi: 10.1109/TEVC.2004.826069.
- [21] A. J. M. Rani and L. Parthipan, “Fuzzified PSO and K- Harmonic Means Algorithm for Electrical Data Clustering,” *International Conference on Recent Trends in Information Technology (ICRTIT)*, pp. 546–550, 2013, doi: <https://doi.org/10.1109/ICRTIT.2013.6844261>.
- [22] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. New Jersey: Prentice-Hall Inc., 1988.
- [23] N. P. Shalsabila, N. Amalita, D. Vionanda, and D. Permana, “Grouping Level of Poverty Based on District/City in Indonesia Using K-Harmonic Means,” *UNP Journal of Statistics and Data Science*, vol. 1, no. 3, pp. 157–163, May 2023, doi: 10.24036/ujsds/vol1-iss3/60.



- [24] G. R. Suraya and A. W. Wijayanto, “Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting,” *Indonesian Journal of Statistics and Its Applications*, vol. 6, no. 2, pp. 180–201, Aug. 2022, doi: 10.29244/ijsa.v6i2p180-201.
- [25] J. Wijaya and M. Stevany Manurung, “Penggerombolan Kabupaten/Kota Di Sumatera Utara Berdasarkan Produktivitas Tanaman Pangan,” *JOURNAL OF ANALYTICAL RESEARCH, STATISTICS AND COMPUTATION*, vol. 2, no. 2, pp. 36–51, Sep. 2023, doi: <https://doi.org/10.4590/jarsic.v2i2.23>.
- [26] B. Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [27] B. Oktaviandi, T. O. Mukhti, Y. Kurniawati, and Z. Martha, “Implementation of the Fuzzy C-Means Clustering Method in Grouping Provinces in Indonesia based on the Types of Goods Sold in E-commerce Businesses in 2022,” *UNP Journal of Statistics and Data Science*, vol. 2, no. 3, pp. 360–365, Aug. 2024, doi: 10.24036/ujsds/vol2-iss3/210.
- [28] M. J. Bunkers, J. R. Miller, and A. T. DeGaetano, “Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique,” *J Clim*, vol. 9, no. 1, pp. 130–146, 1996, doi: 10.1175/1520-0442(1996)009<0130:DOCRIT>2.0.CO;2.
- [29] S. I. Christienova and Y. Suparman, “Clustering dengan Metode K-Means dan K-Harmonic Means: Pengelompokan Sister Village Daerah Sulit dalam Survei,” *Prosiding Seminar Nasional Statistika*, vol. 6, pp. 33–40, 2017, doi: <https://doi.org/10.1234/pns.v6i.126>.
- [30] N. Ulinmuh and R. Veriani, “Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward,” *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, 2020, doi: 10.30743/infotekjar.v5i1.2464.