

ANALISIS FAKTOR DAN PENGELOMPOKAN INDIKATOR KESEHATAN LINGKUNGAN PROVINSI DI INDONESIA TAHUN 2021

Sartikasari^{1*}

¹Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia

*sartikasari.2020431@mhs.its.ac.id

ABSTRAK

Lingkungan sebagai tempat tinggal memberikan sumber daya alam yang penting bagi kehidupan manusia. Lingkungan harus selalu dijaga agar ekosistem lingkungan tidak menimbulkan dampak negatif yang dapat mempengaruhi kesehatan manusia. Sehingga, untuk mewujudkan hidup sehat diperlukan upaya untuk menjaga dan meningkatkan kesehatan lingkungan. Kesehatan lingkungan merupakan kondisi lingkungan baik fisik, kimia, biologi, dan sosial dengan kesehatan manusia. Penelitian ini bertujuan untuk mengidentifikasi variabel-variabel yang saling terkait dan menggabungkannya menjadi faktor-faktor tunggal yang mencerminkan aspek-aspek kesehatan lingkungan dan mengelompokkan provinsi berdasarkan variabel kesehatan lingkungan. Metode yang digunakan adalah analisis faktor dan *k-means clustering*. Analisis faktor merupakan metode yang digunakan untuk mereduksi variabel menjadi beberapa faktor baru yang menjelaskan variabel asli semaksimal mungkin. Analisis *cluster* merupakan analisis yang digunakan untuk mengelompokkan observasi atau objek berdasarkan kemiripan observasi atau *objek*. Hasil menunjukkan bahwa dari sepuluh variabel terbentuk menjadi tiga faktor baru, dan pada pengelompokkan sebanyak dua *cluster* diperoleh pada *cluster* 1 terdiri dari 14 provinsi dan *cluster* 2 terdiri dari 20 provinsi. Jumlah *cluster* sebanyak dua merupakan jumlah *cluster* optimum.

Kata-kata kunci: analisis faktor, kesehatan lingkungan, *K-means*

ABSTRACT

The environment as a place to live provides natural resources that are important for human life. The environment must always be maintained so that the environmental ecosystem does not cause negative impacts that can affect human health. Thus, to realize a healthy life, efforts are needed to maintain and improve environmental health. Environmental health is an environmental condition both physical, chemical, biological, and social with human health. This study aims to identify interrelated variables and combine them into single factors that reflect aspects of environmental health and group provinces based on environmental health variables. The methods used are factor analysis and k-means clustering. Factor analysis is a method used to reduce variables to several new factors that explain the original variable as much as possible. Cluster analysis is an analysis used to group observations or objects based on the similarity of observations or objects. The results showed that from ten variables formed into three new factors, and in grouping as many as two clusters obtained in cluster 1 consisting of 14 provinces and cluster 2 consisting of 20 provinces. The number of clusters as many as two is the optimal number of clusters.

Keywords: environmental health, factor analysis, K-means

Pendahuluan

Manusia hidup dalam masyarakat sangat bergantung pada lingkungan. Lingkungan memberikan sumber daya alam yang penting bagi kehidupan manusia, seperti air, udara, tanah, pangan, dan energi yang dapat dimanfaatkan untuk keberlangsungan hidup. Pemanfaatan sumber daya alam menjadi kunci dalam memenuhi kebutuhan sehari-hari dan memajukan peradaban manusia. Namun, pemanfaatan yang tidak terkendali dan kurang bertanggung jawab terhadap lingkungan dapat menyebabkan dampak negatif yang serius terhadap kesehatan manusia dan ekosistem. Kesehatan lingkungan merupakan kondisi lingkungan baik fisik, kimia, biologi, dan sosial dengan kesehatan manusia. Hal ini berarti bahwa, kesehatan manusia juga dipengaruhi oleh kualitas lingkungan tempat tinggal. Sehingga, untuk mewujudkan hidup sehat diperlukan upaya untuk menjaga dan meningkatkan kesehatan lingkungan. Analisis terhadap kesehatan lingkungan, dapat memberikan wawasan penting tentang hubungan antara lingkungan dan kesehatan manusia dan dapat diidentifikasi pola-pola kesehatan yang terkait dengan faktor-faktor lingkungan tertentu, mengenali kelompok populasi yang rentan, dan merancang intervensi yang sesuai. Pada penelitian ini akan digunakan metode multivariat yaitu analisis faktor menggunakan PCA dan analisis *cluster*.

Analisis multivariat merupakan analisis yang digunakan untuk menganalisis data yang terdiri dari banyak atau multi variabel. Analisis multivariat dibagi menjadi dua metode yaitu metode dependen dan metode independen. Beberapa contoh metode independen adalah analisis faktor dan analisis *cluster*. Analisis faktor adalah metode yang digunakan untuk mereduksi dimensi. Analisis faktor merupakan metode untuk menganalisis variabel-variabel yang saling berkorelasi atau dependen yang bertujuan untuk memperoleh sejumlah kecil faktor yang mampu menjelaskan semaksimal mungkin keragaman data dan menghasilkan faktor yang saling independen (Kaharuddin, 2020). Sedangkan analisis *cluster* adalah analisis yang digunakan untuk pengelompokan individu atau observasi ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok dan anggota kelompok memiliki sifat yang relatif homogen atau sama (Talaku, Leleury and Talluta, 2017).

Penelitian ini menganalisis faktor-faktor kesehatan lingkungan provinsi di Indonesia tahun 2021 yang terdiri dari 10 variabel. Variabel tersebut akan direduksi kemudian dibentuk faktor baru dan melakukan pengelompokan provinsi berdasarkan faktor kesehatan lingkungan. Melalui penelitian ini, diharapkan bisa didapatkan pemahaman yang lebih baik tentang faktor-faktor kesehatan lingkungan dan memberikan gambaran lebih jelas tentang kondisi kesehatan lingkungan di provinsi-provinsi di Indonesia tahun 2021. Hasil analisis faktor akan membantu mengidentifikasi variabel-variabel yang saling terkait dan menggabungkan menjadi faktor-faktor tunggal yang mencerminkan aspek-aspek kesehatan lingkungan yang relevan dan dapat digunakan untuk menyusun kebijakan dan intervensi yang lebih efektif dalam meningkatkan kesehatan lingkungan di provinsi-provinsi di Indonesia.

Tinjauan Pustaka

Statistika Deskriptif

Statistika deskriptif adalah suatu metode tentang bagaimana mengumpulkan data, menyajikan data dalam bentuk tabel maupun grafik, mengolah dan menganalisis data kemudian diinterpretasikan hingga diperoleh sebuah kesimpulan. Pada statistika deskriptif suatu data dapat disajikan dalam bentuk tabel maupun grafik. Selain itu, pada statistika deskriptif mempelajari perhitungan ukuran pemusatan dan ukuran penyebaran data. Ukuran pemusatan data terdiri dari rata-rata, modus, dan median. Sedangkan ukuran penyebaran data meliputi jangkauan, varians, dan standar deviasi (Silvia, 2020).

Asumsi PCA dan Analisis Faktor

Suatu data dapat dilakukan analisis komponen utama dan analisis faktor jika memenuhi asumsi distribusi normal multivariat, dependen, data cukup, dan saling berkorelasi antar variabel. Berikut merupakan penjelasan asumsi PCA dan analisis faktor.

1. Distribusi Normal Multivariat

Distribusi normal multivariat merupakan asumsi yang dilakukan untuk mengetahui apakah data mengikuti pola distribusi normal multivariat. Pengujian distribusi normal multivariat dapat menggunakan uji T_{proporsi} dengan hipotesis awal (H_0) yaitu data berdistribusi normal multivariat dan hipotesis alternatif (H_1) data tidak berdistribusi normal multivariat. Data dikatakan berdistribusi normal multivariat jika T_{proporsi} berada antara $45\% \leq T_{\text{proporsi}} \leq 55\%$ (Johnson and Wichern, 2007). Statistik uji T_{proporsi} ditunjukkan pada persamaan 1 berikut.

$$t = \frac{\text{persentase}(d_i^2 < \chi_p^2)}{n} \times 100\% \quad (1)$$

$$d_i^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \quad (2)$$

n merupakan jumlah data, $\bar{\mathbf{x}}$ = rata-rata data, $\boldsymbol{\Sigma}$ = matriks varians-kovarians data, dan d_i^2 = jarak Mahalanobis

2. Dependen

Asumsi dependen merupakan asumsi data yang akan dianalisis saling berhubungan. Pengujian dependensi dapat dilakukan menggunakan uji *Bartlett* yang berasal dari *Likelihood ratio test* dengan asumsi distribusi normal. Hipotesis pada uji *Bartlett* adalah sebagai berikut.

$H_0 : \boldsymbol{\rho} = \mathbf{I}$ (matriks korelasi merupakan matriks identitas atau antar independen)

$H_1 : \boldsymbol{\rho} \neq \mathbf{I}$ (matriks korelasi bukan merupakan matriks identitas atau dependen)

Statistik uji

$$\chi^2 = -\left(n - 1 - \frac{2p + 5}{6}\right) \ln|\mathbf{P}| \quad (3)$$

H_0 ditolak jika $\chi^2 > \chi^2_{(1-\alpha; df)}$ dengan df atau derajat bebas sebesar $p(p-1)/2$ (Arsham and Lovric, 2011).

3. Data Cukup

Asumsi data cukup menunjukkan apakah data yang digunakan telah layak atau dapat dilakukan analisis berdasarkan nilai *Kaiser Meyer Olkin* (KMO). Asumsi data cukup jika nilai KMO lebih besar dari 0.5. Statistik uji KMO ditunjukkan pada persamaan 4 berikut.

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (4)$$

4. Korelasi Antar Variabel

Pemeriksaan korelasi antar variabel menggunakan *Measuring of Sampling Adequacy* (MSA) adalah indeks perbandingan antara koefisien korelasi parsial untuk setiap variabel. Jika nilai MSA lebih kecil dari 0.5 maka variabel dikeluarkan dan analisis menggunakan variabel dengan MSA lebih besar dari 0.5. Statistik uji MSA ditunjukkan pada persamaan 5 berikut.

$$MSA = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (5)$$

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) atau analisis komponen utama adalah metode multivariat yang menganalisis data pada variabel yang saling berkorelasi dengan tujuan mengekstraksi informasi penting dan menggambarkan menjadi satu set variabel baru yang disebut komponen utama. PCA dapat digunakan pada tahapan *Confirmatory Principal Analysis* (CFA). Pendekatan pembentukan faktor dapat dilakukan berdasarkan matriks korelasi atau matriks kovarians. Penentuan jumlah komponen utama pada PCA ditentukan berdasarkan *eigen value* lebih besar dari 1 jika menggunakan matriks korelasi, nilai proporsi kumulatif yang dapat dijelaskan lebih dari 80% dan berdasarkan *scree plot* yaitu titik belok dari *scree plot* (Noya van Delsen, Wattimena and Saputri, 2017).

Analisis Faktor

Analisis faktor merupakan metode untuk menganalisis variabel-variabel yang saling berkorelasi atau dependen yang bertujuan untuk memperoleh sejumlah kecil faktor yang mampu menjelaskan semaksimal mungkin keragaman data dan menghasilkan faktor yang saling

independen (Kaharuddin, 2020). Persamaan model analisis faktor ditunjukkan pada persamaan 6 berikut.

$$\begin{aligned} X_1 - \mu_1 &= \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p \end{aligned} \quad (6)$$

Keterangan :

μ_p = mean dari variabel ke- p

ε_p = faktor spesifik ke- p

F_m = common factor ke- m

ℓ_{pm} = loading factor dari variabel ke- p pada faktor ke- m

Analisis Klaster

Analisis klaster merupakan suatu metode pada analisis multivariat yang bertujuan untuk meng-clusterkan data observasi ataupun variabel-variabel ke dalam *cluster* sedemikian rupa sehingga masing-masing bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Tujuan dari analisis *cluster* adalah mengelompokkan objek yang mirip ke dalam klaster yang sama (Hikmah *et al.*, 2022). Terdapat dua macam analisis *cluster* yaitu metode *cluster* hierarki dan non-hierarki.

K-means clustering merupakan algoritma untuk *cluster* n objek berdasarkan atribut menjadi k *cluster*. Pada *K-means* jumlah *cluster* ditentukan terlebih dahulu di awal, kemudian objek pengamatan digabungkan ke dalam klaster-klaster tersebut. Pada metode klaster non hierarki meliputi metode *sequential threshold*, *parallel threshold*, dan *optimizing partitioning* (Hikmah *et al.*, 2022). Langkah pertama pada *k-means clustering* adalah menentukan jumlah *cluster* sebanyak k *cluster*, kemudian menentukan nilai *centroid*, menghitung jarak antara titik *centroid* dengan titik tiap objek, dan pengelompokan objek. Terdapat beberapa jarak yang digunakan pada analisis *cluster* diantaranya jarak *Euclidean*, *Manhattan*, *Minkowski*, *Jaccard*, *cosine*, *Chebyshev*, dan lain-lain. Jarak yang sering digunakan adalah jarak *Euclidean* dengan rumus sebagai berikut.

$$d_{i,j} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

Kesehatan Lingkungan

Kesehatan lingkungan merupakan salah satu kegiatan kesehatan masyarakat sebagai upaya pencegahan terjadinya penyakit. Kesehatan lingkungan bertujuan untuk mewujudkan kualitas lingkungan yang sehat, secara fisik, kimia, biologi, maupun sosial yang memungkinkan masyarakat mencapai derajat kesehatan setinggi-tingginya. Kesehatan lingkungan meliputi air, udara, tanah, pangan, sarana dan bangunan. Lingkungan sehat mencakup lingkungan pemukiman, tempat kerja, tempat rekreasi, serta tempat dan fasilitas umum (Hadi, 2019).

Metode

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh melalui publikasi Kementerian Kesehatan tahun 2021 yaitu “Profil Kesehatan Indonesia 2021” yang dapat diakses pada laman www.kemkes.go.id. Pengambilan data dilakukan pada Selasa, 30 Mei 2023 di Surabaya. Variabel yang digunakan pada penelitian ini adalah sebagai berikut.

X_1 : Persentase Sarana Air Minum Sesuai Standar (%)

X_2 : Persentase Keluarga dengan Akses Terhadap Fasilitas Sanitasi yang Layak (%)

X_3 : Persentase Desa/Kelurahan Stop Buang Air Besar Sembarangan (%)

X_4 : Persentase Desa/Kelurahan Melaksanakan Sanitasi Total Berbasis Masyarakat (%)

X_5 : Persentase Tempat dan Fasilitas Umum Dilakukan Pengawasan Sesuai Standar (%)

X_6 : Persentase Tempat Pengelolaan Pangan yang Memenuhi Syarat Sesuai Standar (%)

X_7 : Persentase Fasyankes yang Memiliki Pengelolaan Limbah Medis Sesuai Standar (%)

X_8 : Persentase Kabupaten/Kota Melaksanakan Kebijakan Gerakan Masyarakat Hidup Sehat (%)

X_9 : Persentase Rumah Tangga yang Menempati Rumah Layak Huni (%)

X_{10} : Persentase Kabupaten/Kota Sehat (%)

Hasil dan Pembahasan

Karakteristik Data

Karakteristik variabel yang diduga mempengaruhi kesehatan lingkungan di Indonesia berdasarkan provinsi tahun 2021 ditunjukkan pada Tabel 1 berikut.

Tabel 1. Karakteristik Data

Variabel	Mean	Median	Minimum	Maximum	Standard Deviation
X ₁	69.01	80	7.3	100	26.32
X ₂	83.64	87.7	3.7	100	16.85
X ₃	44.06	40.7	3.8	100	27.15
X ₄	78.59	84.9	17.9	100	21.78
X ₅	53.66	53.2	17	84.3	17.16
X ₆	51.59	50.6	16.5	81.1	13.88
X ₇	22.43	16.35	0.9	78.1	20.41
X ₈	49.99	44.2	0	100	31.10
X ₉	60.14	62.05	27.6	85.15	12.16
X ₁₀	45.39	33.8	0	100	37.10

Berdasarkan Tabel 1 dapat diketahui bahwa rata-rata tertinggi dari faktor kesehatan lingkungan adalah X₂ sebesar 83.64% yang artinya mayoritas sebesar 83.64% keluarga di Indonesia memiliki akses terhadap fasilitas sanitasi layak (X₂), hal ini menunjukkan kondisi yang baik terhadap kepedulian masyarakat terhadap kesehatan. Persentase kabupaten/kota sehat dari 34 provinsi di Indonesia memiliki kesenjangan atau perbedaan yang cukup besar antar provinsi berdasarkan deviasi standar sebesar 37.10.

Asumsi PCA dan Analisis Faktor

Asumsi yang digunakan pada PCA dan analisis faktor adalah distribusi normal multivariat, kecukupan data, dependensi, dan korelasi antar variabel berdasarkan *Measuring of Sampling Adequacy* (MSA).

Distribusi Normal Multivariat

Distribusi normal multivariat adalah distribusi dari banyak atau multi variabel yaitu berdistribusi normal. Hasil pengujian asumsi distribusi normal adalah sebagai berikut.

H₀ : Faktor kesehatan lingkungan berdistribusi normal multivariat

H₁ : Faktor kesehatan lingkungan tidak berdistribusi normal multivariat

Daerah penolakan : Tolak H₀ jika T_{proporsi} < 45% atau T_{proporsi} > 55%

Statistik uji : T_{proporsi} yang diperoleh sebesar 0.5 atau 50%

Berdasarkan nilai statistik uji T_{proporsi} sebesar 0.5 atau 50% yang berada di antara 45% ≤ T_{proporsi} ≤ 55% sehingga diputuskan gagal tolak H₀ yang artinya faktor kesehatan lingkungan berdistribusi normal multivariat.

Kecukupan Data

Asumsi kecukupan data dilakukan untuk mengetahui apakah data yang digunakan cukup atau layak untuk dilakukan PCA dan analisis faktor berdasarkan nilai *Kaiser Meyer Olkin*. Hasil pemeriksaan kecukupan data diperoleh nilai *Kaiser Meyer Olkin* (KMO) sebesar 0.664 lebih besar dari 0.5 sehingga diperoleh kesimpulan bahwa 10 variabel kesehatan lingkungan telah cukup atau layak untuk dilakukan PCA dan analisis faktor.

Depedensi

Uji depedensi menggunakan uji *Bartlett* digunakan untuk mengetahui ada tidaknya hubungan dari faktor kesehatan lingkungan. Hasil pengujian dijelaskan sebagai berikut.

Hipotesis

H₀ : $\rho = \mathbf{I}$ (faktor kesehatan lingkungan Indonesia tahun 2021 saling independen)

H₁ : $\rho \neq \mathbf{I}$ (faktor kesehatan lingkungan Indonesia tahun 2021 saling dependen)

Taraf signifikan : 5%

Daerah penolakan : tolak H₀ jika $\chi^2 > \chi^2_{(0.05; df)}$ atau jika $p\text{-value} < 5\%$.

Tabel 2. Uji *Bartlett*

χ^2	$\chi^2_{(0.05; 45)}$	$p\text{-value}$
85.259	61.656	0.00

Berdasarkan Tabel 2 hasil statistik uji *Bartlett* sebesar 85.259 lebih besar dari $\chi^2_{(0.05; 45)}$ sebesar 61.656 dan diperkuat dengan nilai *p-value* sebesar 0 sehingga diputuskan tolak H_0 yang artinya faktor kesehatan lingkungan di Indonesia tahun 2021 saling dependen atau berhubungan.

Korelasi Antar Variabel

Pemeriksaan korelasi antar variabel dilakukan untuk mengetahui apakah korelasi antar variabel kurang dari atau lebih dari 0.5, jika nilai MSA kurang dari 0.5 maka variabel tersebut dikeluarkan. Hasil pemeriksaan korelasi antar variabel ditunjukkan pada Tabel 3 berikut.

Tabel 3. Nilai MSA

Variabel	MSA	Variabel	MSA
X ₁	0.599	X ₆	0.575
X ₂	0.565	X ₇	0.689
X ₃	0.700	X ₈	0.663
X ₄	0.738	X ₉	0.636
X ₅	0.514	X ₁₀	0.824

Tabel 3 menunjukkan nilai *measuring of sampling adequacy* (MSA) dari faktor kesehatan lingkungan Indonesia tahun 2021, di mana sepuluh variabel tersebut memiliki nilai MSA lebih dari 0.5 sehingga asumsi korelasi antar variabel terpenuhi.

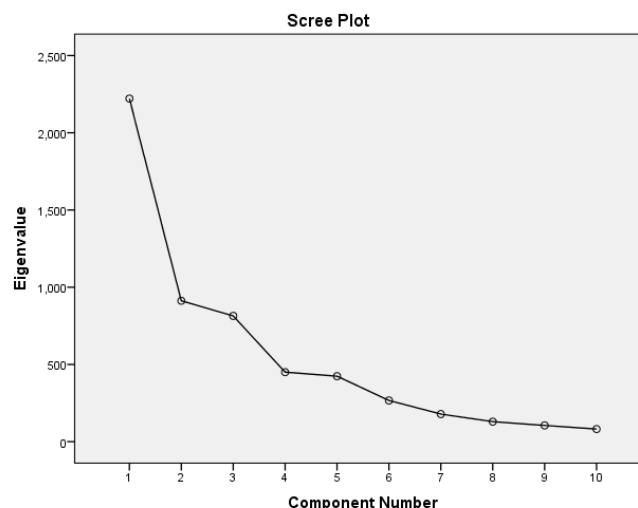
Analisis Komponen Utama (*Principal Component Analysis*)

Analisis komponen utama digunakan untuk mereduksi dimensi menjadi variabel yang lebih sedikit tanpa menghilangkan karakteristik variabel asli. Analisis komponen utama merupakan metode yang digunakan pada analisis faktor. Hasil ekstraksi komponen berdasarkan nilai *eigen* dan proporsi menggunakan matriks kovarians ditunjukkan pada Tabel 4. berikut.

Tabel 4. Nilai *Eigen* dan Proporsi

Komponen	<i>Eigen</i>	% varians	% kumulatif	Komponen	<i>Eigen</i>	% varians	% kumulatif
1	2221.664	39.792	39.792	6	266.580	4.775	91.135
2	912.176	16.338	56.129	7	912.176	3.200	94.336
3	814.270	14.584	70.714	8	814.270	2.324	96.659
4	449.884	8.058	78.771	9	449.884	1.881	98.540
5	423.724	7.589	86.360	10	423.724	1.460	100

Berdasarkan Tabel 4 dapat diketahui bahwa berdasarkan ekstraksi komponen menggunakan matriks varians kovarians terbentuk 3 komponen dengan persentase kumulatif dari 3 komponen sebesar 70.714% yang artinya 3 komponen yang terbentuk mampu menjelaskan variabel asli sebesar 70.714%. Hasil tersebut didukung berdasarkan *scree plot* pada Gambar 1 berikut.



Gambar 1. *Scree Plot*

Gambar 1 menunjukkan bahwa dari *scree plot* dapat diketahui bahwa titik belok komponen terjadi pada komponen 3, di mana setelah komponen 3 perbedaan nilai *eigen* tidak terlalu signifikan, sehingga digunakan 3 komponen baru yang menjelaskan 10 variabel.

Analisis Faktor

Analisis faktor bertujuan untuk mereduksi variabel. Berdasarkan hasil PCA dari 10 variabel terbentuk 3 komponen. Setelah dilakukan ekstraksi dapat diketahui komunalitas masing-masing variabel yang ditunjukkan pada Tabel 5 berikut.

Tabel 5. Komunalitas

Variabel	Ekstraksi	Variabel	Ekstraksi
X ₁	0.764	X ₆	0.240
X ₂	0.226	X ₇	0.353
X ₃	0.756	X ₈	0.936
X ₄	0.631	X ₉	0.122
X ₅	0.169	X ₁₀	0.968

Berdasarkan Tabel 5 dapat diketahui bahwa dari nilai komunalitas tertinggi adalah pada X₁₀ yang artinya dari 3 komponen dapat menjelaskan variabel X₁₀ sebesar 96.8%, sedangkan komunalitas terkecil adalah X₉ sebesar 0.122 yang artinya 3 komponen menjelaskan X₉ sebesar 12.2%. Analisis selanjutnya adalah *component matrix*. Rotasi bertujuan untuk mengetahui variabel yang belum tepat atau belum sesuai dengan faktor. Hasil rotasi faktor menggunakan *varimax* ditunjukkan pada Tabel 6 berikut.

Tabel 6. Rotated Component Matrix

Variabel	Komponen			Variabel	Komponen		
	1	2	3		1	2	3
X ₁	0.217	0.016	0.847	X ₆	0.488	0.032	0.028
X ₂	0.335	0.219	0.256	X ₇	0.445	0.344	0.192
X ₃	0.361	0.510	0.604	X ₈	0.943	0.102	-0.190
X ₄	0.601	0.483	0.190	X ₉	0.262	0.214	0.087
X ₅	0.106	-0.013	-0.397	X ₁₀	0.095	0.978	-0.048

Tabel 6 menunjukkan hasil *rotated component matrix* yang digunakan untuk mengelompokkan variabel berdasarkan bobot terbesar dari masing-masing variabel ke dalam salah satu komponen. Komponen 1 terdiri dari X₂, X₄, X₆, X₇, X₈, dan X₉, komponen 2 terdiri dari X₁₀, dan komponen 3 terdiri dari X₁, X₃, dan X₅. Berdasarkan pengelompokan diperoleh komponen baru atau faktor baru pada Tabel 7 berikut.

Tabel 7. Faktor Baru

Faktor	Variabel	Keterangan	Nama Faktor Baru
1	X ₂ , X ₄ , X ₆ , X ₇ , X ₈ , X ₉	Sanitasi Layak, Sanitasi Total, Pengelolaan Pangan, Pengelolaan Limbah Medis, Germas, Rumah Layak	Faktor lingkungan dan masyarakat
2	X ₁₀	Kabupaten/Kota Sehat	Faktor kab/kota sehat
3	X ₁ , X ₃ , X ₅	Air Minum Standar, Stop BAB Sembarangan, Tempat dan Fasilitas Umum	Faktor fasilitas dan masyarakat

Tabel 7 menunjukkan hasil faktor baru di mana, faktor 1 terdiri dari 6 variabel dengan nama faktor lingkungan dan masyarakat, faktor 2 terdiri dari 1 variabel yaitu faktor kabupaten/kota sehat, dan faktor 3 terdiri dari 3 variabel yaitu faktor fasilitas dan masyarakat.

Analisis Cluster Nonhierarki

Analisis *cluster* nonhierarki pada faktor kesehatan lingkungan di Indonesia tahun 2021 menggunakan metode *k-means*. Pada *cluster* nonhierarki, jumlah *cluster* ditentukan di awal sejumlah *k cluster*. Pada penelitian ini akan dibentuk sebanyak 2 *cluster*. Hasil pengelompokan provinsi menjadi 2 *cluster* ditunjukkan pada Tabel 8 berikut.

Tabel 8. Hasil Analisis Cluster Metode K-Means

Cluster	Jumlah Anggota	Anggota Cluster
1	14	Aceh, Sumut, Sumbar, DKI Jakarta, NTT, Kalbar, Kalteng, Sultengah, Sultengg, Sulbar, Maluku, Malut, Papua Barat, Papua
2	20	Riau, Jambi, Sumsel, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DIY, Jatim, Banten, Bali, NTB, Kalsel, kaltim, Kaltara, Sulsel, Gorontalo

Tabel 9 menunjukkan hasil pengelompokan provinsi berdasarkan faktor kesehatan lingkungan tahun 2021 menggunakan *k-means* sebanyak 2 *cluster*. Pada *cluster* 1 terdiri dari 14 provinsi dan *cluster* 2 terdiri dari 20 provinsi. Pada analisis *cluster* nonhierarki banyaknya *cluster* ditentukan diawal, sehingga diperlukan pemilihan jumlah *cluster* optimum berdasarkan nilai *Pseudo-F*. Tabel 10 berikut merupakan hasil nilai *Pseudo-F*.

Tabel 10. Pemilihan *Cluster* Optimum

Jumlah <i>Cluster</i>	<i>Pseudo-F</i>
2	12.096
3	10.786
4	9.572

Tabel 10 menunjukkan hasil nilai *Pseudo-F* dengan ditentukan terdapat 2, 3, dan 4 *cluster*. Berdasarkan tabel tersebut dapat diketahui bahwa dari ketiga pengelelompokkan diperoleh nilai *Pseudo-F* tertinggi yaitu dengan jumlah *cluster* sebanyak 2 sebesar 12.096. sehingga, *cluster* optimum dari pengelompokan kesehatan lingkungan di Indonesia tahun 2021 menggunakan *k-means* adalah 2 *cluster*.

Kesimpulan dan Saran

Kesimpulan yang diperoleh dari penelitian ini adalah berdasarkan hasil PCA dan analisis faktor dari 10 variabel tereduksi menjadi 3 faktor baru yaitu faktor lingkungan dan masyarakat, faktor kab/kota, dan faktor fasilitas dan masyarakat. Hasil analisis *cluster* menggunakan *K-Means* sebanyak 2 *cluster*, pada *cluster* 1 terdiri dari 14 provinsi dan *cluster* 2 terdiri dari 20 provinsi dengan jumlah *cluster* optimum berdasarkan *Pseudo-F* adalah sebanyak 2 *cluster*.

Saran dari penelitian ini untuk penelitian selanjutnya adalah dapat melakukan *clustering* berdasarkan hasil faktor baru dari analisis faktor dan dapat menggunakan metode *cluster* lain seperti *hierarchy clustering* atau metode yang digunakan pada *machine learning* seperti DBSCAN atau GMM.

Daftar Pustaka

- Arsham, H. and Lovric, M. (2011) 'Bartlett's Test', *International Encyclopedia of Statistical Science*, (December 2013). doi: 10.1007/978-3-642-04898-2.
- Hadi, S. (2019) 'Pengarusutamaan Kesehatan Lingkungan Dalam Peningkatan Kualitas Hidup Manusia', *Media Penelitian dan Pengembangan Kesehatan*, 12(2), pp. 77–78.
- Hikmah, H. *et al.* (2022) 'Analisis Klaster Pengelompokan Kecamatan di Sulawesi Barat Berdasarkan Indikator Pendidikan', *Saintifik*, 8(2), pp. 188–196. doi: 10.31605/saintifik.v8i2.383.
- Johnson, A. R. and Wichern, D. W. (2007) *Applied Multivariate Statistical Analysis*. 6th edn. London: Pearson Education, Inc.
- Kaharuddin (2020) 'Analisis Faktor Yang Mempengaruhi Kinerja Karyawan Pada Yayasan Sahabat Bunda Kota Makassar', *FORECASTING: Jurnal Ilmiah Ilmu Manajemen*, 2(1), pp. 60–75.
- Noya van Delsen, M. S., Wattimena, A. Z. and Saputri, S. (2017) 'Penggunaan Metode Analisis Komponen Utama Untuk Mereduksi Faktor-Faktor Inflasi Di Kota Ambon', *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 11(2), pp. 109–118. doi: 10.30598/barekengvol11iss2pp109-118.
- Silvia, V. (2020) *Statistika Deskriptif*. Yogyakarta: Penerbit Andi.
- Talakua, M. W., Leleury, Z. A. and Talluta, A. W. (2017) 'Analisis Cluster Dengan Menggunakan Metode Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014', *Jurnal Ilmu Matematika dan Terapan*, 11(2), pp. 119–128.