

Penerapan Metode *K-means* Pada Klasterisasi Provinsi di Indonesia Berdasarkan Indikator Indeks Kebahagiaan

Ni Wayan Rita Damayanthi

Program Studi Matematika, FMIPA, Universitas Udayana
e-mail: ritadamayanthi05@gmail.com

Ni Luh Putu Suciptawati

Program Studi Matematika, FMIPA, Universitas Udayana
e-mail: suciptawati@unud.ac.id

Ketut Jayanegara

Program Studi Matematika, FMIPA, Universitas Udayana
e-mail: ktjayanegara@unud.ac.id

I Komang Gde Sukarsa

Program Studi Matematika, FMIPA, Universitas Udayana
e-mail: gedesukarsa@unud.ac.id

Eka N. Kencana

Program Studi Matematika, FMIPA, Universitas Udayana
e-mail: i.putu.enk@unud.ac.id

I Gusti Ngurah Lanang Wijayakusuma

Jurusan Matematika, FMIPA, Universitas Udayana
e-mail: lanang_wijaya@unud.ac.id

Abstract: *The happiness index is a measure that reflects individual well-being, thus playing an important role in the development of a region. The level of happiness in Indonesia is still significantly lower compared to other ASEAN countries. In relation to efforts to improve the happiness index in Indonesia, this study applies the K-means method to cluster the 34 provinces of Indonesia based on the indicators of the happiness index for the year 2021. The data used is sourced from publications by the Indonesian Central Bureau of Statistics with seven happiness index indicators and employing the Minkowski distance. The clustering results of the 34 provinces using the K-means method obtained four clusters with a cluster accuracy value of 71 percent. Cluster 1 consists of seven provinces with a fairly high average of seven attributes, cluster 2 consists of seven provinces is a cluster with a low average level of internal and external satisfaction, cluster 3 consists of four provinces with a high average of seven attributes, and cluster 4 consists of 16 provinces is a cluster with provinces with a fairly high average level of external satisfaction, but a low level of internal satisfaction.*

Keywords: *Happiness index, K-means, Cluster*

Abstrak: Indeks kebahagiaan merupakan suatu ukuran yang mencerminkan kesejahteraan individual sehingga berperan penting dalam pembangunan suatu daerah. Tingkat kebahagiaan di Indonesia masih jauh lebih rendah dibandingkan negara-negara ASEAN lainnya. Terkait upaya peningkatan indeks kebahagiaan di Indonesia, penelitian ini mengaplikasikan metode K-means untuk mengelompokkan 34 provinsi Indonesia berdasarkan indikator indeks kebahagiaan tahun 2021. Data yang digunakan bersumber dari publikasi Badan Pusat Statistik Indonesia dengan tujuh indikator indekskebahagiaan dan menggunakan jarak Minkowski. Hasil klasterisasi 34 provinsi dengan menggunakan metode K-means diperoleh empat klaster dengan nilai akurasi klaster sebesar 71 persen. Klaster 1 beranggotakan tujuh provinsi dengan rata-rata ketujuh atribut yang cukup tinggi, klaster 2 beranggotakan tujuh provinsi merupakan klaster dengan rata-rata tingkat kepuasan internal dan eksternal yang rendah, klaster 3 beranggotakan empat provinsi yang rata-rata ketujuh atributnya tinggi, dan klaster 4 terdiri dari 16 provinsi merupakan klaster dengan provinsi-provinsi yang rata-rata tingkat kepuasan eksternalnya cukup tinggi, namun tingkat kepuasan internal yang rendah.

Kata kunci: Indeks kebahagiaan, K-means, Klaster

1. Pendahuluan

Pembangunan suatu negara dapat diukur berdasarkan tingkat kesejahteraan rakyatnya. Selama ini, pembangunan suatu negara lebih banyak dilihat atas faktor ekonomi. Namun, keberhasilan pembangunan suatu negara tidak cukup hanya dilihat dari kondisi ekonomi, karena hal tersebut belum mampu sepenuhnya untuk menggambarkan tingkat kesejahteraan rakyat yang sesungguhnya. Tingkat kesejahteraan mampu diukur menggunakan standar yang berbeda-beda pada setiap individunya (indikator subjektif), salah satunya dengan indeks kebahagiaan. Indeks kebahagiaan merupakan suatu ukuran yang mencerminkan kesejahteraan individual sehingga dijadikan unsur penting yang berperan dalam pembangunan suatu daerah. Indeks kebahagiaan sebagai ukuran pembangunan yang subjektif digunakan untuk melihat persepsi masyarakat terkait perasaan yang dialami di dalam menjalani kehidupan sehari-hari (BPS, 2021).

Berdasarkan *World Happiness Report 2022*, pada tahun 2021 Indonesia menempati peringkat ke-87 dari 146 negara di dunia. Posisi Indonesia ini jauh lebih rendah dibandingkan Negara ASEAN seperti Singapura yang berada pada urutan ke-27, Filipina di urutan ke-60, dan Malaysia berada di urutan ke-70 (Helliwell et al., 2022). Berdasarkan skala 0 sampai 100, indeks kebahagiaan Indonesia pada tahun 2021 sebesar 71,49 (BPS, 2021). Tingkat kebahagiaan seseorang di tiap-tiap daerah di Indonesia memiliki perbedaan apabila dikaitkan dengan indikator indeks kebahagiaan. Provinsi Banten dengan nilai indeks kebahagiaan sebesar 68,08 berada di posisi terendah, sedangkan Provinsi Maluku Utara menempati posisi tertinggi dengan nilai indeks kebahagiaan sebesar 76,34 (BPS, 2021). Perbedaan indeks kebahagiaan di setiap provinsi dipengaruhi

oleh beberapa faktor, diantaranya pendidikan, kesehatan, dan kemiskinan (BPS, 2021). Menurut Rahayu (2016) determinan kebahagiaan di Indonesia adalah pendapatan, pendidikan, kesehatan dan modal sosial. Dengan adanya perbedaan ini, sehingga prioritas program peningkatan kebahagiaan dan kesejahteraan di masing-masing daerah juga berbeda. Berdasarkan hal tersebut, maka dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator indeks kebahagiaan untuk mengetahui provinsi-provinsi mana yang memiliki tingkat kebahagiaan yang rendah sehingga dapat dilakukan upaya peningkatan indeks kebahagiaan guna terciptanya masyarakat yang sejahtera dan mengatasi terjadinya kesenjangan antar daerah. Pengelompokan dapat dilakukan dengan bantuan teknik statistik yaitu menggunakan analisis kluster.

Analisis kluster merupakan salah satu teknik peubah ganda yang bertujuan untuk mengelompokkan objek (responden, produk, atau entitas lain) berdasarkan sekumpulan karakteristik yang dipilih pengguna (variabel pengklasteran). Kluster yang dihasilkan harus menunjukkan keragaman intra kluster (*within cluster*) yang minimum dan keragaman antar kluster (*between cluster*) yang maksimum. Metode dalam analisis kluster secara umum ada dua, yaitu metode berhierarki dan metode tak berhierarki. Metode berhierarki menggabungkan pengamatan ke dalam hierarki atau struktur seperti pohon melalui serangkaian dari $n - 1$ (n merupakan jumlah pengamatan) keputusan pengelompokan. Sedangkan metode tak berhierarki atau metode partisi tidak melibatkan proses konstruksi menyerupai struktur pohon seperti yang dilakukan dalam metode berhierarki. Sebaliknya, metode tak berhierarki menetapkan objek ke dalam kluster setelah jumlah kluster ditentukan (Hair et al., 2018).

Mubarak & Kholijah (2023) mengaplikasikan metode *K-means* untuk mengelompokkan kabupaten/kota di Provinsi Jambi berdasarkan penyakit menular. Hasil pengklasteran terbentuk tiga kluster yaitu kluster dengan penyakit menular rendah terdiri dari enam kabupaten/kota, kluster dengan penyakit menular sedang beranggotakan tiga kabupaten, dan Kabupaten Muaro Bungo dan Kota Jambi termasuk dalam kluster yang tingkat penyakit menularnya tinggi. Selain itu, pengelompokan menggunakan metode *K-means* juga dilakukan oleh Sari & Sukestiyarno (2021) pada persebaran kasus Covid-19 di Indonesia. Tujuan penelitian yaitu untuk mengetahui tingkat persebaran kasus Covid-19 dalam kategori tinggi, sedang, dan rendah pada setiap provinsi di Indonesia tahun 2020. Berdasarkan hasil analisis diperoleh pengelompokan 34 provinsi menjadi tiga kluster. Kluster I merupakan kluster dengan kasus yang tinggi beranggotakan dua provinsi, Kluster II beranggotakan tiga provinsi termasuk kluster dengan kasus yang sedang, dan kluster III beranggotakan 29 provinsi merupakan kluster dengan kasus yang rendah.

Terdapat beberapa teknik pengklasteran dalam analisis kluster, salah satunya yang paling sering digunakan karena kesederhanaan dan prosesnya yang cepat adalah metode *K-means*. *K-means* merupakan teknik klasterisasi tak berhierarki yang mempartisi objek ke dalam dua atau lebih kluster berdasarkan nilai kedekatan objek terhadap titik pusat

klaster. Dalam metode *K-means*, pola k akan ditentukan secara acak sebagai titik awal pusat klaster. Setelah k pusat klaster awal ditentukan, selanjutnya akan dihitung jarak terdekatnya terhadap setiap objek. Jarak antara objek dengan pusat klaster yang paling minimum akan bergabung membentuk sebuah klaster baru. Proses iterasi terus berlanjut sampai tidak ada objek yang berpindah klaster (Agrawal & Gupta, 2013).

Berdasarkan permasalahan tersebut, maka dilakukan riset untuk mengetahui pengelompokan provinsi di Indonesia berdasarkan indikator indeks kebahagiaan tahun 2021 menggunakan metode *K-means*. Hasil pengelompokan diharapkan dapat memberikan gambaran kepada masyarakat dan pemerintah agar dapat meningkatkan indeks kebahagiaan di setiap provinsi serta mengatasi kesenjangan antar provinsi di Indonesia.

2. Metode Penelitian

Jenis data pada penelitian ini menggunakan data kuantitatif yang diperoleh dari publikasi Badan Pusat Statistik (BPS) Indonesia tahun 2021 sehingga termasuk dalam jenis data sekunder. Data yang digunakan adalah data menurut provinsi di Indonesia pada tahun 2021 yaitu sebanyak 34 provinsi dengan tujuh indikator indeks kebahagiaan yang terdiri dari variabel pengembangan diri, kepuasan terhadap pendidikan dan keterampilan, pendapatan rumah tangga, keadaan lingkungan, kondisi keamanan, keharmonisan keluarga, dan kesehatan (BPS, 2021).

Tahapan analisis data pada penelitian ini untuk mengelompokkan provinsi di Indonesia menurut indikator indeks kebahagiaan menggunakan *software R* yaitu sebagai berikut.

1. Melakukan pengumpulan data yang diperoleh dari publikasi Badan Pusat Statistik Indonesia tahun 2021.
2. Eksplorasi data menggunakan statistik deskriptif untuk melihat karakteristik dalam data. Salah satu bentuk penyajian data secara ringkas yang dapat menampilkan nilai-nilai statistik dari suatu variabel yaitu dengan penyajian numerik.
3. Jika satuan data yang diamati berbeda serta nilai varian dari variabel yang diamati cukup besar, maka dilakukan standarisasi data dengan *Z-score*. Standarisasi data menjadikan semua variabel mempunyai kontribusi yang sama dalam proses klasterisasi. Dengan standarisasi klaster yang terbentuk menjadi homogen dan terpisah dengan baik. Hair et al. (2018) menyatakan bahwa standarisasi data dengan *Z-score* merupakan salah satu cara yang sering digunakan yaitu dengan mengkonversi setiap variabel terhadap nilai standar.

$$z = \frac{x_i - \bar{x}}{s} \quad (1)$$

x_i menunjukkan data ke- i , \bar{x} adalah rata-rata dari data, dan s merupakan simpangan

baku dari data (Walpole & Myers, 1995).

4. Melakukan pengklasteran menggunakan metode *K-means*. MacQueen berpendapat bahwa *K-means* digunakan untuk mendeskripsikan algoritma *K-means* dengan cara memasukkan setiap objek ke dalam suatu kluster yang mempunyai rata-rata terdekat (Johnson & Wichern, 2007). Tahapan pengklasteran menggunakan metode *K-means* yaitu sebagai berikut.

- i. Menentukan jumlah kluster yang akan dibentuk dengan bantuan metode pendugaan jumlah kluster, diantaranya yaitu metode *elbow*, *silhouette*, dan *gap statistic*. Metode *Elbow* merupakan salah satu metode untuk menduga jumlah kluster optimal dengan melihat grafik perbandingan antara jumlah kluster yang digunakan dan nilai dari jumlah kuadrat intra kluster atau dikenal dengan istilah

$$WSS_{(k)} = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (2)$$

Grafik WSS umumnya akan menunjukkan pola yang turun curam kemudian melandai membentuk siku pada suatu titik. Pada titik yang membentuk siku itulah jumlah kluster dianggap sebagai kluster yang optimal karena penambahan kluster tidak akan memberikan perubahan nilai WSS yang signifikan (Bholowalia & Kumar, 2014). Selanjutnya gabungan antara metode kohesi dan separasi yang mampu menentukan kualitas suatu kluster serta mengukur kedekatan relasi antara objek dalam kluster disebut metode *Silhouette* (Kodinariya & Makwana, 2013). Kohesi mengacu pada ukuran seberapa dekat setiap data dalam kluster terhadap pusat klasternya, sedangkan separasi mengacu pada jarak antara *centroid* kluster yang berlainan. Secara umum, metode *Silhouette* mengukur seberapa layak suatu objek bergabung dalam kluster saat ini (C_i), dibandingkan saat bergabung dalam kluster lain (C_j) dan dirumuskan dengan persamaan

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

dimana $a(i)$ merupakan rata-rata jarak objek i dengan objek lainnya yang merupakan elemen dari C_i dan $b(i)$ merupakan nilai rata-rata minimum dari objek i terhadap kluster lain yang bukan kluster induknya. Ketiga yaitu metode *gap statistic* yang menduga jumlah kluster optimal dengan rumus

$$Gap_n(k) = E_n^*(\log(W_k)) - \log(W_k) \quad (4)$$

Nilai $E_n^*(\log(W_k))$ merupakan ekspektasi dari $\log(W_k)$ pada distribusi referensi dari n pengamatan. Nilai $Gap_n(k)$ yang semakin besar menunjukkan bahwa

semakin baik k kluster untuk dijadikan sebagai jumlah kluster yang optimum dalam matriks data $A_{n \times p}$ (Kencana, 2020).

- i. Menentukan k pusat awal kluster secara acak.
- ii. Menghitung jarak setiap objek ke pusat kluster dari masing-masing kluster menggunakan salah satu metrik jarak. Melihat pertimbangan dari beberapa ukuran kemiripan, jarak *Euclidean* merupakan jarak yang paling populer dan sering digunakan untuk mengelompokkan objek pada sebuah matriks data, namun jarak ini tidak cocok digunakan pada data yang mengandung pencilan. Oleh karena terdapat pencilan pada data yang digunakan dalam penelitian ini, sehingga diperlukan ukuran jarak yang lebih tahan terhadap pengaruh pencilan dalam data, salah satunya dapat menggunakan jarak *Minkowski*. Jarak *Minkowski* merupakan suatu metrik dalam ruang vektor bernorma yang digunakan untuk mengukur jarak antara objek dalam ruang geometri dengan mempertimbangkan perbedaan pada setiap dimensi komponen. Jika $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ dan $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ merupakan dua buah pengamatan yang berada pada R^p , maka jarak *Minkowski* dapat dituliskan dengan rumus berikut (Han et al., 2012).

$$d(i, j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \dots + |x_{ip} - x_{jp}|^h} \quad (5)$$

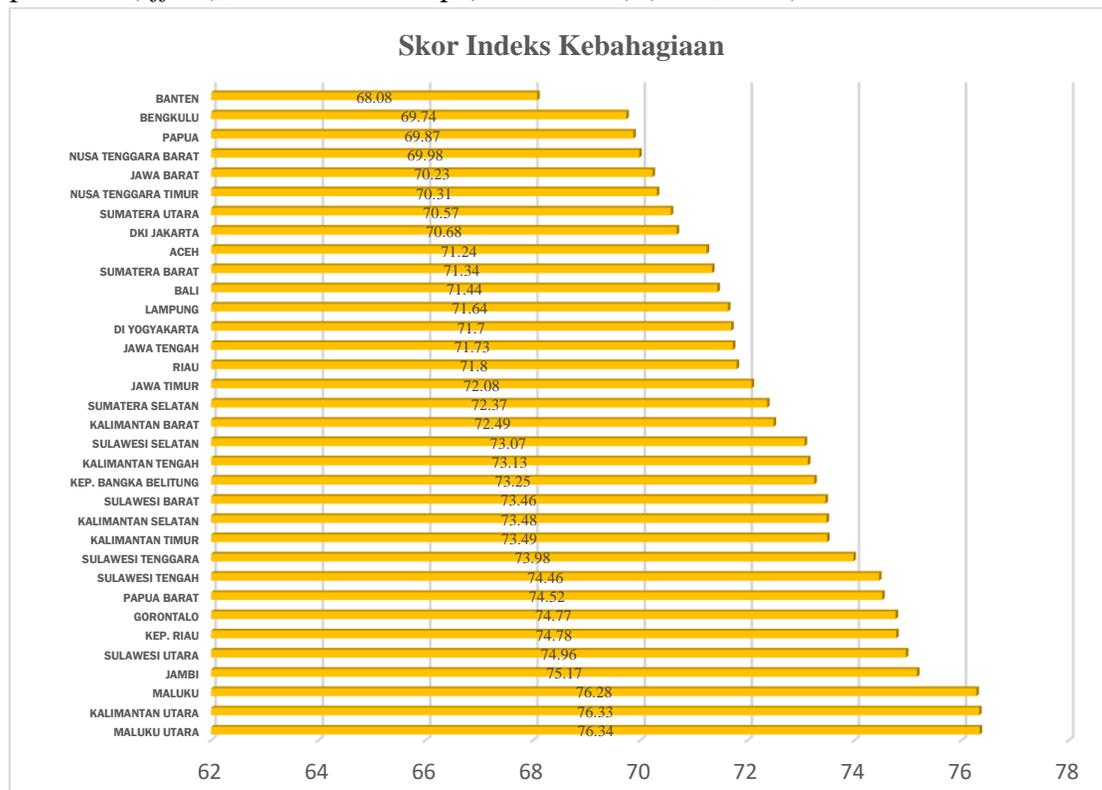
- iii. Mengalokasikan setiap objek ke kluster terdekat. Kelompokkan objek berdasarkan jarak minimum dilihat dari hasil perhitungan jarak *Minkowski*.
- iv. Menghitung pusat kluster baru dengan mencari rata-rata dari setiap kluster menggunakan persamaan

$$C_i = \frac{\sum_{i=1}^n x_i}{n} \quad (6)$$

- v. Lakukan iterasi dengan mengulangi langkah iii, iv, dan v hingga tidak ada lagi objek yang berpindah kluster.
- vi. Mengukur akurasi dari kluster yang terbentuk. Nilai akurasi dari hasil pengelompokan menunjukkan seberapa baik kluster yang dibentuk bisa merepresentasikan matriks observasi maupun hubungan antara data yang sebenarnya. Akurasi kluster dapat diperoleh dengan menghitung rasio dari jumlah kuadrat antarkluster (*between sum squares*) dengan jumlah kuadrat total (*total sum squares*).
- vii. Mendeskripsikan kluster hasil klusterisasi 34 provinsi dengan metode *K-means*.

3. Hasil dan Pembahasan

Objek pengamatan pada penelitian ini terdiri dari 34 provinsi di Indonesia dengan tujuh atribut yang menjadi indikator indeks kebahagiaan. Indeks kebahagiaan adalah suatu ukuran yang mampu menggambarkan tingkat kesejahteraan suatu daerah yang diukur berdasarkan persepsi subjektif dari masing-masing individu terhadap seluruh aspek yang memengaruhi kehidupannya. Indeks kebahagiaan Indonesia merupakan indeks komposit yang dapat dihitung menggunakan dimensi dan indikator dengan skala 0-100. Dimensi penyusun indeks kebahagiaan tersusun atas dimensi kepuasan hidup (*life satisfaction*), perasaan (*affect*), dan makna hidup (*eudaimonia*) (BPS, 2021).



Gambar 1. Skor Indeks Kebahagiaan 34 Provinsi di Indonesia Tahun 2021

Berdasarkan gambar di atas terlihat bahwa Provinsi Maluku, Kalimantan Utara, dan Maluku Utara merupakan tiga besar provinsi dengan indeks kebahagiaan tertinggi di Indonesia. Zainal Arifin Paliwang selaku Gubernur Kalimantan Utara mengatakan masyarakat Kaltara memiliki indeks kebahagiaan yang tinggi karena pemerintah yang dekat dengan rakyatnya sehingga masyarakat Kaltara mudah dalam menyampaikan aspirasinya. Selain itu, jajaran kepolisian dan masyarakat mampu menciptakan rasa aman dan kerukunan antar daerah di Provinsi Kalimantan Utara (BIROADPIM, 2022).

Provinsi Papua, Bengkulu, dan Banten merupakan tiga besar provinsi dengan indeks

kebahagiaan terendah di Indonesia. Gubernur Papua, Lukas Enembe mengatakan bahwa masyarakat Papua tidak bahagia karena kebutuhan dasar seperti kebutuhan fisiologis, rasa cinta, aktualisasi diri, keamanan, dan harga diri yang belum terpenuhi sehingga menyebabkan sering terjadinya konflik (Mara, 2022)

3.1 Eksplorasi data

Sebelum melakukan analisis kluster, akan dilakukan eksplorasi data menggunakan statistik deskriptif untuk mengetahui karakteristik dalam data. Analisis statistik deskriptif dilakukan secara numerik dengan menampilkan nilai minimum, nilai maksimum, mean, standar deviasi, dan varians.

Tabel 1. Statistik Deskriptif

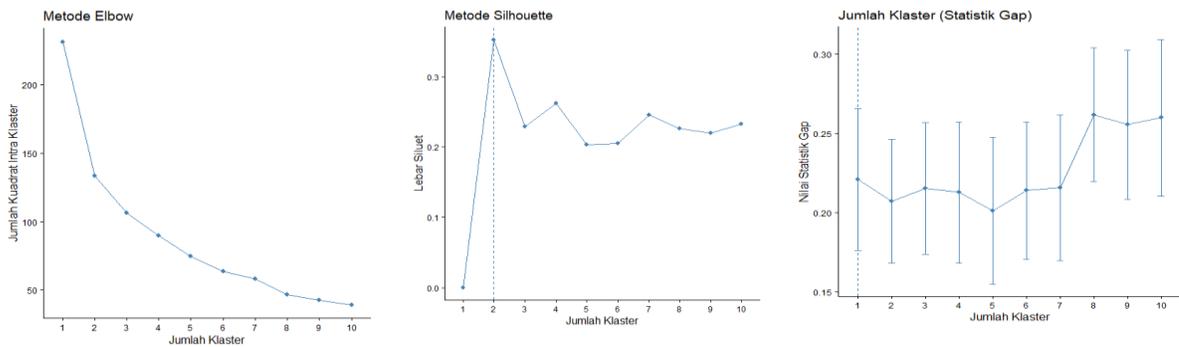
	N	Min	Max	Mean	Std. Deviation	Variance
Pendidikan	34	56.81	71.25	63.95353	3.735	13.948
Pendapatan	34	62.97	74.14	67.77176	2.928	8.574
Lingkungan	34	78.39	85.71	81.92912	1.832	3.357
Keamanan	34	76.89	86.39	82.06176	2.490	6.201
Keharmonisan	34	79.04	89.02	84.25265	2.477	6.133
Pengembangan diri	34	60.64	74.62	67.33206	3.273	10.710
Kesehatan	34	73.38	80.99	77.08441	1.881	3.537

Sumber: data diolah (2023)

Tabel 1 menunjukkan tingkat kepuasan terhadap pendidikan dan keterampilan tertinggi sebesar 71,25 dari Provinsi Maluku, sedangkan nilai terendah sebesar 56,81 berasal dari Provinsi Sulawesi Barat. Selain itu, tingkat kepuasan terhadap kesehatan tertinggi yaitu sebesar 80,99 berasal dari Provinsi Kalimantan Utara, sedangkan nilai terendah berasal dari Provinsi Banten dengan nilai sebesar 73,38. Hal tersebut juga menunjukkan bahwa setiap variabel dari indikator indeks kebahagiaan memiliki rentang nilai yang cukup besar, sehingga perlu dilakukan standarisasi data. Standarisasi data bertujuan untuk meminimalkan rentang nilai yang terlalu lebar dalam data dan setiap variabel memiliki kontribusi yang setara dalam proses klusterisasi, sehingga terbentuk kluster yang homogen dan terpisah dengan baik. Standarisasi dengan *Z-score* dilakukan sesuai dengan persamaan (1).

3.2 Pendugaan jumlah kluster

Terdapat beberapa metode yang dapat membantu dalam menentukan jumlah kluster, diantaranya dengan menggunakan metode *elbow*, *silhouette*, dan *gap statistic*.



Gambar 2. Plot pendugaan jumlah kluster optimal

Berdasarkan plot *elbow* terlihat bahwa penurunan jumlah kuadrat intra kluster yang mulai melandai pada saat jumlah kluster $k = 6$. Pada plot siluet menunjukkan bahwa lebar siluet tertinggi pada $k = 2$. Selanjutnya plot statistik gap menunjukkan nilai gap mengalami perbedaan atau kenaikan yang paling signifikan terjadi pada jumlah kluster 7 menuju 8, sehingga jumlah kluster $k = 8$.

Berdasarkan ketiga metode pendugaan jumlah kluster optimal yang telah dicobakan, diperoleh hasil yang berbeda-beda dari setiap metodenya. Mempertimbangkan jumlah objek sebanyak 34 provinsi, jika menggunakan hasil 6 dan 8 kluster seperti yang ditampilkan oleh plot *elbow* dan plot *gap statistic* maka terlalu banyak kluster yang harus dibentuk, kemudian jika menggunakan 2 kluster seperti yang ditampilkan oleh plot *siluet* maka terlalu sedikit kluster yang dibentuk. Jika terlalu sedikit kluster yang terbentuk maka, hasil pengklasteran belum mampu merepresentasikan matriks data yang sebenarnya. Sedangkan, jika terlalu banyak kluster yang terbentuk maka akan kesulitan dalam proses interpretasi. Dengan demikian, dicobakan 3, 4, 5, dan 6 kluster untuk mengakomodir 2 hingga 8 kluster yang disarankan dari ketiga metode penduga jumlah kluster.

Tabel 2. Klasterisasi metode K-means

Jumlah Kluster	Banyak Anggota Kluster	Akurasi
K=3	17, 10, 7	59,9%
K=4	16, 7, 7, 4	71,0%
K=5	11, 9, 7, 4, 3	75,9%
K=6	11, 7, 6, 4, 4, 2	79,7%

Sumber: data diolah (2023)

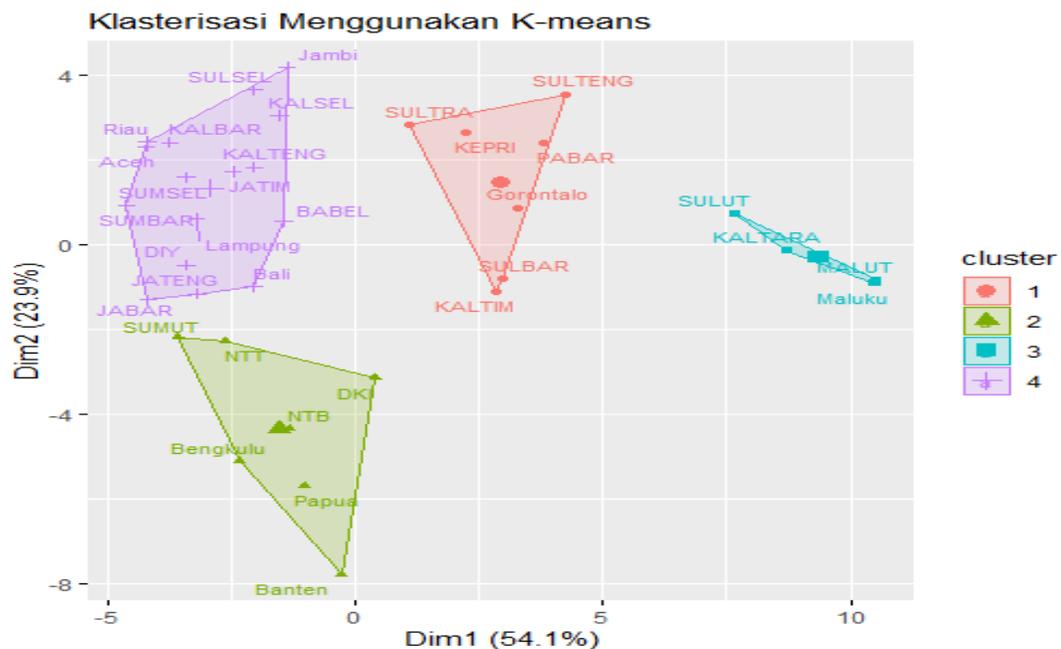
Berdasarkan jumlah kluster 3, 4, 5, dan 6 yang telah dicobakan, dipilih jumlah kluster sebanyak empat kluster yang akan digunakan untuk mengelompokkan 34 provinsi di Indonesia mempertimbangkan akurasi dari kluster yang meningkat signifikan pada $k=4$ serta berdasarkan atas ketiga metode pendugaan jumlah kluster.

Setelah menentukan jumlah kluster yang akan dibentuk, kemudian menentukan empat pusat kluster (*centroid*) awal secara acak. Selanjutnya menghitung jarak antar objek terhadap *centroid* menggunakan ukuran jarak *Minkowski* dengan persamaan (5), kemudian gabungkan objek berdasarkan jarak terdekat objek terhadap *centroid*.

Jika masih terdapat objek yang berpindah kluster atau pusat kluster berubah, maka dihitung kembali pusat kluster yang baru. Untuk menghitung *centroid* baru, dapat menggunakan persamaan (6) atau menghitung nilai rata-rata dari kluster yang baru terbentuk. Selanjutnya lakukan pengelompokan dengan langkah yang sama seperti pada iterasi sebelumnya. Proses iterasi dilakukan sampai tidak ada objek yang berpindah kluster atau saat *centroid* sudah tetap.

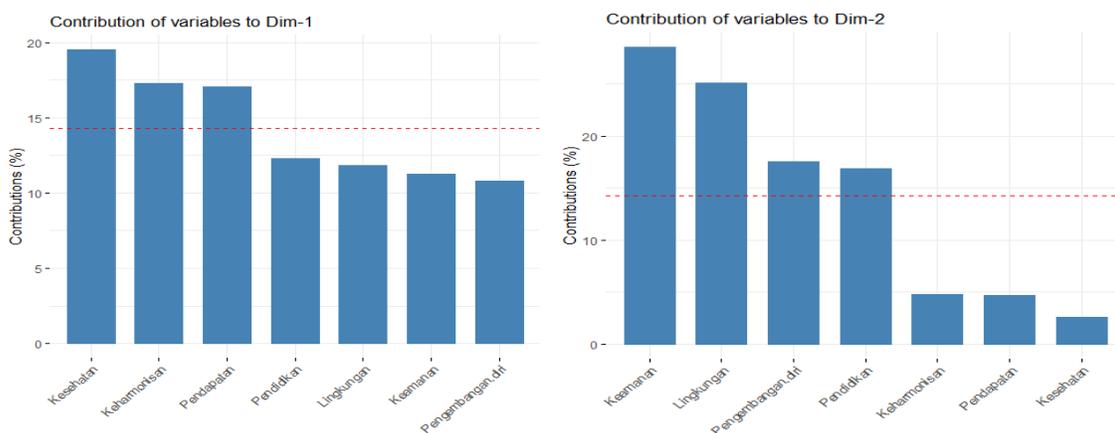
3.3 Pengelompokan menggunakan metode *K-means*

Hasil klasterisasi 34 provinsi di Indonesia menjadi empat kluster menggunakan *K-means clustering* dapat divisualisasikan pada Gambar 3.



Gambar 3. Plot klasterisasi provinsi menurut indikator indeks kebahagiaan

Sumbu x dan y merupakan komponen-komponen utama (*principal component*) yang terbentuk dari tujuh atribut. Komponen utama pertama (Dim1) menjelaskan variansi data sebesar 54.1% dan komponen utama kedua (Dim2) sebesar 23.9%. Kedua dimensi ini mampu mengekstraksi total matriks data sebesar 78.0%. Kontribusi relatif dari setiap atribut terhadap *eigen value* Dim1 dan Dim2 dapat dilihat pada Gambar 4.



Gambar 4. Kontribusi variabel pada komponen utama pertama dan kedua

Pada Gambar 4 menunjukkan ragam unit dari Dim1 atau komponen utama pertama dominan berasal dari kontribusi atribut kepuasan terhadap kesehatan, keharmonisan keluarga, dan pendapatan rumah tangga, sehingga komponen utama pertama disebut sebagai kepuasan internal. Sedangkan pada Dim2 atau komponen utama kedua, kontribusi dominan berasal dari atribut pengembangan diri, kepuasan terhadap kondisi keamanan, keadaan lingkungan, pendidikan dan keterampilan. Maka, komponen utama kedua dapat disebut sebagai kepuasan eksternal.

Dengan demikian, plot pada Gambar 3 memvisualisasikan bahwa klaster 3 memiliki tingkat kepuasan internal tertinggi dibandingkan klaster lainnya, klaster 2 memiliki tingkat kepuasan internal dan eksternal yang rendah. Kemudian klaster 4 memiliki tingkat kepuasan eksternal yang cukup tinggi, namun tingkat kepuasan internal yang rendah, serta klaster 1 memiliki tingkat kepuasan eksternal yang cukup tinggi.

Karakteristik dari masing-masing klaster dapat diketahui dengan menentukan rata-rata atribut dari setiap klaster. Rata-rata klaster hasil pengelompokan *K-means clustering* dengan tujuh atribut dapat dituliskan dalam Tabel 3.

Tabel 3. Pusat Klaster Hasil Pengelompokan K-means

Klaster	Pendi-dikan	Penda-patan	Ling-kungan	Keamanan	Keharmonisan	Pengem-bangan diri	Kesehatan
1	0.4156	0.4453	0.3263	0.5293	1.0280	0.5092	0.6523
2	-0.6542	-0.9153	-0.9545	-0.9874	-1.0163	-0.8479	-1.0110
3	1.6270	1.8188	1.4944	1.3034	1.2910	1.4416	1.8465
4	-0.3023	-0.2490	-0.0987	-0.1254	-0.3278	-0.2122	-0.3047

Sumber: data diolah (2023)

3.4 Interpretasi hasil klaster

- a) Klaster 1 beranggotakan tujuh provinsi merupakan klaster dengan rata-rata atributnya menunjukkan nilai yang cukup tinggi atau klaster dengan rata-rata

indeks kebahagiaan tertinggi kedua. Provinsi-provinsi yang tergabung dalam klaster 1 memiliki tingkat kepuasan eksternal yang cukup tinggi.

- b) Klaster 2 juga beranggotakan tujuh provinsi merupakan klaster dengan rata-rata atributnya memiliki nilai terendah dibandingkan ketiga klaster lainnya. Provinsi-provinsi pada klaster 2 memiliki rata-rata tingkat kepuasan internal dan eksternal yang sangat rendah.
- c) Klaster 3 beranggotakan empat provinsi yang dicirikan oleh rata-rata setiap atribut memiliki nilai tertinggi dibandingkan tiga klaster lainnya. Keempat provinsi yang tergabung dalam klaster 3 memiliki rata-rata tingkat kepuasan internal dan eksternal yang sangat tinggi.
- d) Klaster 4 beranggotakan 16 provinsi dengan rata-rata setiap atributnya cukup rendah. Salah satu provinsi dalam klaster 4 yaitu Provinsi Bali dengan tingkat kepuasan internal yang cukup rendah.

3.5 Akurasi klaster

Hasil pengelompokan 34 provinsi ke dalam empat klaster diperoleh nilai akurasi klaster sebesar 0,71 atau 71.0%. Nilai tersebut menunjukkan bahwa keragaman antar klaster sebesar 71 persen. Hal ini berarti dari keempat klaster yang terbentuk sudah cukup baik mewakili matriks data indikator indeks kebahagiaan di Indonesia.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan hasil analisis, pengklasteran 34 provinsi di Indonesia menggunakan metode *K-means* terbentuk empat klaster dengan akurasi klaster sebesar 71 persen. Klaster 1 merupakan provinsi-provinsi yang penduduknya dikategorikan bahagia beranggotakan tujuh provinsi dengan rata-rata ketujuh atribut yang cukup tinggi, klaster 2 beranggotakan tujuh provinsi merupakan klaster dengan rata-rata tingkat kepuasan internal dan eksternal yang rendah sehingga dikategorikan penduduknya sangat tidak bahagia, sedangkan klaster 3 beranggotakan empat provinsi yang rata-rata ketujuh atributnya tinggi klaster ini penduduknya dikategorikan sangat bahagia, dan klaster 4 merupakan klaster dengan provinsi-provinsi yang rata-rata tingkat kepuasan eksternalnya cukup tinggi, namun tingkat kepuasan internal yang rendah terdiri dari 16 provinsi.

4.2 Saran

Bagi provinsi-provinsi yang termasuk dalam klaster dengan tingkat kebahagiaan yang rendah, diharapkan pemerintah dapat membuat kebijakan serta upaya yang mampu

meningkatkan indeks kebahagiaan di provinsi-provinsi tersebut. Misalnya dengan meningkatkan kualitas pendidikan dan menyediakan berbagai fasilitas kesehatan yang memadai, serta melakukan pembangunan infrastruktur yang merata ke seluruh provinsi di Indonesia. Selain itu, setiap individu masyarakat bersikap toleransi, saling menghargai antar umat beragama, selalu bersyukur, dan melibatkan keluarga pada setiap kegiatan.

Penelitian berikutnya disarankan untuk mencoba menggunakan metode kluster tak berhierarki lainnya, seperti metode *K-medoids* yang lebih tahan terhadap pencilan. Selain itu, diharapkan menggunakan ukuran jarak lainnya seperti jarak mahalanobis karena terdapat pencilan dalam data indikator indeks kebahagiaan.

Daftar Pustaka

- Agrawal, A., & Gupta, H. (2013). Global K-Means (GKM) Clustering Algorithm: A Survey. *International Journal of Computer Applications*, 79(2), 20–24.
<https://doi.org/10.5120/13713-1472>
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 17–24.
- BIROADPIM. (2022). Kaltara Teratas soal Indeks Kebahagiaan di Indonesia, Simak Tanggapan Gubernur. Diakses pada 29 Juli 2023, dari <https://adpim.kaltaraprov.go.id/2022/01/14/kaltara-teratas-soal-indeks-kebahagiaan-di-indonesia-simak-tanggapan-gubernur/>
- BPS. (2021). *Indeks Kebahagiaan 2021*. Jakarta: Badan Pusat Statistik RI.
- Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis* (8 th ed.). United Kingdom: Cengage Learning EMEA.
<https://doi.org/10.1002/9781119409137.ch4>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques. Data Mining: Concepts and Techniques* (3 rd). Waltham: Morgan Kaufmann.
<https://doi.org/10.1016/C2009-0-61819-5>
- Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., & Aknin, L. B., & Wang, S. (2022). *World Happiness Report 2022*. New York: Sustainable Development Solutions Network. Diambil dari <http://worldhappiness.report/>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6 th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New Jersey: John Wiley & Sons, Inc.
- Kencana, E. N. (2020). Sains Data dengan R: Klusterisasi Menggunakan K-Means Clustering. Diambil dari <https://www.researchgate.net/publication/341787776> Sains
- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90–95.
- Mara, S. R. E. (2022). Bedah Pernyataan Gubernur Provinsi Papua, Lukas Enembe :

- Orang Papua Tidak Happy “Bahagia”, Kenapa?. Diakses pada 29 Juli 2023, dari <https://sentaninews.id/2022/02/14/bedah-pernyataan-gubernur-provinsi-papua-lukas-enembe-orang-papua-tidak-happy-bahagia-kenapa/>
- Mubarak, H., & Kholijah, G. (2023). Analisis Klaster Dalam Pengelompokan Kabupaten/Kota Di Provinsi Jambi Berdasarkan Penyakit Menular Menggunakan Metode K-Means. *Jurnal Statistika dan Komputasi*, 2(1), 20–30. <https://doi.org/10.32665/statkom.v2i1.1597>
- Rahayu, T. P. (2016). Determinan Kebahagiaan di Indonesia. *Jurnal Ekonomi dan Bisnis*, 19(1), 149–170. <https://doi.org/10.24914/jeb.v19i1.485>
- 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. Diambil dari <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Walpole, R. E., & Myers, R. H. (1995). *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan* (4 th). Bandung: ITB.