



## Penerapan *Silhouette Coefficient*, *Elbow Method* dan *Gap Statistics* untuk Penentuan *Cluster Optimum* dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan

Affifa Atira<sup>1</sup>, Betha Nurina Sari<sup>2</sup>

<sup>1,2</sup>Universitas Singaperbangsa Karawang

---

### Abstract

Received: 10 Juli 2023  
Revised: 14 Agustus 2023  
Accepted: 21 Agustus 2023

*Currently, the assessment of regional development success is still limited to economic growth and poverty factors. However, evaluation based solely on economic factors is not accurate enough because high economic growth does not always guarantee the happiness of society and can even worsen social inequality. Therefore, one of the important priorities in development is to create economic growth that can improve the welfare of society equally, without creating gaps between social groups. The well-being and social progress in a region can be influenced by the happiness of its people. Happiness can be used as a measure to assess the welfare obtained by individuals. In this study, clustering of happiness index based on provinces in Indonesia was conducted using the K-Means algorithm in RStudio. Through the K-Means clustering analysis, the results of this study can be used as a reference by the government in formulating strategic plans or making improvements to increase the level of happiness and welfare of the Indonesian society. There are three methods to determine the optimal number of clusters: Silhouette method, Elbow method, and Gap Statistics. In the comparison of methods, it is observed that the optimal cluster formation occurs with 2 clusters, where the Elbow and Silhouette methods yield the best results. The data was processed using the K-Means method, resulting in 2 groups: 16 provinces with a high-level happiness group and 18 provinces with a low-level happiness group. From the evaluation using the Silhouette Index, a value of 0.4005945 was obtained at  $k=2$ , indicating that this cluster falls into the weak structure category.*

**Keywords:** *K-Means Algorithm, Clustering, RStudio, Happiness*

(\*) Corresponding Author: [1910631170061@student.unsika.ac.id](mailto:1910631170061@student.unsika.ac.id)

**How to Cite:** Atira, A., & Sari, B. N. (2023). Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan. <https://doi.org/10.5281/zenodo.8282638>

---

### PENDAHULUAN

Kebahagiaan adalah kondisi perasaan positif dan kepuasan yang dirasakan oleh seseorang dalam kehidupannya. Kebahagiaan dapat diartikan sebagai upaya memenuhi keinginan dan aspirasi dengan cara yang benar dan memuaskan secara nyata dan signifikan (Banusu & Firmanto, 2020). Kondisi kebahagiaan dapat tercipta dari berbagai hal, seperti pencapaian tujuan hidup, hubungan sosial yang positif, kesehatan fisik dan mental yang baik, serta pemenuhan kebutuhan dasar seperti pangan, papan, dan sandang. Kebahagiaan sering diukur melalui indikator subjektif seperti kepuasan hidup, tingkat stres, dan emosi positif.

Berbagai banyak faktor yang mempengaruhi tingkat kebahagiaan suatu wilayah seperti pendidikan, pekerjaan, kesehatan, dan lainnya. Menurut penelitian yang telah dilakukan oleh (Putra & Sudibia, 2019), faktor yang memengaruhi tingkat kebahagiaan seseorang adalah pendapatan, harapan, interaksi sosial,

keyakinan, sikap bersyukur, perilaku peduli lingkungan, kesehatan, jenis kelamin, modal sosial, dan budaya. Selain itu, Kumalasari & Yasa juga mengatakan bahwa jumlah penduduk, tingkat kebebasan dari korupsi, dan migrasi netto memiliki dampak yang signifikan terhadap tingkat kebahagiaan secara bersamaan (Kumalasari & Yasa, 2020).

Saat ini, penilaian kesuksesan pembangunan suatu wilayah masih terbatas pada faktor-faktor pertumbuhan ekonomi dan kemiskinan. Namun, evaluasi berdasarkan faktor ekonomi saja tidak cukup akurat, karena pertumbuhan ekonomi yang tinggi tidak selalu menjamin kebahagiaan masyarakat, bahkan dapat memperburuk ketimpangan sosial. Maka dari itu, salah satu prioritas penting dalam pembangunan adalah menciptakan pertumbuhan ekonomi yang dapat meningkatkan kesejahteraan masyarakat secara merata, tanpa menimbulkan kesenjangan antara kelompok masyarakat. Kesejahteraan dan kemajuan sosial di suatu wilayah dapat dipengaruhi oleh kebahagiaan masyarakat. Kebahagiaan dapat dijadikan sebagai ukuran untuk menilai kesejahteraan yang diperoleh oleh individu. Negara yang berhasil mencapai kesejahteraan yang tinggi umumnya membuat warganya merasa bahagia karena kesejahteraan menandakan bahwa kebutuhan masyarakat telah tercukupi, sehingga kesempatan untuk merasa bahagia pun semakin besar (Kumalasari & Yasa, 2020).

Indeks Kebahagiaan adalah suatu ukuran yang mengindikasikan sejauh mana tingkat kebahagiaan masyarakat dalam suatu negara. Indeks ini mencerminkan kesejahteraan subyektif berdasarkan beberapa aspek kehidupan yang dianggap penting dan memiliki makna bagi masyarakat, serta dapat digunakan sebagai pelengkap dari indikator obyektif lainnya (Badan Pusat Statistik, 2022). Dalam kurun waktu tahun 2017 hingga 2021, Indeks Kebahagiaan dihitung berdasarkan tiga dimensi utama, yakni Kepuasan Hidup, Perasaan, dan Makna Hidup.

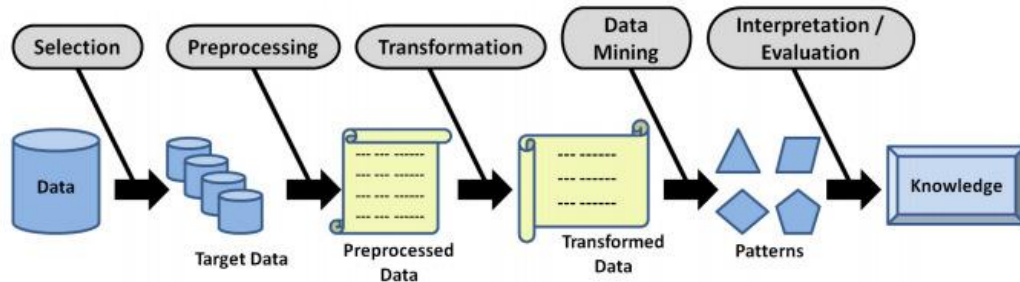
Penelitian sebelumnya dilakukan oleh Handoyono menggunakan algoritma *K-Means* dalam mengelompokkan kualitas perekonomian dari 34 provinsi Indonesia menghasilkan 3 *cluster* dibuat untuk mengelompokkan kumpulan objek dengan karakteristik yang serupa (Handoyono, 2022).

Perbandingan antara algoritma *K-Means* dan *K-Medoids* dalam pengelompokkan tingkat kebahagiaan provinsi di Indonesia yang dilakukan oleh (Fathia Palembang et al., 2022). Penelitian ini membahas tentang pengelompokan provinsi-provinsi berdasarkan indeks kebahagiaan tahun 2021 dengan menggunakan dataset provinsi-provinsi berdasarkan indeks kebahagiaan dengan jumlah 34 provinsi melalui perbandingan 2 algoritma.

Penelitian ini menggunakan algoritma *K-Means* dalam pengelompokkan provinsi-provinsi di Indonesia berdasarkan indeks kebahagiaan tahun 2014, 2017, dan 2021. Dalam penentuan klaster terbaik, dilakukan perbandingan dengan menggunakan metode *Silhouette Coefficient*, *Elbow Method*, dan *Gap Statistics*. Hasil *clustering* kemudian dievaluasi menggunakan metode *Silhouette*. Hasil pengelompokkan tersebut akan digunakan sebagai dasar dalam pengambilan keputusan dalam pembangunan wilayah yang lebih berorientasi pada kesejahteraan dan kebahagiaan masyarakat.

## **METODOLOGI PENELITIAN**

Penelitian ini menggunakan pendekatan metodologi *Knowledge Discovery in Database* (KDD) yang terdiri dari lima fase siklus hidup, seperti yang terlihat pada Gambar 1.



**Gambar 1.** Tahapan-tahapan KDD  
(Sumber: (Gustientiedina et al., 2019))

*a. Selection*

Tahap ini melibatkan pemilihan data yang akan digunakan dalam analisis. Tujuan dari tahap ini adalah untuk memilih subset data yang relevan dan penting agar proses analisis dapat lebih efisien dan efektif.

*b. Preprocessing*

Tahap ini melibatkan membersihkan data dari kesalahan, *missing values*, dan *noise*. Selain itu, tahap *preprocessing* juga dapat melibatkan transformasi data, seperti *scaling* atau normalisasi data, dan reduksi dimensi data, seperti *Principal Component Analysis* (PCA) atau *Feature Selection*.

*c. Transformation*

Tahap ini melibatkan transformasi data ke dalam format yang lebih sesuai untuk proses analisis. Contohnya, transformasi data dapat dilakukan dengan mengubah data numerik ke dalam bentuk kategori atau mengubah data kategori ke dalam bentuk numerik.

*d. Data Mining*

Tahap ini melibatkan penerapan algoritma *data mining* untuk mengidentifikasi pola atau hubungan dalam data. Tujuan dari tahap ini adalah untuk menemukan pengetahuan baru dari data yang telah diproses pada tahap sebelumnya.

*e. Interpretation/Evaluation*

Pada tahapan ini, dilakukan evaluasi terhadap hasil yang telah ditemukan sebelumnya, dan kemudian menginterpretasikan hasil tersebut. Tujuan tahap ini dilakukan adalah untuk memverifikasi bahwa hasil yang ditemukan telah memenuhi tujuan bisnis yang telah ditetapkan sebelumnya.

**KAJIAN TEORI**

***Data Mining***

Dalam *data mining*, informasi yang berharga dan berguna dapat diambil dari kumpulan data yang besar melalui suatu proses yang tepat. Tujuan utama dari *data mining* adalah untuk menemukan pola dan hubungan yang tersembunyi di dalam data yang dianalisis. *Data mining* adalah suatu bidang yang menggabungkan teknik-teknik dari berbagai disiplin ilmu, termasuk pembelajaran mesin, pengenalan pola, statistika, basis data, dan visualisasi data. Dalam konteks ini, tujuan yang ingin dicapai adalah untuk menangani kesulitan dalam mengambil

informasi dari sebuah *database* yang kompleks dan besar (Utomo & Mesran, 2020). Oleh karena itu, *data mining* sangat penting dalam mengembangkan pemahaman yang lebih baik tentang data dan membantu organisasi dalam pengambilan keputusan yang lebih tepat dan efektif.

### **Clustering**

*Clustering* adalah suatu metode dalam *data mining* yang digunakan untuk mengelompokkan data yang memiliki karakteristik serupa atau memiliki pola yang mirip. Tujuan utama dari *clustering* adalah untuk mempartisi sekelompok objek menjadi beberapa kelompok, dimana kelompok yang paling optimal adalah kelompok yang memiliki tingkat kesamaan yang tinggi di antara objek-objek dalam kelompok tersebut dan memiliki tingkat perbedaan yang signifikan dengan objek-objek pada kelompok lainnya (Ali, 2019).

### **Algoritma K-Means**

*K-Means* merupakan algoritma *clustering* yang paling populer dan sering digunakan dalam berbagai aplikasi seperti analisis data, pengenalan pola, dan pengelompokan data. Algoritma ini bekerja dengan cara membagi suatu dataset menjadi  $k$  kelompok berdasarkan jarak antara objek-objek dalam dataset tersebut. Kelompok dibentuk untuk data yang memiliki karakteristik serupa, sementara data dengan karakteristik yang berbeda dikelompokkan secara terpisah. Tujuan utamanya adalah meminimalkan variasi dalam kelompok, sehingga data yang ada dalam satu kelompok memiliki kesamaan yang tinggi (Ali, 2019). Langkah-langkah berikut ini dapat diikuti untuk melakukan clustering dengan menggunakan teknik *K-Means* (Fatmawati & Windarto, 2018):

1. Menentukan jumlah kelompok ( $k$ ) yang diinginkan.
2. Melakukan inisialisasi pusat cluster (*centroid*).
3. Objek atau data yang ada akan dikelompokkan ke dalam kluster yang terdekat. Untuk menghitung jarak antara setiap data dengan *centroid*, digunakan rumus *Euclidean distance*. Dengan menggunakan rumus ini, kita dapat menemukan jarak terpendek antara setiap data dengan *centroid*. Di bawah ini adalah persamaan pertama rumus *Euclidean Distance*:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Keterangan:

$D(i, j)$  = Jarak antara data pada titik  $i$  dan  $j$

$X_{ki}$  = Data ke  $i$  atribut data ke  $k$

$X_{kj}$  = Titik data *centroid* ke  $k$  pada atribut data ke  $j$

4. Dilakukan perhitungan ulang untuk menentukan pusat kelompok dengan mempertimbangkan anggota kelompok saat ini. Pusat kelompok dihitung dengan cara mengambil rata-rata dari seluruh data/objek yang termasuk dalam kelompok tertentu, menggunakan persamaan 2 sebagai berikut:

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk})$$

Keterangan:

$R_k$  = Rata rata  $k$

$N_k$  = Jumlah data pada *cluster*  $k$

$X_{nk}$  = Pola pada urutan ke  $n$  yang termasuk dalam *cluster*  $k$

5. Ulangi langkah nomor 3 jika terdapat perubahan pada pusat *cluster*.

#### **Metode Elbow**

Metode *elbow* (siku) adalah sebuah pendekatan yang digunakan dalam analisis klustering untuk menentukan jumlah optimal dari klaster dalam data. Jika terdapat sudut atau penurunan yang paling signifikan antara nilai klaster pertama dan klaster kedua dalam grafik, maka nilai klaster tersebut dapat dianggap yang terbaik (Sulistiyawan et al., 2021).

#### **Silhouette Coefficient**

Koefisien *silhouette* adalah sebuah metrik yang digunakan untuk mengevaluasi kualitas pengelompokan (*clustering*) dalam analisis data. Metrik ini mengukur seberapa baik setiap objek data cocok dengan kelompoknya sendiri dibandingkan dengan kelompok lainnya. Koefisien *silhouette* menggabungkan konsep kohesi (*cohesion*) dan pemisahan (*separation*) dalam pengelompokan data. Rentang nilai koefisien *silhouette* adalah dari -1 hingga 1, dan sistem pengelompokan data dikatakan baik ketika nilai koefisien *silhouette* mendekati 1 (Syamhuri et al., 2022).

**Tabel 1** Kategori Evaluasi Nilai Koefisien *Silhouette*

Nilai <i>Silhouette Coefficient</i>	Kategori
$0,7 < s \leq 1$	<i>Strong structure</i>
$0,5 < s \leq 0,7$	<i>Medium structure</i>
$0,25 < s \leq 0,5$	<i>Weak structure</i>
$s \leq 0,25$	<i>No structure</i>

## **HASIL DAN PEMBAHASAN**

Data yang digunakan dalam penelitian ini meliputi indeks kebahagiaan tahun 2014, 2017, dan 2021 dan data tersebut diperoleh dari *website* Badan Pusat Statistik (BPS) (<https://www.bps.go.id/>) dalam bentuk data sekunder. Total jumlah *record* dari data yang dipakai sebanyak 34. Data tersebut diproses dengan memanfaatkan bahasa pemrograman R dan *software* RStudio agar sesuai dengan kebutuhan penelitian.

### **a. Selection**

Tahapan ini merupakan proses awal untuk pemilihan data sebelum diproses dan dianalisis ke tahap *clustering*. Pemilihan atribut ini disesuaikan dengan kebutuhan dan tujuan yang telah ditetapkan dalam proses data mining. Data yang digunakan yaitu indeks kebahagiaan dari 34 provinsi di Indonesia. Atribut yang digunakan adalah indeks kebahagiaan tahun 2014, 2017, dan 2021. Data tersebut tersimpan dalam format *file* excel (.xlsx atau .xls). Sebagian dataset yang telah dipilih dapat dilihat dalam Gambar 2.

Provinsi	ik_2014	ik_2017	ik_2021
ACEH	67.48	71.96	71.24
SUMATERA UTARA	67.65	68.41	70.57
SUMATERA BARAT	66.79	72.43	71.34
RIAU	68.85	71.89	71.8
JAMBI	71.1	70.45	75.17

**Gambar 2** Data Awal Indeks Kebahagiaan 2014, 2017, dan 2021

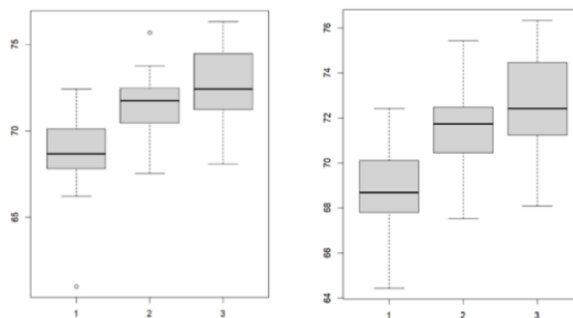
### **b. Preprocessing**

Pada langkah *preprocessing*, terdapat 1 data yang bernilai *null* pada atribut indeks kebahagiaan 2014. *Missing value* diatasi dengan imputasi *mean* pada data. Alasan data tersebut tidak dihapus adalah dikarenakan data yang dimiliki bersifat sensitif. Keterangan atribut yang terdapat *missing value* dapat dilihat pada Gambar 3.

```
> summary(data)
  Provinsi      ik_2014      ik_2017      ik_2021
Length:34      Min.   :60.97    Min.   :67.52    Min.   :68.08
Class :character 1st Qu.:67.81    1st Qu.:70.49    1st Qu.:71.27
Mode  :character Median :68.66    Median :71.74    Median :72.43
      Mean :68.79    Mean   :71.47    Mean   :72.61
      3rd Qu.:70.11  3rd Qu.:72.47    3rd Qu.:74.34
      Max.   :72.42    Max.   :75.68    Max.   :76.34
      NA's   :1
```

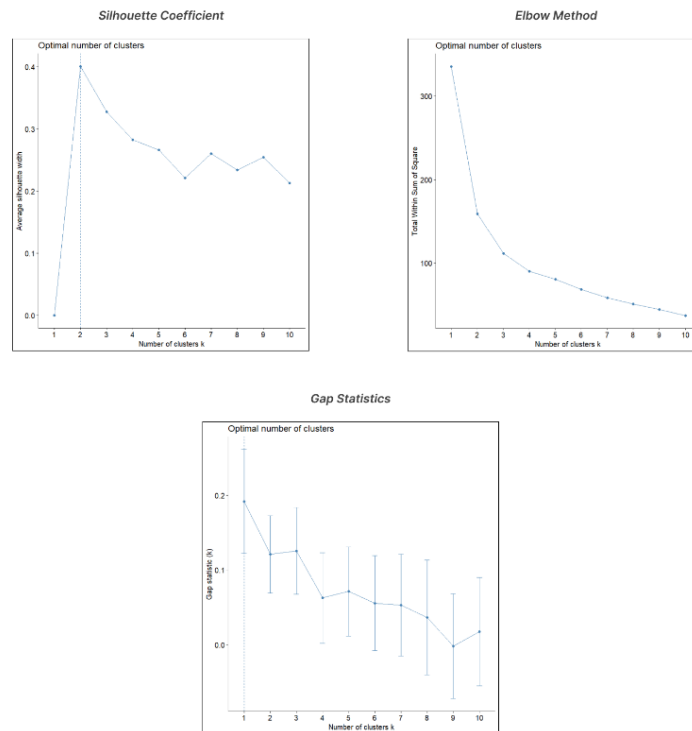
**Gambar 3** Atribut yang mempunyai *Missing Value*

Setelah mengatasi *missing value*, selanjutnya dilakukan tahap pengecekan adanya *outlier* pada atribut yang digunakan. *Outlier* sering kali dianggap sebagai nilai yang tidak representatif atau anomali dalam dataset. Pengecekan *outlier* dilakukan dengan menggunakan boxplot. Hasilnya terdapat dua provinsi yang mengandung *outlier*, diantaranya yaitu Papua (indeks kebahagiaan 2014) dan Maluku Utara (indeks kebahagiaan 2017). *Outlier* tersebut kemudian diatasi dengan menggunakan metode *Interquartile Range (IQR)*. Gambar 4 menunjukkan visualisasi *outlier* dengan keadaan sebelum dan setelah diatasi dengan IQR melalui boxplot.



**Gambar 4** Visualisasi *outlier* sebelum dan setelah diatasi dengan IQR melalui boxplot

Langkah selanjutnya adalah menentukan jumlah kluster optimal dengan menggunakan 3 metode yakni *silhouette coefficient*, *elbow method*, dan *gap statistics* untuk mendapatkan pemahaman perbandingan yang lebih komprehensif tentang struktur dan karakteristik data yang sedang dianalisis.



**Gambar 5** Jumlah Kluster Optimal Berdasarkan *Silhouette Coefficient*, *Elbow Method*, dan *Gap Statistics*

Gambar 5 merupakan informasi grafik jumlah kluster optimal berdasarkan 3 metode. Hasil dari ketiga metode tersebut menghasilkan jumlah  $k$  yang berbeda-beda. Dalam metode *silhouette*, nilai  $k$  yang optimal dapat ditentukan dengan melihat grafik di mana garis memiliki tinggi paling tinggi dan berada di garis putus-putus. Dalam kasus ini, grafik menunjukkan bahwa titik kedua memiliki nilai tertinggi, yang menunjukkan bahwa nilai  $k$  yang optimal adalah 2. Serupa dengan *silhouette*, pada *elbow method* terlihat bahwa terdapat sebuah garis yang mengalami patahan yang membentuk siku (*elbow*) pada saat  $k=2$ . Sedangkan pada metode *gap statistic* menunjukkan titik awal di mana terjadi peningkatan yang pertama kali terindikasi terletak pada  $k=1$ .

**c. Transformation**

Tahap transformasi data melibatkan proses perubahan data dengan maksud agar data dapat diproses menggunakan algoritma *K-Means Clustering*. Pada data penelitian ini, tidak dilakukan transformasi data dikarenakan data sudah dalam skala yang sama atau tidak memiliki perbedaan skala yang signifikan antara atributnya, sehingga normalisasi tidak diperlukan.

**d. Data Mining**

Langkah selanjutnya, pengklasteran akan dilakukan dengan metode *K-Means*. Untuk mengembangkan model algoritma *K-Means*, penulis menggunakan bahasa pemrograman R dan RStudio sebagai alat pengolah datanya. Pada perbandingan jumlah kluster sebelumnya, diambil  $k=2$  untuk melanjutkan proses clustering.

```

K-means clustering with 2 clusters of sizes 16, 18

Cluster means:
   ik_2014 ik_2017 ik_2021
1  70.16701 72.55313 74.35688
2  67.76222 70.50000 71.05833

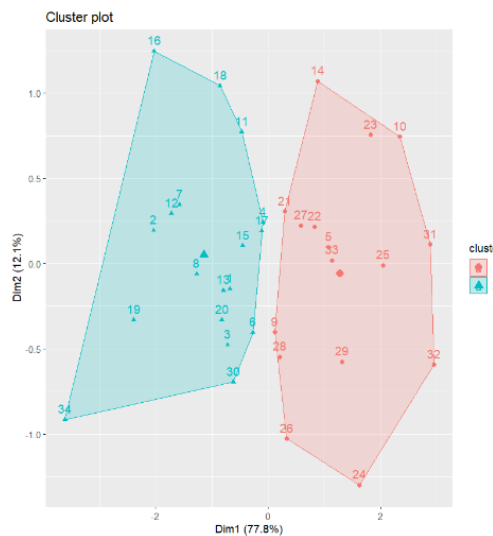
Clustering vector:
[1] 2 2 2 2 1 2 2 2 1 1 2 2 2 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 2 1 1 1 2

Within cluster sum of squares by cluster:
[1] 76.69499 82.01936
 (between_SS / total_SS =  52.7 %)

Available components:
 [1] "cluster"  "centers"  "totss"    "withinss" "tot.withinss" "betweenss"
 [7] "size"     "iter"     "ifault"   "silinfo"  "nbclust"  "data"
    
```

**Gambar 6** Clustering dengan menggunakan K-Means pada k=2

Dari hasil yang ditampilkan pada Gambar 6, didapatkan hasil pengelompokan 2 cluster dimana cluster dengan tingkat indeks kebahagiaan tingkat tinggi (C1) terdiri dari 16 item, dan cluster dengan tingkat indeks kebahagiaan tingkat rendah (C2) terdiri dari 18 item. Gambar 7 menunjukkan cluster plot hasil analisis kluster menggunakan metode K-Means.



**Gambar 7** Cluster Plot Indeks Kebahagiaan

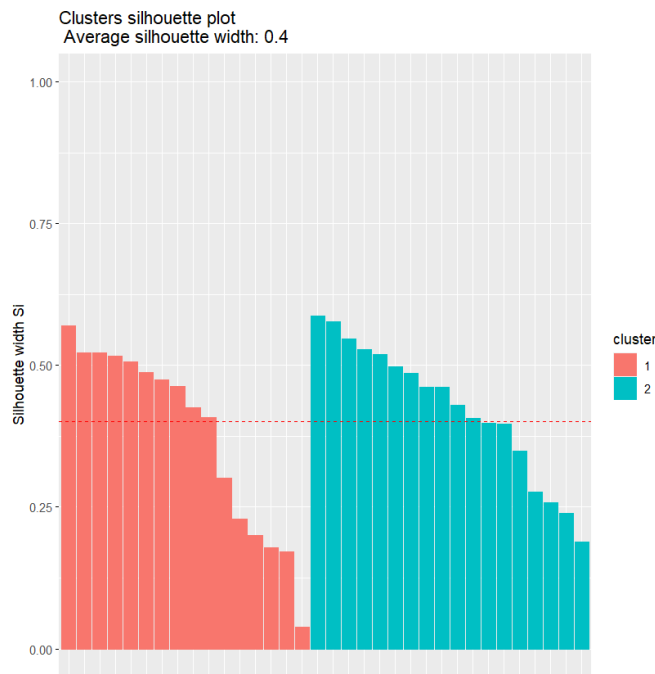
Dengan penentuan 2 klaster, penjabaran dari cluster provinsi berdasarkan indeks kebahagiaan di Indonesia dapat dilihat pada Gambar 8.

Cluster	Jumlah	Provinsi
Cluster 1	16	Jambi, Kep. Bangka Belitung, Kep. Riau, DI Yogyakarta, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat
Cluster 2	18	Aceh, Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Bengkulu, Lampung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Sulawesi Barat, Papua

**Gambar 8** Hasil Pengelompokan Clustering

**e. Evaluation**

Dari tiga metode yang telah diterapkan untuk menentukan jumlah kluster, diperoleh hasil bahwa nilai k optimal untuk analisis kluster menggunakan metode *K-Means* adalah 2 kluster. Berdasarkan hasil tersebut, dilakukan validasi kluster untuk memastikan apakah nilai k optimal 2 telah sesuai. Dalam penelitian ini, validasi dilakukan melalui evaluasi dengan menggunakan metode *Silhouette Coefficient*.



**Gambar 9** Hasil Visualisasi *Cluster Silhouette Plot*

Berdasarkan Gambar 9, dapat dilihat tampilan visualisasi *cluster silhouette plot* dari hasil pengelompokan data indeks kebahagiaan dengan rata-rata nilai *silhouette* yang didapatkan sebesar 0,4. Warna merah mengartikan *cluster 1* dan warna hijau toska mengartikan *cluster 2*.

**Tabel 2** Nilai *silhouette* k=2 hingga k=10

<b>Cluster</b>	<b>Nilai Silhouette</b>
K2	0.4005945
K3	0.3273596
K4	0.2605719
K5	0.2306365
K6	0.2306941
K7	0.2275777
K8	0.2269647
K9	0.2188343
K10	0.1997154

Proses evaluasi dilakukan untuk memeriksa apakah nilai k yang digunakan dalam tahap klusterisasi dengan algoritma *K-Means* sudah merupakan kluster yang paling optimal. Sebagai perbandingan, evaluasi dilakukan pada kluster bernilai 2 hingga 10. Dari hasil yang tercantum dalam Tabel 2, ditemukan nilai *silhouette*

maksimum sebesar 0,4005945 dimana termasuk kedalam kategori *weak structure*. Nilai *silhouette* terjadi penurunan secara berurutan, dimana nilai *silhouette* terendah terjadi saat jumlah *cluster* 10, dengan nilai *silhouette* sebesar 0,1997154.

## KESIMPULAN

Dari hasil penelitian yang dilakukan dapat disimpulkan bahwa penerapan metode *K-Means* untuk *clustering* provinsi di Indonesia berdasarkan indeks kebahagiaan berhasil diterapkan. Pemilihan jumlah kluster optimal dilakukan dengan menggunakan metode *Silhouette Coefficient*, *Elbow Method*, dan *Gap Statistic*. Hasil yang didapat dari perbandingan metode untuk menentukan jumlah kluster optimal menunjukkan bahwa metode *silhouette* dan *elbow* paling efektif pada penelitian ini. Pada proses klusterisasi, ditemukan bahwa nilai dari *silhouette* dan *elbow* yang paling tinggi terjadi pada klusterisasi *K-Means* dengan  $k = 2$ . Pada tahap evaluasi, didapatkan hasil sebesar 0,4005945 dengan nilai *silhouette* maksimum dimana termasuk kedalam kategori *weak structure*. Dari *cluster* tersebut diperoleh 16 provinsi berada di *cluster* indeks kebahagiaan tingkat tinggi yakni Jambi, Kalimantan Tengah, Kalimantan Selatan, DI Yogyakarta, Kalimantan Timur, Kalimantan Utara, Gorontalo, Maluku, Maluku Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Kep. Bangka Belitung, Kep. Riau, dan Papua Barat. Selanjutnya sebanyak 18 provinsi berada di *cluster* indeks kebahagiaan tingkat rendah yakni Aceh, Sumatera Utara, Sumatera Barat, Riau, Sumatera Selatan, Kalimantan Barat, Sulawesi Barat, Papua, Jawa Barat, Nusa Tenggara Barat, dan Nusa Tenggara Timur, Jawa Tengah, Jawa Timur, Bengkulu, Lampung, DKI Jakarta, Banten, dan Bali. Dari hasil yang telah didapatkan, diharapkan informasi tersebut dapat digunakan sebagai acuan oleh pemerintah dalam merumuskan rencana strategis atau melakukan perbaikan untuk meningkatkan tingkat kebahagiaan masyarakat Indonesia.

## DAFTAR PUSTAKA

- Ali, A. (2019). Klusterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 186–195. <https://doi.org/10.30812/matrik.v19i1.529>
- Badan Pusat Statistik, (2022). *Statistik Indonesia 2022*. Jakarta: Badan Pusat Statistik. ISSN: 0126-2912
- Banusu, Y., & Firmanto, A. D. (2020). *Kebahagiaan Dalam Ruang Keseharian Manusia*.
- Fathia Palembang, C., Yahya Matdoan, M., & Permatasari Palembang, S. (2022). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Tingkat Kebahagiaan Provinsi Di Indonesia. *Jurnal Multidisiplin Ilmu*, 01(5), 830–839. <https://journal.mediapublikasi.id/index.php/bullet/article/download/1135/550>
- Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi (Vol. 3, Issue 2). <https://doi.org/https://doi.org/10.24114/cess.v3i2.9661>

- Gustientiedina, Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Handoyono, N. A. (2022). Apakah Semakin Tinggi Ipm Akan Semakin Bahagia? Analisis Kluster Ditinjau Dari Kualitas Perekonomian. *AKUNTANSI DEWANTARA*, 6(3), 1–11. <https://doi.org/https://doi.org/10.26460/ad.v6i3.12220>
- Kumalasari, D. A., & Yasa, I. G. W. M. (2020). Pengaruh Faktor-Faktor Yang Mempengaruhi Tingkat Kebahagiaan Negara Di Dunia. *E-JURNAL EKONOMI PEMBANGUNAN UNIVERSITAS UDAYANA*, 9(5), 963–992.
- Putra, G. B. B., & Sudibia, I. K. (2019). Faktor-Faktor Penentu Kebahagiaan Sesuai Dengan Kearifan Lokal Di Bali.
- Sulistiyawan, E., Hapsery, A., & Junita Ayu Arifahanum, L. (2021). Perbandingan Metode Optimasi Untuk Pengelompokan Provinsi Berdasarkan Sektor Perikanan Di Indonesia (Studi Kasus Dinas Kelautan dan Perikanan Indonesia). *JURNAL GAUSSIAN*, 10(1), 76–84. <https://doi.org/https://doi.org/10.14710/j.gauss.10.1.76-84>
- Syamhuri, W. B., Furqon, M. T., & Dewi, C. (2022). Pengelompokan Film Berdasarkan Alur Cerita menggunakan Metode Self Organizing Maps dan Silhouette Coefficient (Vol. 6, Issue 12). <http://j-ptiik.ub.ac.id>
- Utomo, D. P., & Mesran. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(2), 437–444. <https://doi.org/10.30865/mib.v4i2.2080>