

# Perbandingan Analisis Kluster *K-Means* dan *Average Linkage* untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah

Dahayu Widyadhana <sup>a,\*</sup>, Rina Budi Hastuti <sup>a</sup>, Iqbal Kharisudin <sup>a</sup>, Fatkhurokhman Fauzi <sup>b</sup>

<sup>a</sup> Universitas Negeri Semarang, Semarang 50229, Indonesia

<sup>b</sup> Universitas Muhammadiyah Semarang, Semarang 50272, Indonesia

\* Alamat Surel: [dahayuwidya@students.unnes.ac.id](mailto:dahayuwidya@students.unnes.ac.id)

## Abstrak

Kemiskinan mendefinisikan suatu kondisi dimana ketidakmampuan untuk memenuhi kelangsungan hidup minimum. Kemiskinan merupakan persoalan multi dimensi, dan merupakan suatu persoalan yang dihadapi di banyak negara berkembang. Di Indonesia pada tiap daerahnya memiliki data kemiskinan yang berbeda, salah satunya adalah Jawa Tengah. Badan Pusat Statistik (BPS) melaporkan presentase kemiskinan di Jawa Tengah sebesar 10,8% dari jumlah penduduk. Pada penelitian ini digunakan beberapa variabel diantaranya garis kemiskinan, jumlah penduduk miskin, jumlah keluarga pra sejahtera dan pengeluaran perkapita. Penelitian ini bertujuan untuk membandingkan pengklasteran kemiskinan di Jawa Tengah dengan metode *K-Means* dan *Average Linkage*. Evaluasi metode dilakukan dengan membandingkan nilai *Silhouette Coefficient* masing-masing metode. Berdasarkan hasil penelitian, jumlah kluster terbaik yang dihasilkan adalah sebanyak 2 kluster, dengan nilai *Silhouette Coefficient* yang diperoleh menggunakan metode *Average Linkage* adalah 0,35, sedangkan dengan metode *K-Means* adalah sebesar 0,2. Dengan demikian, model terbaik yang diperoleh untuk pengklasteran kemiskinan di Jawa Tengah adalah dengan metode *Average Linkage* dengan nilai *Silhouette Coefficient* yang lebih tinggi dibandingkan metode *K-Means* yaitu 0,35.

## Kata kunci:

Kemiskinan, kluster, *K-means*, *average linkage*.

© 2021 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

## 1. Pendahuluan

Dalam rangka mendorong masyarakat untuk meningkatkan kesejahteraan masyarakat miskin agar dapat menikmati pertumbuhan ekonomi yang lebih berkualitas, pengentasan kemiskinan menjadi prioritas utama pembangunan nasional saat ini. Menurut Che *et al.* (2021) kemiskinan mendefinisikan bahwa rumah tangga tidak memiliki kemampuan yang cukup untuk memenuhi kelangsungan hidup minimum dan pembangunan. Kemiskinan merupakan persoalan multi dimensi, dan merupakan suatu persoalan yang dihadapi di banyak negara berkembang. Di Indonesia sendiri pada tahun 2019 memiliki jumlah penduduk yang besar yaitu 267 juta jiwa, Badan Pusat Statistik (BPS) mengatakan pada tahun 2019 banyaknya penduduk miskin di Indonesia adalah sebesar 24,79 juta jiwa atau 9,2% dari total penduduk di Indonesia. Di Pulau Jawa sendiri Provinsi Jawa Tengah menempati peringkat kedua dengan presentase kemiskinan 10,8%, yang mana pada peringkat pertama diperoleh oleh Provinsi Daerah Istimewa Yogyakarta dengan presentase kemiskinan 11,7%. Oleh sebab itu, fokus kita pada penelitian ini adalah Provinsi Jawa Tengah karena presentase kemiskinannya yang cukup tinggi.

Ada beberapa indikator yang dapat menentukan tingkat kemiskinan suatu daerah, menurut Itang (2017) garis kemiskinan merupakan tolok ukur yang ditetapkan sebagai kriteria kemiskinan. Saleh (2002) menggunakan salah satu faktor penentu tingkat kemiskinan dengan variabel pengeluaran pemerintah untuk sumber daya fisik perkapita per provinsi. Pada tahun 2004 Badan Koordinasi Keluarga Berencana Nasional

## To cite this article:

Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Kluster *K-Means* dan *Average Linkage* untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika* 4, 584-594

(BKKBN) menggunakan standar kesejahteraan keluarga untuk mengukur kemiskinan dan membaginya ke dalam lima kelompok dan kelompok yang perlu mendapat perhatian ekstra adalah keluarga prasejahtera dan Badan Pusat Statistik (BPS) untuk mengukur kemiskinan dengan menggunakan jumlah penduduk miskin.

Analisis *cluster* adalah metode multivariat yang dirancang untuk mengklasifikasikan objek berdasarkan karakteristik yang sama (Hair *et al.*, 2006). Ada dua metode dalam analisis *cluster*, yaitu metode hirarki dan non hirarki. Metode hirarki merupakan metode yang membuat sebuah dekomposisi berhirarki (tingkatan) dari himpunan data berdasarkan kesamaan karakteristik objeknya. Sedangkan metode non hirarki digunakan untuk pengelompokan objek, dimana jumlah *cluster* yang akan dibentuk dapat ditentukan sebelumnya. Terdapat beberapa jenis analisis *cluster* yang menggunakan metode hirarki, antara lain *single linkage*, *complete linkage*, *average linkage*, *centroid*, *ward* dan *median cluster*. Sedangkan dalam analisis *cluster* non hirarki, metode paling umum digunakan adalah metode *K-Means*.

Metode-metode analisis *cluster* hirarki dan non hirarki sudah banyak digunakan oleh berbagai penelitian, seperti yang dilakukan oleh Desy *et al.* (2016) melakukan penelitian dengan melakukan analisis *cluster* dengan menggunakan metode *K-Means* dan *Fuzzy C-Means* untuk mengelompokkan obligasi korporasi, menghasilkan kesimpulan bahwa pengelompokan data obligasi dengan menggunakan metode *K-Means* lebih tepat dibandingkan dengan metode *Fuzzy C-Means* karena memiliki nilai rasio Sw/Sb yang lebih kecil yaitu 0,6651. Kemudian, Silvia N., Wahyuningsih *et.al* (2016) menentukan hasil analisis *cluster* dengan studi kasus produksi palawija di Provinsi Kalimantan Timur 2014/2015 adalah dengan membandingkan metode *complete linkage* dan *average linkage*. Disimpulkan bahwa dengan menggunakan *complete linkage* dan *average linkage* keduanya membentuk 4 kelompok, dengan nilai rasio simpangan baku Sw terhadap Sb memperlihatkan nilai rasio simpangan baku (R) metode *average linkage* lebih kecil dibandingkan pada *complete linkage* yaitu 0,056. Oleh karena itu, dalam hal ini metode *average linkage* merupakan metode terbaik dibandingkan *complete linkage*.

Perbandingan kelebihan dan kekurangan metode *Average Linkage* dan *K-Means* menurut Govender & Sivakumar (2020).

**Tabel 1.** Kelebihan dan Kekurangan Tiap Metode

Metode Clustering	Kelebihan	Kekurangan
<i>K-Means</i>	<ul style="list-style-type: none"> <li>▪ Kompleksitas rendah</li> <li>▪ Perhitungannya cepat</li> <li>▪ Dapat menangani data yang besar</li> <li>▪ Anggota dalam klaster dapat disesuaikan</li> </ul>	<ul style="list-style-type: none"> <li>▪ Perlu menentukan jumlah klaster terlebih dahulu</li> <li>▪ Sensitif terhadap pencilan</li> <li>▪ Tidak mampu digunakan untuk klaster yang bervariasi</li> <li>▪ Sensitif terhadap skala data</li> <li>▪ Sentroid awal yang berbeda menghasilkan hasil yang berbeda</li> </ul>
<i>Average Linkage</i>	<ul style="list-style-type: none"> <li>▪ Tidak mengharuskan penentuan berapa banyak klaster</li> <li>▪ Dendrogram memberikan gambaran grafis</li> <li>▪ Dapat mendeteksi macam-macam bentuk dan ukuran klaster</li> </ul>	<ul style="list-style-type: none"> <li>▪ Kompleksitas tinggi</li> <li>▪ Proses perhitungannya lambat</li> <li>▪ Setelah klaster terbentuk tidak dapat dilakukan penyesuaian</li> <li>▪ Sulit dalam menentukan klaster yang dianggap tidak bermakna</li> <li>▪ Klaster bergantung pada jarak metrik yang digunakan</li> </ul>

Pada penelitian ini dibandingkan hasil analisis *cluster* dengan metode *K-Means* dan *Average Linkage* pada kasus kemiskinan di Jawa Tengah. Evaluasi metode dilakukan dengan membanding tingkat kebaikan masing-masing metode. Dari hasil tingkat kebaikan tersebut kemudian ditentukan mana metode *cluster* tepat.

## 2. Metode

### 2.1. *K-Means*

*K-Means* merupakan metode yang mencoba mempartisi data menjadi dua atau lebih kelompok menggunakan nilai rata-rata sebagai pusat *cluster* (Oktarina, Notodiputro *et.al*, 2020). *K-Means* umum digunakan pada *data mining* (Long *et al.*, 2021). Sebagai contoh  $X = \{x_1, x_2, \dots, x_n\}$  adalah data yang akan

dianalisis dan  $V = \{v_1, v_2, \dots, v_n\}$  adalah titik pusat dari kelompok data  $X$  pada dimensi ( $\mathbb{R}^p$ ). Dalam hal ini  $n$  menyatakan banyak objek,  $p$  menyatakan banyak variabel, dan  $k$  adalah jumlah dari partisi atau kelompok (Cebeci & Yildiz, 2015). Titik pusat dari *cluster* dapat dihitung menggunakan persamaan (1) berikut.

$$v_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{kj} \quad (1)$$

dengan  $v_{ij}$  adalah titik pusat dari kelompok ke- $i$  pada variabel  $j$ ,  $n_i$  adalah jumlah objek yang termasuk dalam kelompok ke- $i$ ,  $x_{kj}$  adalah nilai pengamatan dari objek  $k$  hingga variabel  $j$ . Pengelompokan data dapat ditulis sebagai berikut.

$$\mu_{ik} = \begin{cases} 1, & d = \min\{d^2_{ik}(x_k, v_i)\} \\ 0, & d \text{ yang lainnya} \end{cases} \quad (2)$$

dengan  $\mu_{ik}$  adalah nilai dari objek  $k$  anggota dalam *cluster*  $i$ ,  $v_i$  adalah nilai rata-rata pusat dari kelompok ke  $i$ ,  $x_k$  adalah objek pengamatan ke- $k$ . Tujuan dari pengelompokan ini untuk mengurangi keragaman dalam suatu kelompok dan memaksimalkan keanekaragaman antar kelompok. Dengan kata lain, pengelompokan ini bertujuan untuk meminimalkan fungsi objektif (Cebeci & Yildiz, 2015). Secara matematis fungsi dari tujuan tersebut dapat ditulis dengan persamaan berikut.

$$J(X, V) = \sum_{k=1}^{n_i} \sum_{i=1}^c \mu_{ik} d^2_{ik} \quad (3)$$

dengan  $n_i$  adalah jumlah dari objek yang termasuk dalam kelompok ke- $i$ ,  $c$  adalah jumlah dari *cluster*,  $d^2_{ik}$  adalah nilai jarak *eulidean* antara objek ke pusat *cluster* ke- $i$ .

## 2.2 Average Linkage

Salah satu metode *cluster* hirarki adalah *average linkage* yang merupakan metode berdasarkan pada jarak rata-rata antar seluruh objek dalam sebuah *cluster* dengan seluruh objek pada *cluster* lain, yang dapat dihitung dengan menggunakan persamaan sebagai berikut.

$$d_{(AB)C} = \frac{\sum_u \sum_v d_{uv}}{N_{(AB)}N_C} \quad (4)$$

menurut Johnson (2007)  $d_{uv}$  merupakan jarak antar objek  $u$  pada *cluster* ( $AB$ ) dan objek  $v$  pada *cluster*  $C$ . Sedangkan  $N_{(AB)}$  dan  $N_C$  berturut-turut merupakan jumlah objek dalam *cluster* ( $AB$ ) dan ( $C$ ).

## 2.3 Silhouette Coefficient

Dalam mengavaluasi hasil pengelompokan hal yang dapat dilakukan adalah validasi *cluster*. Validasi *cluster* berguna untuk mengukur tingkat kebaikan dari *cluster* yang terbentuk, salah satu *internal validation index* adalah *Silhouette Coefficient* (Azuri et al., 2016).

## 3. Hasil dan Pembahasan

Data yang digunakan untuk analisis diperoleh melalui situs resmi Badan Pusat Statistik (BPS) Jawa Tengah. Terdapat beberapa variabel yang diamati, yaitu di antaranya variabel garis kemiskinan, jumlah penduduk miskin, jumlah keluarga pra sejahtera dan pengeluaran perkapita sebagai input dalam melakukan analisis *cluster*. Sebelum melakukan *clustering* dilakukan terlebih dahulu standarisasi data dengan tujuan untuk menyamakan satuan jadi nilai standar dan tidak lagi bergantung pada satuan pengukuran, dan juga uji multikolinearitas agar menghindari terjadinya korelasi yang kuat antara dua atau lebih variabel kelompok.

	GK	PM	Prasejahtera	Pengeluaran
GK	1.00000000	-0.2808036	-0.2689415	0.04547019
PM	-0.28080357	1.00000000	0.4922094	0.14361144
Prasejahtera	-0.26894148	0.4922094	1.00000000	0.21635246
Pengeluaran	0.04547019	0.1436114	0.2163525	1.00000000

**Gambar 1.** Hasil uji multikolinearitas

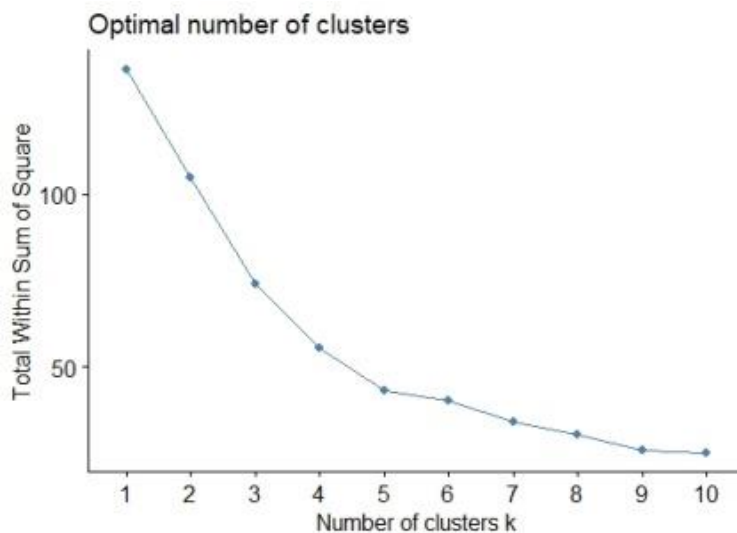
Jika nilai korelasi melebihi 0,8 maka hal itu menandakan bahwa terjadi multikolinearitas (Gujarati, 2003). Berdasarkan hasil analisis mengenai uji multikolinearitas pada Gambar 1 dapat disimpulkan bahwa tidak terdapat multikolinearitas pada data dalam penelitian ini. Dengan demikian, karena data sudah dilakukan standarisasi dan bebas dari asumsi multikolinearitas maka selanjutnya dapat dilakukan analisis *cluster* dengan metode *K-Means* dan *Average Linkage*.

### 3.1 Metode *K-Means*

Sebelum memasuki analisis *cluster*, hal pertama yang harus dilakukan adalah menentukan banyak *cluster*  $k$  optimal. Terdapat 3 metode yang dapat digunakan dalam menentukan  $k$  optimal sebagai berikut.

- Metode *elbow*

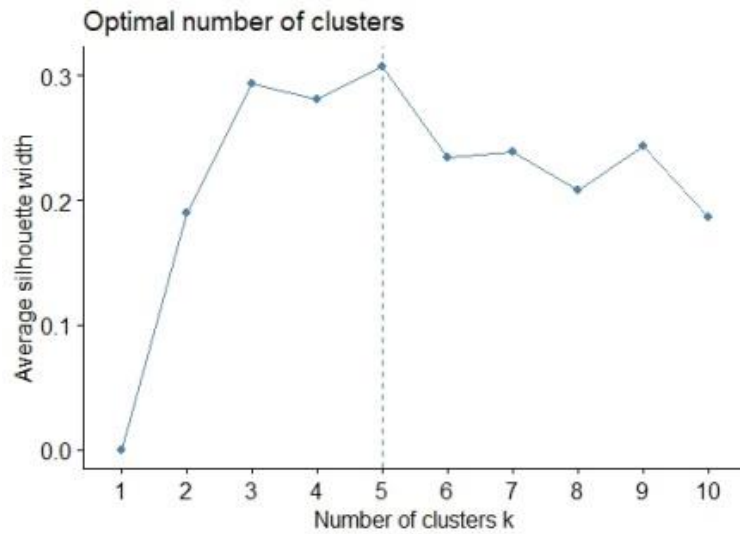
Pada metode *elbow* nilai *Sum of Square Error* (SSE) yang mengalami penurunan signifikan dan berbentuk siku diambil untuk menentukan nilai *cluster* terbaik.



**Gambar 2.** Visualisasi  $k$  optimal metode *elbow*

- Metode *silhouette*

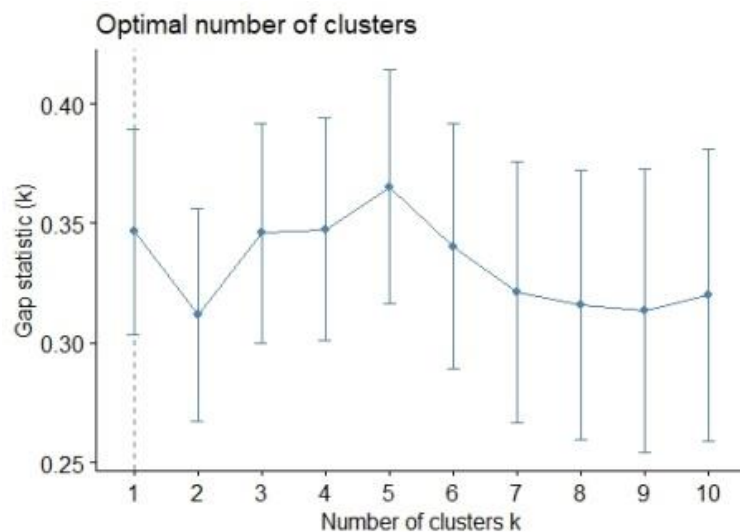
Metode ini digunakan untuk mengukur seberapa baik suatu objek diposisikan dalam suatu *cluster* yang digunakan untuk melihat kualitas dan kekuatan *cluster*. Untuk menentukan jumlah  $k$  optimalnya dengan cara melihat nilai yang paling tinggi.



**Gambar 3.** Visualisasi k optimal metode *silhouette*

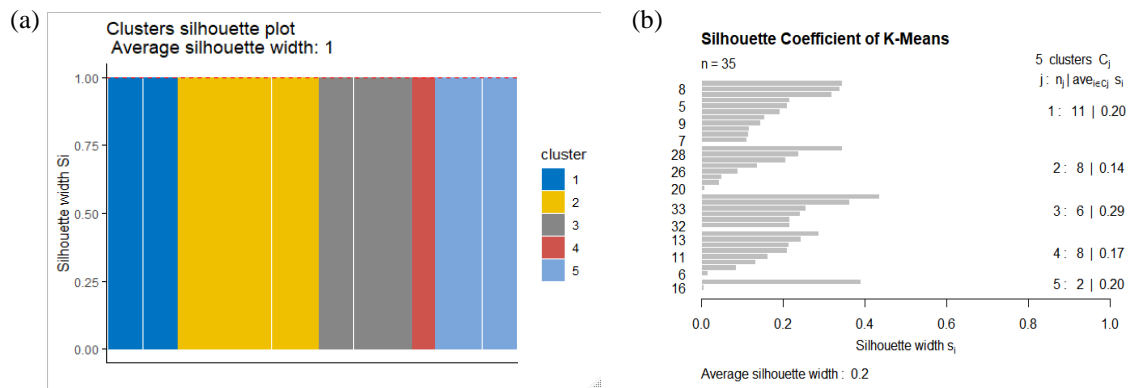
- Metode *gap statistic*

Metode *gap statistic* bertujuan untuk menentukan jumlah *cluster* lebih konstan dibandingkan pengukuran lainnya. Untuk penentuan nilai k optimal menggunakan metode *gap statistic* adalah dapat dilihat dari garis yang paling tinggi.



**Gambar 4.** Visualisasi k optimal metode *gap statistic*

Dari ketiga metode yang digunakan dalam menentukan banyak *cluster* di dapat hasil nilai k optimal untuk analisis *cluster K-Means* adalah sebesar 5 *cluster*. Berdasarkan hasil tersebut, dilakukan validasi klaster untuk menguji apakah nilai k optimal sebesar 5 sudah tepat.



**Gambar 5.** (a) visualisasi nilai *silhouette*; (b) visualisasi *silhouette coefficient*

Berdasarkan hasil pada Gambar 5 di atas menunjukkan rata-rata nilai *silhouette* nya adalah 1 yang artinya semua wilayah di Jawa Tengah berada di *cluster* yang tepat, dengan tingkat kebaikan dari *clusternya* ditinjau dari *silhouette coefficient* yang terbentuk adalah 0,2.

Setelah mendapatkan jumlah *cluster* optimal dibuat *cluster* dari kemiskinan di Provinsi Jawa Tengah. Diperoleh plot dari analisis *cluster* menggunakan *K-Means* adalah sebagai berikut.



**Gambar 6.** *Cluster plot* kemiskinan

Berdasarkan Gambar 6, didapatkan hasil pengklasteran kemiskinan di Jawa Tengah sebagai berikut.

**Tabel 2.** Hasil pengklasteran kemiskinan di Jawa Tengah dengan metode *K-Means*

Cluster	Wilayah	Identifikasi
1	Kabupaten Cilacap, Kabupaten Banyumas, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Kebumen, Kabupaten Wonosobo, Kabupaten Magelang, Kabupaten Boyolali, Kabupaten Wonogiri, Kabupaten Sragen , Kabupaten Temanggung.	<i>Cluster</i> ini memiliki angka garis kemiskinan paling rendah.

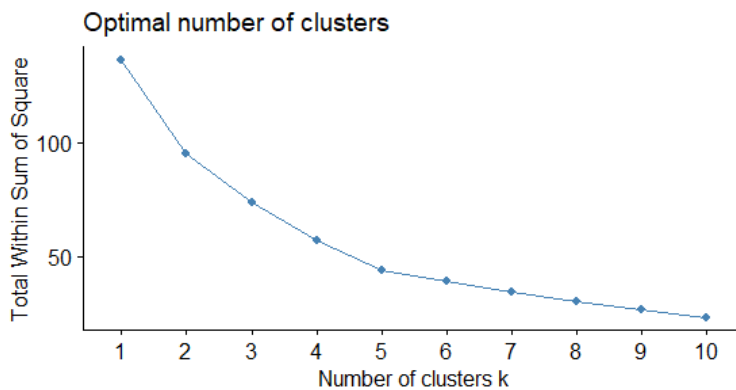
2	Kabupaten Klaten, Kabupaten Jepara, Kabupaten Demak, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Pemalang, Kabupaten Tegal, Kabupaten Brebes	Jumlah penduduk miskin dan pengeluaran per kapita di <i>cluster</i> ini paling tinggi di atas rata-rata.
3	Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal	Berkebalikan dengan <i>cluster</i> 1, wilayah di <i>cluster</i> ini memiliki angka garis kemiskinan paling tinggi di atas rata-rata, namun memiliki jumlah penduduk miskin paling rendah di antara yang lain.
4	Kabupaten Purworejo, Kabupaten Sukoharjo, Kabupaten Karanganyar, Kabupaten Rembang, Kabupaten Pati, Kabupaten Kudus, Kabupaten Semarang, Kabupaten Kendal	Garis kemiskinan di <i>cluster</i> ini memiliki nilai yang tinggi di atas rata-rata setelah <i>cluster</i> 3.
5	Kabupaten Grobogan, Kabupaten Blora	Di <i>cluster</i> ini memiliki angka keluarga pra sejahtera paling tinggi.

### 3.2 Metode Average Linkage

Pertama ditentukan terlebih dahulu jumlah *cluster* yang optimal/ideal dengan menggunakan 3 metode sama seperti metode *K-Means* sebelumnya.

- Metode *elbow*

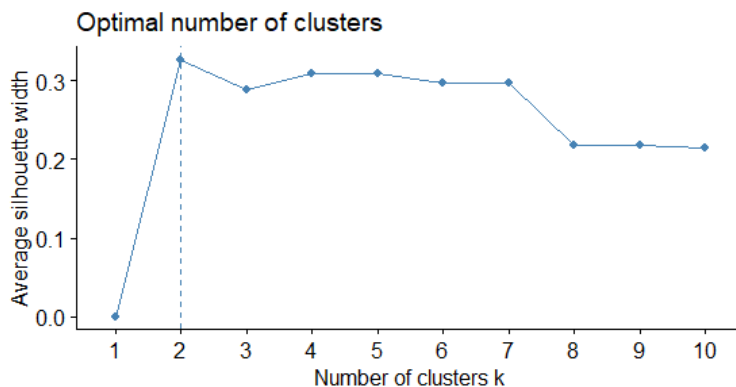
Dalam menentukan nilai *k* optimal dengan metode *elbow* dapat dilihat dari grafik yang penurunannya mulai landai.



**Gambar 7.** Visualisasi *k* optimal metode *elbow*

- Metode *silhouette*

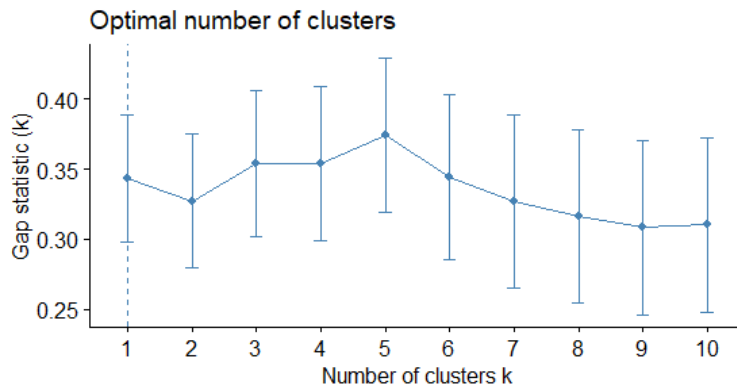
Menentukan nilai *k* optimal pada metode *silhouette* dapat ditentukan melalui garis paling tinggi atau dengan melihat garis yang paling optimal.



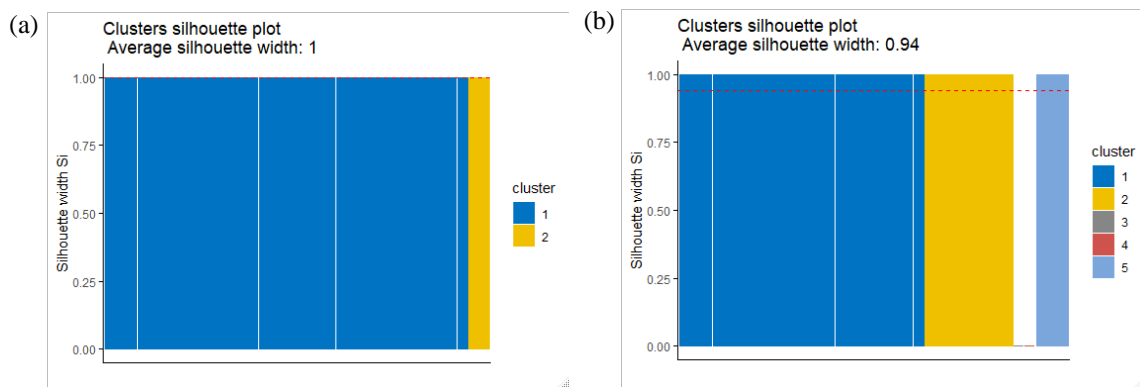
**Gambar 8.** Visualisasi k optimal metode *silhouette*

- Metode *gap statistic*

Penentuan nilai k optimal metode *gap statistic* dapat dilihat dari garis yang tertinggi.

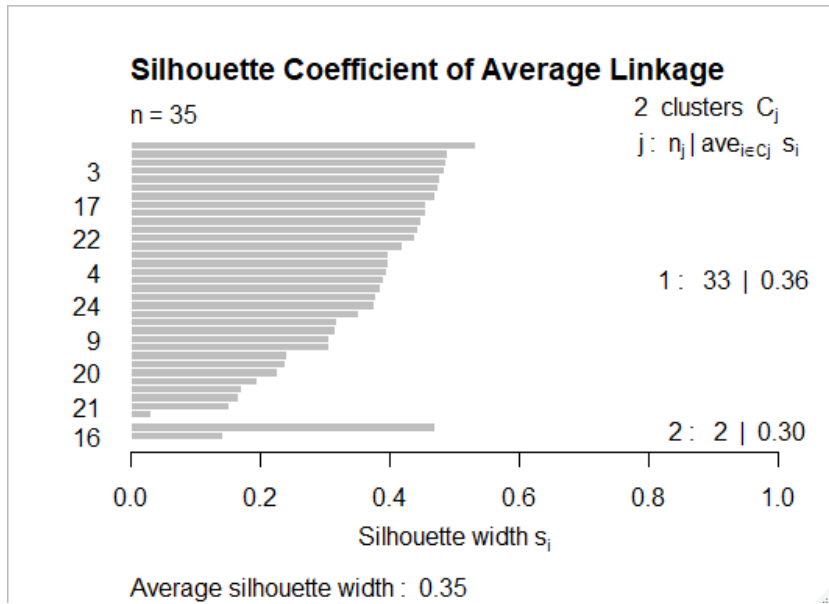
**Gambar 9.** Visualisasi k optimal metode *gap statistic*

Setelah didapatkan nilai k optimal dengan ketiga metode untuk menentukan jumlah *cluster* optimalnya, diperoleh nilai k optimal yang diperoleh dengan metode *elbow* dan *gap statistic* adalah 5 *cluster*, sedangkan dengan *silhouette* adalah 2 *cluster*. Selanjutnya adalah melakukan validasi untuk memastikan bahwa hasil *cluster* mewakili populasi secara umum.

**Gambar 10.** (a) nilai *silhouette* k=2; (b) nilai *silhouette* k=5

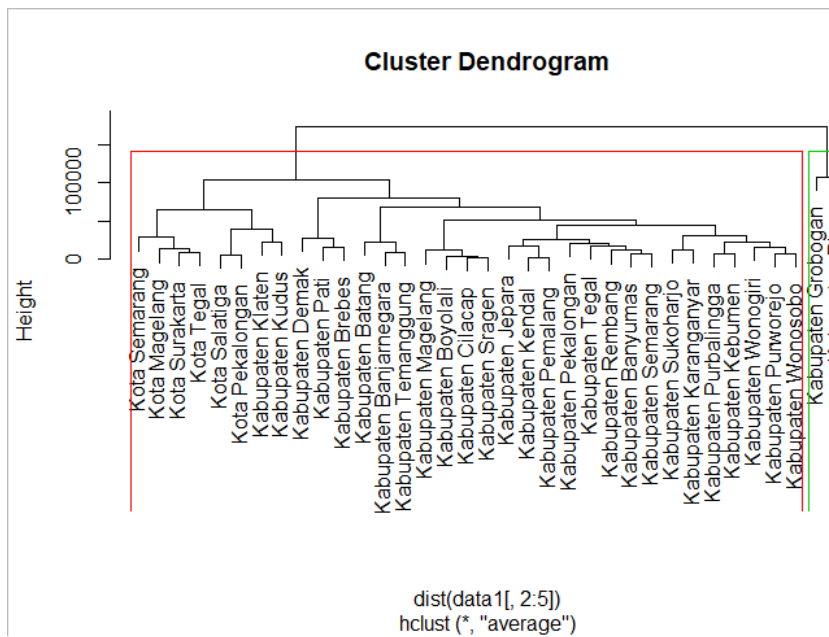
Hasil perbandingan dari nilai *silhouette* dengan k=2 dan k=5, jumlah k optimal yang tepat adalah 2 hal ini dikarenakan rata-rata nilai *silhouettenya* adalah 1 yang artinya semua objek berada pada kluster yang tepat. Langkah selanjutnya adalah menentukan *silhouette coefficient* yang terbentuk dengan k=2.





**Gambar 11.** Visualisasi *silhouette coefficient*

Ditinjau dari *silhouette coefficient*nya dengan metode *average linkage* maka tingkat kebaikan *cluster* nya adalah 0,35. Selanjutnya adalah membentuk *cluster dendrogram* kemiskinan di Jawa Tengah.



**Gambar 12.** Cluster dendrogram kemiskinan di Jawa Tengah

Berikut ini merupakan tabel identifikasi dari *cluster* yang terbentuk dengan menggunakan metode *Average Linkage*.

**Tabel 3.** Hasil pengklasteran kemiskinan di Jawa Tengah dengan metode *Average Linkage*

Cluster	Wilayah	Identifikasi
1	Kabupaten Pemalang, Kabupaten Demak, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Rembang, Kabupaten Jepara, Kabupaten Pekalongan, Kabupaten Tegal, Kabupaten Temanggung, Kabupaten Karanganyar, Kabupaten Purworejo, Kabupaten Wonogiri, Kabupaten Magelang, Kabupaten Boyolali, Kabupaten Sragen, Kabupaten Semarang, Kabupaten	Wilayah di <i>cluster</i> ini memiliki tingkat garis kemiskinan paling tinggi di atas rata-rata dibandingkan dengan klaster 2.

	Kendal, Kabupaten Cilacap, Kabupaten Kebumen, Kabupaten Banjarnegara, Kabupaten Purbalingga, Kabupaten Wonosobo, Kabupaten Banyumas, Kabupaten Pati. Kota Tegal, Kabupaten Kudus, Kota Salatiga, Kota Pekalongan, Kota Magelang, Kota Surakarta, Kota Semarang, Kota Pekalongan, Kabupaten Batang, Kabupaten Brebes.	
2	Kabupaten Grobogan, Kabupaten Blora	Jumlah penduduk miskin, keluarga prasejahtera dan pengeluaran per kapita di klaster ini paling tinggi dibanding dengan <i>cluster</i> 1.

Setelah dilakukan analisis *cluster* kemiskinan di Jawa Tengah dengan menggunakan metode *K-Means* dan *Average Linkage*, berikut ini merupakan perbandingan hasil *cluster* yang terbentuk.

**Tabel 4.** Perbandingan hasil *cluster K-Means* dan *Average Linkage*

Metode <i>K-Means</i>	Metode <i>Average Linkage</i>
Menghasilkan k optimal sebesar 5 <i>cluster</i>	k optimal yang dihasilkan adalah 2 <i>cluster</i>
Nilai <i>silhouette coefficient</i> nya adalah 0,2	Nilai <i>silhouette coefficient</i> nya adalah 0,35

#### 4. Simpulan

Pada penelitian ini dilakukan perbandingan klusterisasi dengan menggunakan metode *K-Means* dan *Average Linkage* untuk mengetahui metode mana yang paling tepat dalam pengelompokan data kemiskinan pada kabupaten dan kota di Jawa Tengah. Sebelum melakukan *clustering* terlebih dahulu dilakukan uji multikolinearitas. Berdasarkan hasil analisis menunjukkan nilai korelasinya kurang dari 0,8 yang artinya tidak terdapat multikolinearitas pada data kemiskinan, kemudian penentuan jumlah *cluster* optimal dengan 3 metode, yaitu metode *silhouette*, metode *elbow* dan metode *gap statistic*. Diperoleh jumlah *cluster* optimal untuk metode *K-Means* adalah 5 *cluster*, sedangkan dengan metode *Average Linkage* adalah 2 *cluster*. Setelah membandingkan kedua metode, didapatkan *Average Linkage* adalah metode terbaik untuk pengklasteran data kemiskinan di Jawa Tengah hal ini dikarenakan nilai *Silhouette Coefficient* lebih tinggi dibandingkan dengan tingkat kebaikan metode *K-Means* yaitu 0,35 sedangkan tingkat kebaikan untuk metode *K-means* adalah 0,2. Karakteristik kabupaten/kota yang terbentuk dengan metode *Average Linkage* yakni *cluster* 1 termasuk wilayah yang garis kemiskinannya tinggi, sedangkan *cluster* 2 tergolong wilayah yang jumlah penduduk miskin, keluarga prasejahtera dan pengeluarannya tinggi.

#### Daftar Pustaka

- Azuri, D. F., Zulhanif, & Pontoh, R. S. (2016). Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Gender Menggunakan Bisecting K-Means. *Peran Penelitian Ilmu Dasar Dalam Menunjang Pembangunan Berkelanjutan*, 78–83.
- Cebeci, Z., & Yildiz, F. (2015). Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics*, 6(3), 13–23.
- Che, X., Zhu, B., & Wang, P. (2021). Assessing global energy poverty: An integrated approach. *Energy Policy*, 149(June 2020), 112099.
- Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). In *Atmospheric Pollution Research* (Vol. 11).
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2006). *Multivariate Data Analysis Pearson International Edition Edition 6*. New Jersey.
- Itang, I. (2017). Faktor Faktor Penyebab Kemiskinan. *Tazkiya*, 16(01), 1–30.
- Johnson, R. A. (2007). *Applied Multivariate Statistics Analysis* (p. 680). p. 680.

- Long, Z. Z., Xu, G., Du, J., Zhu, H., Yan, T., & Yu, Y. F. (2021). Flexible Subspace Clustering: A Joint Feature Selection and K-Means Clustering Framework. *Big Data Research*, 23, 100170.
- Ningrat, D. R., Maruddani, D. A. I., & Wuryandari, T. (2016). Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi. *None*, 5(4), 641–650.
- Ningsih, S., Wahyuningsih, S., & Nasution, Y. N. (2016). Perbandingan kinerja metode complete linkage dan average linkage dalam menentukan hasil analisis cluster ( Studi Kasus : Produksi Palawija Provinsi Kalimantan Timur 2014 / 2015 ). *Prosiding Seminar Sains Dan Teknologi FMIPA Unmu*, 1(1).
- Oktarina, C., Notodiputro, K. A., & Indahwati, I. (2020). Comparison of K-Means Clustering Method and K-Medoids on Twitter Data. *Indonesian Journal of Statistics and Its Applications*, 4(1), 189–202.
- Saleh, S. (2002). *FAKTOR-FAKTOR PENENTU TINGKAT KEMISKINAN REGIONAL DI INDONESIA*. 7(2), 87–102.