



Comparison of Hierarchical and Non-Hierarchical Methods in Clustering Cities in Java Island using the Human Development Index Indicators year 2018

Alvia Rossa Damayanti^a, Arie Wahyu Wijayanto^b

^aPoliteknik Statistika STIS, Jalan Otto Iskandardinata No. 64C, Jakarta Timur 13330, Indonesia. Email: 211709532@stis.ac.id

^bPoliteknik Statistika STIS, Jalan Otto Iskandardinata No. 64C, Jakarta Timur 13330, Indonesia. Email: ariewahyu@stis.ac.id

ABSTRACT

The Human Development Index (HDI) is a composite index to assess the developmental level of life quality in a particular region. In 2018, Java Island, which geographically has the most regencies/ municipalities in Indonesia, achieved human development with “high” status and was followed by all its regencies which have also achieved human development with “high” status. Therefore, research was carried out on how the characteristics inherent in the high HDI have been achieved in regencies on Java Island and grouping them so that it is easy to interpret regencies/ municipalities with homogeneous characteristics. This study used the hierarchical cluster method (single linkage, average linkage, and ward) and non-hierarchical cluster methods (K-Means and FCM). The results show that the best hierarchical cluster method is the average linkage method which forms four clusters where the regencies/ municipalities with the best characteristics (dimensions of education, health, and high purchasing power) are Kepulauan Seribu, Bogor, and 78 other regencies/ municipalities. Then, the best non-hierarchical method is the FCM method which forms two clusters, with a prominent characteristic is those city areas have better characteristics than district areas.

Keywords: Cluster Analysis, Agglomerative Method, K-Means, Fuzzy C-Means, Human Development Index

ABSTRAK

Indeks Pembangunan Manusia (IPM) merupakan suatu indeks komposit untuk mengevaluasi tingkat perkembangan kualitas hidup masyarakat di suatu wilayah. Pada tahun 2018, Pulau Jawa yang secara geografis memiliki kabupaten/kota terbanyak di Indonesia mencapai pembangunan manusia dengan status “tinggi” dan diikuti dengan seluruh kabupaten/kotanya yang juga telah mencapai pembangunan manusia dengan status “tinggi”. Oleh karena itu, dilakukan penelitian tentang bagaimana karakteristik yang melekat pada tingginya IPM yang telah dicapai pada kabupaten/kota di Pulau Jawa serta dilakukan pengelompokan agar mudah dalam menginterpretasikan kabupaten/kota yang memiliki karakteristik homogen. Dalam penelitian ini digunakan metode kluster hirarki (*single*

* Corresponding author.

Alamat e-mail: 211709532@stis.ac.id

linkage, *average linkage*, dan *ward*) dan metode kluster non hirarki (K-Means dan FCM). Hasilnya menunjukkan bahwa metode kluster hirarki terbaik adalah metode *average linkage* yang membentuk empat buah kluster di mana kabupaten/kota dengan karakteristik terbaik (dimensi pendidikan, kesehatan, dan daya beli yang tinggi) adalah Kepulauan Seribu, Bogor, dan 78 kabupaten/kota lainnya. Kemudian, metode non hirarki terbaik adalah metode FCM yang membentuk dua kluster, dengan karakteristik yang menonjol adalah wilayah kota memiliki karakteristik yang lebih baik dibandingkan wilayah kabupaten.

Keywords: Analisis Kluster, *Agglomerative Method*, *K-Means*, *Fuzzy C-Means*, Indeks Pembangunan Manusia

Diserahkan: 10-12-2020; Diterima: 26-12-2021;

Doi: <https://doi.org/10.29303/emj.v4i1.89>

1. Pendahuluan

Suatu negara pada umumnya menginginkan melakukan pembangunan yang baik sehingga masyarakat dapat mencapai kesejahteraan. Pada jaman dahulu kesejahteraan seringkali hanya dipandang dari segi ekonomi. Hingga pada tahun 1991, Bank Dunia menyatakan bahwa tantangan utama pembangunan manusia muncul untuk memperbaiki kualitas kehidupan. Artinya, tingkat pertumbuhan ekonomi yang tinggi yang berhasil dicapai oleh suatu negara belum tentu bisa berhasil dalam memperbaiki taraf hidup sebagian besar penduduknya. Oleh karena itu, muncullah konsep Indeks Pembangunan Manusia yang digunakan untuk mengukur kesejahteraan manusia dari dimensi pendidikan, kesehatan, dan daya beli.

Di Indonesia, IPM sudah dihitung sejak tahun 1996 dan dilakukan secara berkala. Menurut publikasi BPS (2019), IPM Indonesia mencapai 71,39 dan meningkat sebesar 0,58 dari tahun sebelumnya. Sehingga pada tahun 2018 menjadikan Indonesia berstatus pembangunan manusia kategori “tinggi”. Jika ditelusuri lebih jauh ke dalam kabupaten/kota di Indonesia, di antara 23 kabupaten/kota yang berstatus “sangat tinggi”, sebagian besar berada di Pulau Jawa. Kemudian, peningkatan status pembangunan manusia pada 2017 ke 2018 di seluruh kabupaten/kota di Pulau Jawa menunjukkan bahwa tidak ada wilayah dengan status pembangunan manusia “rendah lagi”. Hal ini merupakan prestasi yang baik bagi Pulau Jawa dan bisa dijadikan acuan untuk pulau lainnya di Indonesia. Meski begitu, tinggi rendahnya IPM hanya ditunjukkan melalui indeks komposit, dan tidak ditunjukkan secara detail indikator IPM apa saja yang dominan berkontribusi dalam penyusunan IPM.

Permasalahan utama yang dibahas dalam penelitian ini adalah menguraikan karakteristik kabupaten/kota di Pulau Jawa yang seluruhnya masuk dalam kategori pembangunan berstatus tinggi, berdasarkan indikator penyusunan IPM melalui analisis kluster.

Melalui penelitian ini diharapkan dapat memberi manfaat di masa mendatang. Adapun tujuan dari penelitian ini di antaranya:

- Membandingkan metode pengelompokan hirarki dengan metode non hirarki. Metode hirarki diwakilkan dengan metode *single linkage*, metode *average linkage*, dan metode *ward*. Sedangkan metode non hirarki diwakilkan dengan metode *K-Means* dan *Fuzzy C-Means* dan diterapkan pada seluruh kabupaten/kota dengan menggunakan indikator Indeks Pembangunan Manusia tahun 2018.
- Mengetahui karakteristik pada kluster kabupaten/kota di Pulau Jawa yang terbentuk.

2. Metodologi

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui publikasi BPS Provinsi D.I. Yogyakarta, Jawa Barat, Jawa Tengah, Jawa Timur, dan Jawa Barat dalam angka tahun 2019 mengenai indikator Indeks Pembangunan Manusia. Dataset yang digunakan terdiri dari 129 kabupaten/kota dan lima variabel.

Pada penelitian ini, dilakukan analisis secara deskriptif maupun inferensia. Dalam analisis inferensia digunakan metode analisis kluster metode hirarki dan non hirarki yang kemudian masing-masing dipilih metode terbaik. Analisis deskriptif digunakan dalam melakukan *profiling* hasil analisis kluster yang terbentuk.

2.1 Principal Component Analysis

Menurut Johnson dan Wichern (2007), *Principal Component Analysis* (PCA) digunakan untuk mengatasi adanya multikolinearitas. Prinsip utama dalam PCA adalah adanya korelasi yang tinggi antarvariabel sehingga variabel-variabel tersebut dapat direduksi. PCA dapat dilakukan dengan cara menghilangkan korelasi antarvariabel bebas melalui transformasi variabel bebas asal ke variabel bebas baru yang tidak berkorelasi sama sekali atau sering disebut sebagai *principal component*. Sebelum melakukan PCA, dilakukan pengukuran nilai *Measure of Sampling Adequacy* (MSA). PCA dapat dilanjutkan ketika nilai MSA masing-masing variabel > 0,5 (menunjukkan bahwa variabel bisa diprediksi dan bisa dianalisis lebih lanjut). Kemudian untuk

menentukan banyak komponen utama bisa digunakan *scree plot* (ketika titik kurva tidak lagi menurun tajam atau mulai melandai) atau menggunakan proporsi kumulatif varians terhadap total varians.

2.2 Analisis Kluster

Analisis kluster merupakan salah satu metode objek *mining* yang bersifat tanpa latihan (*unsupervised analysis*). Analisis kluster merupakan salah satu analisis peubah ganda (*multivariate analysis*) yang dapat digunakan untuk mengklusterkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat himigen sesuai dengan faktor yang digunakan untuk melakukan pengklusteran dan heterogen antarkluster yang terbentuk (Gudono, 2011).

Adapun ciri-ciri kluster adalah:

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu kluster (*within-cluster*)
2. Heterogenitas (ketidaksamaan) yang tinggi antarkluster yang satu dengan kluster lainnya (*between-cluster*)

Proses pembentukan kluster dapat dilakukan dengan dua cara, yaitu metode hirarki dan non hirarki. Dalam metode hirarki belum ada informasi jumlah kluster yang akan dipilih. Metode hirarki secara umum dibedakan menjadi metode aglomeratif (penggabungan) dan divisif (pemecahan). Metode aglomeratif terdiri dari metode *linkage* (*single linkage*, *complete linkage*, dan *average*), *variance* (*ward*), dan *centroid*. Sedangkan pada metode non hirarki bertujuan untuk mengelompokkan n objek ke dalam k kluster ($k < n$), di mana nilai k telah ditentukan sebelumnya. Metode non hirarki terdiri dari *sequential threshold*, *parallel threshold*, dan *optimizing partition*. Metode *K-Means* dan *Fuzzy C-Means* sendiri masuk ke dalam metode non hirarki.

2.3 Metode Single Linkage

Metode ini dikenal dengan nama pendekatan tetangga terdekat dan didasarkan pada jarak minimum antarkluster. Dimulai dengan dua objek yang dipisahkan dengan jarak paling pendek maka keduanya akan ditempatkan pada kluster pertama, dan seterusnya.

$$d_{(ij)k} = \min(d_{IK}, d_{JK}) \quad (1)$$

Besaran d_{IK} dan d_{JK} adalah jarak terpendek antara kluster-kluster I dan K dan juga kluster-kluster J dan K .

Tahap melakukan metode *single linkage* adalah sebagai berikut.

1. Menemukan jarak minimum dalam $D = \{d_{ij}\}$.
2. Menghitung jarak antar kluster yang telah dibentuk pada langkah 1 dengan objek lainnya.
3. Dari algoritma di atas jarak-jarak antara objek I dan J dan kluster K .

2.4 Metode Average Linkage

Metode ini didasarkan pada jarak rata-rata antarobservasi. Pengelompokkan dimulai dari pasangan observasi dengan jarak paling mendekati jarak rata-rata.

$$d_{(ij)k} = \frac{\sum_a \sum_b d_{ab}}{N_{IJ} N_K} \quad (2)$$

Keterangan:

d_{ab} : jarak antara obyek i pada kluster IJ dan obyek b pada kluster K

N_{IJ} : jumlah item pada kluster IJ

N_K : jumlah item pada kluster IJ dan K

Tahap melakukan metode *average linkage* adalah sebagai berikut (Johnson dan Wichern, 2007).

1. Mencari dua objek yang memiliki kesamaan paling dekat pada matriks similaritas. Misal kedua objek itu adalah objek I dan J .
2. Menggabungkan objek I dan J dalam satu kluster.
3. Penggabungan kedua merupakan penggabungan kluster dengan kluster lain yang memiliki kesamaan paling dekat, misal kluster K .
4. Mengulangi kembali langkah 3 sebanyak $N-1$ kali, di mana N adalah jumlah objek

2.5 Metode Ward

Metode ini bertujuan untuk memperoleh kluster yang memiliki varian dalam kluster sekecil mungkin. Dalam mengelompokkan objek didasarkan pada kenaikan *sum square error* (SSE). Pada tiap tahap, dua kluster yang memiliki kenaikan SSE paling kecil maka akan digabungkan.

$$SSE = \sum_{j=1}^N (x_j - \bar{x})' (x_j - \bar{x}) \quad (3)$$

Keterangan:

x_j : nilai atau data dari objek ke- j

\bar{x} : rata-rata nilai dari keseluruhan objek dalam kluster

Tahap melakukan metode *ward* adalah sebagai berikut (Johnson dan Wichern, 2007).

1. Dimulai dari setiap objek dianggap sebagai sebuah kluster tersendiri, maka didapat N kluster yang mempunyai satu objek. Pada tahap ini SSE bernilai nol.

2. Menghitung *SSE* untuk setiap kombinasi dua pasang kluster dari N kluster, lalu memilih dua pasang kluster yang memiliki nilai *SSE* terkecil untuk digabungkan menjadi satu kluster. Secara sistematis, N kluster akan berkurang 1 pada setiap tahap ($N-1$).
3. Membuat kombinasi dua pasang kluster baru yang terdiri dari satu kluster yang telah terbentuk dan kluster yang lain, lalu menghitung nilai *SSE* kembali. Memilih dua pasang kluster yang memiliki nilai *SSE* terkecil untuk digabungkan menjadi satu kluster.
4. Mengulangi langkah 3 sampai semua objek bergabung menjadi satu kluster.

2.6 Metode *K-Means*

Metode *K-Means* digunakan sebagai alternatif metode kluster untuk data dengan ukuran yang lebih besar. Hal ini dikarenakan metode ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan metode hirarki (Johnson dan Wicherr, 2007). Langkah-langkah metode *K-Means* adalah sebagai berikut (Johnson dan Wichern, 2007).

1. Menentukan nilai k awal sebagai jumlah kluster yang ingin dibentuk (menggunakan *Elbow Plot*, ditentukan pada titik ketika garis sikunya berbelok (landai)).
2. Menentukan k *centroid* (titik pusat kluster awal secara acak).
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan rumus korelasi antardua objek yaitu jarak *Euclidean*.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (5)$$

Keterangan:

d_{ij} : jarak objek ke- i dan objek ke- j

p : jumlah kluster

x_{ik} : data dari subjek ke- i pada variabel ke- k

x_{jk} : data dari subjek ke- j pada variabel ke- k

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
5. Tentukan posisi *centroid* baru dengan mengitung nilai rata-rata dari data yang ada pada kelompok *centroid* yang sama.

$$C_k = \left(\frac{1}{n_k}\right) \sum d_i \quad (6)$$

Dengan n_k adalah jumlah data dalam kluster k dan d_i adalah data dalam kluster k .

6. Kemudian kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

2.7 Metode *Fuzzy C-Means (FCM)*

Dalam metode FCM, objek data dapat menjadi milik lebih dari satu kluster, dengan nilai keanggotaan tertentu. FCM dapat menentukan keanggotaan dan masing-masing secara iteratif memperbarui nilai keanggotaan dari objek data dengan jumlah kluster yang ditentukan. Langkah-langkah metode FCM adalah sebagai berikut.

1. Menentukan jumlah kluster menggunakan *silhouette plot*.
2. Membangkitkan derajat keanggotaan awal.
3. Menghitung pusat kluster, memperbarui nilai keanggotaan, dan menghitung fungsi objektif iterasi ke- t .
4. Jika $|P_t - P_{t-1}| < \epsilon$, maka iterasi berhenti.

2.8 Validitas Metode Hirarki

Untuk melakukan validitas pada metode hirarki digunakan uji validitas *cophentic*. Menurut Kassambara (2017), untuk mengukur seberapa baik dendrogram yang dihasilkan oleh fungsi *hclust()* mencerminkan data, salah satu caranya adalah menghitung koefisien korelasi *cophentic*. Semakin mendekati satu berarti semakin akurat kluster dapat merefleksikan data. Nilai di atas 0,75 dianggap sudah baik dalam merefleksikan data.

2.9 Validitas Metode Non Hirarki

Menurut Gordon (1999) salah satu validasi dalam analisis kluster adalah menggunakan validasi internal. Validasi internal yaitu proses yang menghitung hubungan antara struktur kluster yang diperoleh dengan data itu sendiri. Data yang digunakan untuk proses kluster dan validasi adalah sama. Pendekatan ini mempunyai beberapa macam indeks kualitas internal yang digunakan untuk menentukan jumlah kluster yang optimal yaitu indeks konektivitas, indeks *daunn*, dan *silhouette*.

2.10 Variabel yang Digunakan

Variabel yang digunakan dalam melakukan pengelompokan kabupaten/kota berdasarkan indikator IPM adalah sebagai berikut.

Tabel 1- Variabel Penelitian

Variabel	Definisi	Skala
Pengeluaran per kapita (X_1)	Biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi	Rasio

Variabel	Definisi	Skala
	dengan banyaknya anggota rumah tangga.	
Rata-rata lama sekolah (X_2)	Rata-rata jumlah tahun yang ditempuh penduduk berumur 15 tahun ke atas untuk menempuh semua jenjang pendidikan yang pernah dijalani.	Rasio
Umur harapan hidup saat lahir (X_3)	Rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir.	Rasio
Harapan lama sekolah (X_4)	Lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang.	Rasio
Jumlah dokter (X_5)	Banyaknya jumlah dokter di suatu kabupaten/kota.	Rasio

3. Analisis dan Pembahasan

3.1 Pengujian Asumsi

Asumsi Non-Multikolinearitas

Dalam melakukan analisis kluster asumsi non-multikolinearitas tidak dapat diabaikan karena jumlah observasi yang tidak terlalu besar. Berdasarkan pengolahan dengan aplikasi R didapatkan korelasi antar variabel sebagai berikut.

Tabel 2 - Korelasi Antar Variabel

	X_1	X_2	X_3	X_4	X_5
X_1	1	0,819	0,311	0,605	0,737
X_2	0,819	1	0,417	0,717	0,517
X_3	0,311	0,417	1	0,316	0,275
X_4	0,605	0,717	0,316	1	0,389
X_5	0,737	0,517	0,275	0,389	1

Pada Tabel 2, terlihat bahwa masih banyak korelasi yang tinggi ($>0,7$) antarvariabel bebas. Yaitu

pada variabel X_1 dan X_2 serta variabel X_2 dan X_4 . Selain itu, nilai p -value pada *Bartlett's Test of Sphercity*, menunjukkan hasil yang signifikan menolak H_0 . Hal ini mengindikasikan adanya multikolinearitas antarvariabel. Maka dari itu asumsi multikolinearitas dalam analisis kluster tidak terpenuhi. Sehingga, dilakukan *Principal Component Analysis* (PCA) terlebih dahulu untuk mengatasi masalah multikolinearitas. Selain itu, karena satuan variabel berbeda, maka dilakukan standarisasi data sebelum melakukan PCA.

Principal Component Analysis (PCA)

Tahapan selanjutnya adalah pembentukan PCA, berdasarkan *output* aplikasi R, proporsi kumulatif varians menggunakan tiga komponen utama mencapai 91,83%, sehingga dalam penelitian ini digunakan tiga buah komponen utama sebagai pembentuk kluster nantinya. Berikut ini persamaan komponen utama yang terbentuk.

$$PC1 = -0,52X_1 - 0,51X_2 - 0,29X_3 - 0,44X_4 - 0,43X_5 \quad (7)$$

$$PC2 = 0,26X_1 + 0,004X_2 - 0,91X_3 - 0,02X_4 + 0,33X_5 \quad (8)$$

$$PC3 = 0,12X_1 - 0,27X_2 + 0,28X_3 - 0,64X_4 + 0,65X_5 \quad (9)$$

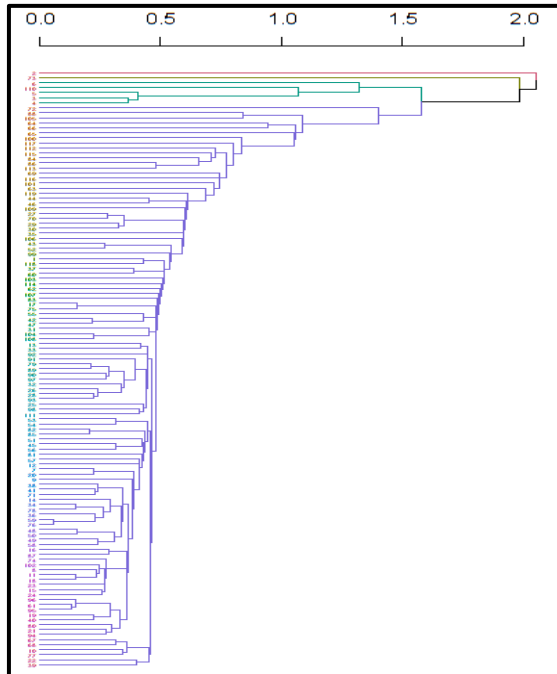
Persamaan (7), (8), dan (9) digunakan untuk mencari nilai komponen utama. Skor tersebut yang dijadikan dasar pengelompokan kabupaten/kota di Pulau Jawa tahun 2018 menggunakan metode hirarki maupun metode non hirarki, karena ketiga komponen tersebut sudah bebas dari multikolinearitas.

3.2 Analisis Kluster Hirarki

Dalam melakukan analisis kluster hirarki digunakan tiga metode, yaitu *single linkage*, *average linkage*, dan *ward* dengan menggunakan jarak *Euclidean*. Berdasarkan pengujian dengan aplikasi R didapatkan kelompok kluster dengan metode *single linkage*, *average linkage*, dan *ward* yang hasil dendogramnya adalah sebagai berikut.

Metode Single Linkage

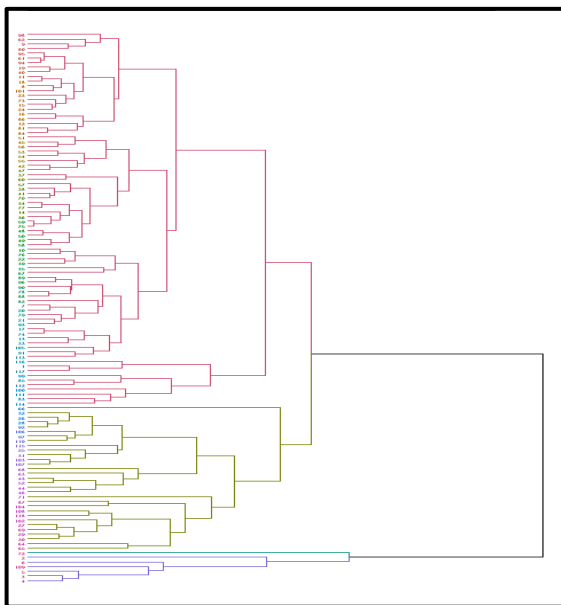
Pada metode *single linkage*, terbentuk empat buah kluster di mana jumlah masing-masing objek pada kluster satu samapai dengan empat berturut-turut adalah 112, 1, 5, dan 1 kabupaten/kota. Berikut adalah dendogram dari pengelompokan menggunakan metode *single linkage*.



Gambar. 1 - Dendrogram hasil pengelompokkan dengan metode *single linkage*

Metode *Average Linkage*

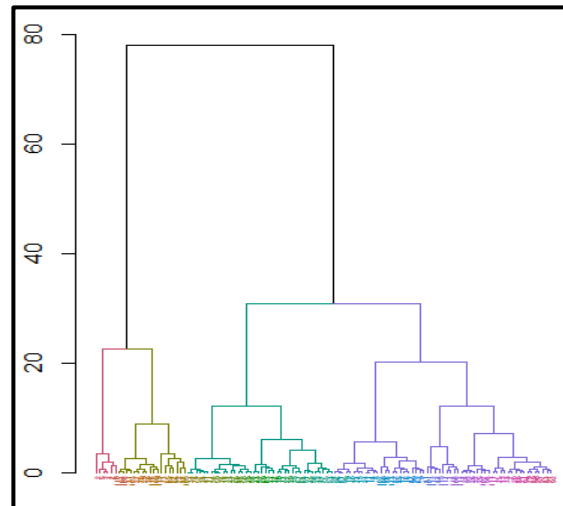
Pada metode *average linkage*, terbentuk empat buah kluster di mana jumlah masing-masing objek pada kluster satu samapai dengan empat berturut-turut adalah 81, 6, 31, dan 1 kabupaten/kota. Berikut adalah dendrogram dari pengelompokkan menggunakan metode *average linkage*.



Gambar. 2 - Dendrogram hasil pengelompokkan dengan metode *average linkage*

Metode *Ward*

Pada metode *ward*, terbentuk empat buah kluster di mana jumlah masing-masing objek pada kluster satu samapai dengan empat berturut-turut adalah 41, 6, 50, dan 22 kabupaten/kota. Berikut adalah dendrogram dari pengelompokkan menggunakan metode *ward*.



Gambar. 3 - Dendrogram hasil pengelompokkan dengan metode *ward*

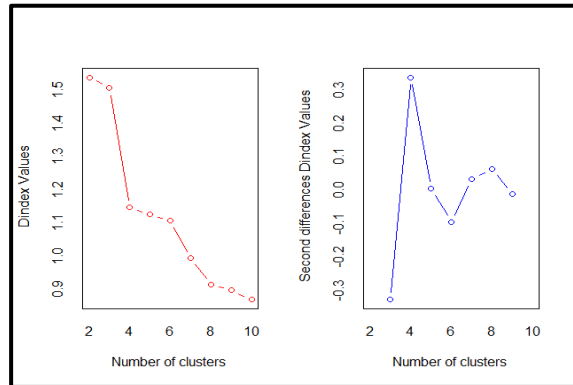
Validitas Metode *Hirarki*

Selanjutnya adalah uji validitas metode kluster hirarki untuk mengukur metode mana yang terbaik dalam membentuk dendrogram dalam mencerminkan data. Pengujian ini dilakukan dengan menghitung koefisien korelasi *cophentic*. Hasilnya adalah sebagai berikut.

Tabel 3 - Nilai koefisien korelasi *cophentic*

Metode	Koefisien Korelasi <i>Cophentic</i>
<i>Single linkage</i>	0,797
<i>Average linkage</i>	0,816
<i>Ward</i>	0,672

Nilai korelasi *cophentic* yang mendekati 1 berarti semakin akurat kluster dapat mencerminkan data. Berdasarkan Tabel 3, menunjukkan bahwa metode *average linkage* terpilih menjadi model analisis kluster hirarki terbaik dalam penelitian ini dalam merefleksikan data. Selanjutnya, untuk menentukan jumlah kluster maksimal terbaik yang akan dibentuk dengan metode *average linkage* digunakan package *NbClust()*, hasilnya diperoleh jumlah kluster terbaik adalah menggunakan empat buah kluster.



Gambar. 4 - Penentuan jumlah kluster

Tabel 4 - Hasil pengelompokkan dengan metode *average linkage*

Kluster	Kabupaten/Kota
Kluster 1	Kep. Seribu, Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Banjar, Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Wonogiri, Sragen, Grobogan, Blora, Rembang, Pati, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Tegal, Gunung Kidul, Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Ngawi, Bojonegoro, Tuban, Lamongan, Bangkalan, Sampang, Pamekasan, Sumenep, Kota Probolinggo, Pandeglang, Lebak, Tangerang, Serang, Kota Cilegon, Kota Serang, dan Kota Tangerang Selatan.
Kluster 2	Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Kota Surabaya.
Kluster 3	Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon,

	Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Klaten, Sukoharjo, Karanganyar, Kudus, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan, Kulon Progo, Bantul, Sleman, Sidoarjo, Magetan, Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu, Kota Tangerang, Kota Semarang\
Kluster 4	Kota Yogyakarta

Dari pembentukan kelompok kabupaten/kota dengan metode *average linkage*, masing-masing jumlah anggota dalam kluster satu hingga kluster empat berturut-turut adalah 81 kabupaten/kota, 6 kabupaten/kota, 31 kabupaten/kota, dan 1 kabupaten/kota. Kota Yogyakarta berdiri sendiri dalam sebuah kluster. Pada umumnya, kabupaten dan kota di Jawa Barat, Jawa Tengah, dan Jawa Timur berkumpul pada satu kluster.

Profiling Hasil Analisis Kluster Hirarki

Berdasarkan pengelompokkan dengan metode *average linkage* didapatkan empat buah kluster. Berikut ini merupakan tabel rata-rata masing-masing indikator dari komponen utama yang telah ditransformasi kembali sesuai dengan data sebelumnya.

Tabel 5 - Nilai rata-rata setiap indikator IPM dalam kluster

Kluster	X_1	X_2	X_3	X_4	X_5
1	10145,6	7,7	71,6	12,5	229,4
2	18878,8	11	73,8	13,4	3009
3	13225,4	9,8	74,1	14,1	452,6
4	18629	11,4	74,5	17,1	2515

Berdasarkan Tabel 5, dapat dianalisa bahwa kluster ke empat yaitu Kota Yogyakarta memiliki karakteristik dengan tiga rata-rata variabel tertinggi jika dibandingkan kluster lainnya. Ketiga variabel tersebut adalah rata-rata lama sekolah (X_2), umur harapan hidup (X_3), dan harapan lama sekolah (X_4). Variabel yang merujuk pada pendidikan ini sejalan dengan kondisi Kota Yogyakarta yang notabane-nya adalah kota pelajar di Indonesia. Sedangkan kluster yang seluruh rata-rata variabelnya rendah terdapat pada kluster pertama yang terdiri dari Kepulauan Seribu, Bogor, Sukabumi, Cianjur, dan 76

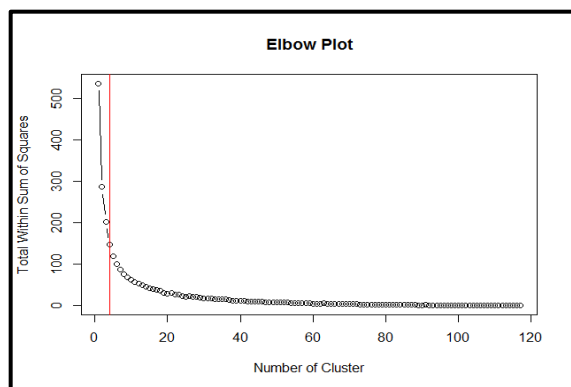
kabupaten/kota lainnya. Kluster pertama tentu memiliki susunan karakteristik variabel terendah jika dibandingkan dengan kluster dua dan kluster tiga. Kluster dua terdiri dari Kota Jakarta dan Kota Surabaya yang memiliki fasilitas pendidikan, kesehatan, dan biaya hidup yang lebih tinggi. Begitu pula pada kluster tiga yang terdiri dari Kota Semarang, Kota Madiun, Kota Megelang, dan 28 kabupaten/kota lainnya. Jika dilihat berdasarkan nilai rata-rata keseluruhan variabel, maka rata-rata tertinggi adalah kluster ke empat dan rata-rata terendah adalah kluster pertama.

3.3 Analisis Kluster Non Hirarki

Analisis kluster metode non hirarki yang digunakan adalah metode *K-Means* dan *Fuzzy C-Means*. Di mana jumlah kluster optimum yang akan digunakan terlebih dahulu ditentukan.

Metode *K-Means*

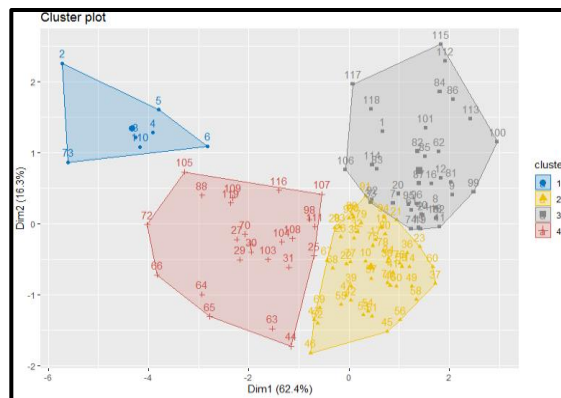
Penentuan jumlah kluster optimum dalam metode ini menggunakan *Elbow Plot*. Jumlah kluster optimum ditentukan berdasarkan titik sebelum kurva melandai dan setelah penurunan tajam. Berdasarkan gambar berikut dihasilkan jumlah kluster optimum adalah empat.



Gambar. 5 - *Elbow Plot*

Pengelompokkan dengan metode *K-Means* menunjukkan terdapat empat kelompok kluster dimana jumlah kabupaten/kota pada kluster satu hingga empat berturut-turut adalah 7, 51, 38, dan 23 kabupaten/kota. Selain itu, berdasarkan *plot K-Means*, ditunjukkan bahwa Kabupaten Pacitan dan Kota Batu berkumpul dengan kluster DKI Jakarta. Pada umumnya, kabupaten dan kota di Jawa Barat, Jawa Tengah, dan Jawa Timur berkumpul pada satu kluster. Kemudian diketahui bahwa varians antarkluster sebesar 397,30 dan masing-masing varians dalam kluster satu hingga empat berturut-turut adalah 16,95; 45,97; 36,13; dan 45,14. Hal ini menunjukkan bahwa kluster yang dibentuk sudah

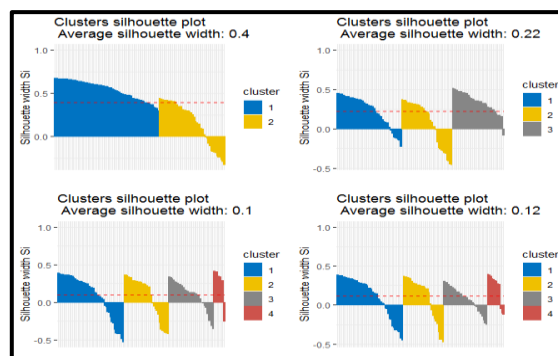
memiliki ciri kluster yang baik, yaitu varians antarkluster tinggi (heterogen) dan varians dalam kluster rendah (homogen). Gambar persebaran kelompok-kelompok kluster yang terbentuk adalah sebagai berikut.



Gambar. 6 - Hasil pengelompokkan dengan metode *K-Means*

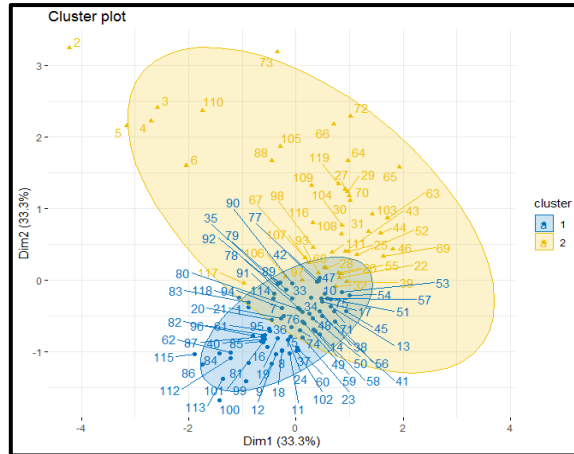
Metode *Fuzzy C-Means (FCM)*

Metode FCM merupakan pengembangan dari metode *K-Means* di mana keberadaan tiap titik data dalam kluster ditentukan oleh derajat keanggotaan. Penentuan jumlah kluster optimum dipilih berdasarkan *silhouette plot* dengan nilai *silhouette* terbesar. Berikut adalah hasil dari *silhouette plot*.



Gambar. 7 - Nilai *silhouette FCM*

Dalam penelitian ini didapatkan jumlah kluster optimum untuk metode FCM adalah dua buah, karena nilai *silhouette-width* dengan dua kluster menunjukkan hasil yang tertinggi. Sedangkan jika menggunakan lebih dari dua kluster nilai *silhouette-width* mengalami penurunan. Nilai *silhouette-width* yang menurun mengindikasikan bahwa data tidak terkluster dengan tepat. Berikut ini merupakan gambar persebaran pengelompokkan dengan metode FCM. Kluster pertama terdiri dari 73 kabupaten/kota dan kluster kedua terdiri dari 46 kabupaten/kota.



Gambar. 8 - Hasil pengelompokkan dengan metode FCM

Hard clustering merupakan suatu pendekatan dalam metode FCM, ketika suatu objek memiliki nilai keanggotaan dalam satu kluster paling besar, maka objek tersebut masuk ke dalam klusternya. Sebagai contoh, Kepulauan Seribu masuk ke dalam kluster pertama dengan nilai keanggotaan 0,57, dan masuk ke dalam kluster dua dengan nilai keanggotaan 0,43. Karena nilai keanggotaan pada kluster satu lebih tinggi, maka Kepulauan Seribu masuk ke dalam kluster pertama.

Validitas Metode Non Hirarki

Selanjutnya adalah uji validitas metode kluster non hirarki melalui uji validitas internal. Untuk metode *K-Means* dengan jumlah kluster sebanyak empat dan metode FCM dengan jumlah kluster sebanyak dua, penghitungan dengan *package cValid()* aplikasi R hasilnya adalah sebagai berikut.

Tabel 6 - Ukuran validitas metode K-means dan FCM

Metode Pengklasteran	Ukuran Validasi	Nilai
(1)	(2)	(3)
<i>K-means</i>	<i>Connectivity</i>	33,07
	<i>Daunn</i>	0,07
	<i>Silhouette</i>	0,34
FCM	<i>Connectivity</i>	25,66
	<i>Daunn</i>	0,04
	<i>Silhouette</i>	0,4

Berdasarkan Tabel 6, dapat disimpulkan bahwa hasil validitas internal menunjukkan metode FCM lebih valid dalam mengelompokkan data

dibandingkan metode *K-Means*. Hal ini ditunjukkan oleh nilai validitas pada indeks *connectivity* yang terkecil dan nilai *silhouette* tertinggi ada pada metode FCM. Oleh karena itu, metode FCM dengan jumlah kluster dua, terpilih sebagai metode kluster non hirarki terbaik.

Tabel 7 - Hasil pengelompokkan dengan metode FCM.

Kluster	Kabupaten/Kota	
Kluster 1	Kep. Seribu, Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bandung Barat, Pangandaran, Kota Banjar, Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Wonosobo, Magelang, Boyolali, Wonogiri, Sragen, Grobogan, Blora, Rembang, Pati, Jepara, Demak, Temanggung, Kendal, Batang, Pekalongan, Tegal, Pemalang, Brebes, Sleman, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Jombang, Nganjuk, Madiun, Magetan, Bojonegoro, Tuban, Lamongan, Sampang, Pamekasan, Sumenep, Kota Kediri, Lebak, Tangerang, Serang, Kota Tangerang, dan Kota Tangerang Selatan.	
	Kluster 2	Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Bekasi, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Purworejo, Klaten, Sukoharjo, Karanganyar, Kudus, Semarang, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Pekalongan, Kota Tegal, Kulon Progo, Bantul, Gunung Kidul, Kota Yogyakarta, Pacitan, Mojokerto, Ngawi, Gresik, Bangkalan, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu, Pandeglang, Kota Cilegon, Kota Serang, dan Kota Semarang.

Profiling Hasil Analisis Kluster Non Hirarki

Berdasarkan pengelompokkan dengan metode FCM didapatkan dua buah kluster. Berikut ini merupakan tabel rata-rata masing-masing indikator dari

komponen utama yang telah ditransformasi kembali sesuai dengan data sebelumnya.

Tabel 8 - Nilai rata-rata setiap indikator IPM dalam klaster

Klaster	X ₁	X ₂	X ₃	X ₄	X ₅
1	9970,4	7,04	71,53	12,45	226,08
2	13822,7	9,78	73,79	13,92	799,37

Berdasarkan Tabel 8, dapat dianalisa bahwa mayoritas anggota pada klaster dua terdiri dari wilayah perkotaan. Karakteristik yang tercermin pada klaster dua ini adalah pengeluaran per kapita, rata-rata lama sekolah, umur harapan hidup, harapan lama sekolah, dan jumlah dokter yang seluruhnya lebih tinggi dibandingkan klaster pertama. Hal ini sesuai dengan ciri wilayah perkotaan yang menggambarkan bahwa fasilitas pendidikan, fasilitas kesehatan, dan biaya hidup yang lebih tinggi jika dibandingkan dengan wilayah kabupaten. Wilayah kota dari masing-masing provinsi di Pulau Jawa tersebar secara merata pada klaster dua, begitu juga wilayah kabupaten masing-masing provinsi di Pulau Jawa juga tersebar merata pada klaster pertama. Jika dihitung rata-rata seluruh variabel maka klaster kedua lebih tinggi dibandingkan klaster pertama.

4. Kesimpulan

Berdasarkan analisis dari 119 kabupaten/kota di Pulau Jawa tahun 2018 yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut.

1. Hasil pengelempokkan terbaik pada metode hirarki adalah metode *average linkage* dengan empat buah klaster. Dari metode pengelompokkan *average linkage* didapatkan bahwa klaster dengan rata-rata seluruh variabel terendah adalah klaster pertama yang terdiri dari Kepulauan Seribu, Bogor, dan 78 kabupaten/kota lainnya. Sedangkan rata-rata klaster tertinggi adalah klaster ke empat yang terdiri dari Kota Yogyakarta.
2. Pada analisis klaster non hirarki, metode FCM lebih valid jika dibandingkan metode *K-means* dalam penelitian ini. Hasilnya menunjukkan bahwa dengan menggunakan metode FCM terbentuk dua klaster di mana klaster pertama mayoritas beranggotakan wilayah kota, dan klaster kedua mayoritas beranggotakan wilayah kabupaten. Hal ini berarti masih terdapat ketimpangan pada wilayah kabupaten dengan kota khususnya dari segi fasilitas dan pendidikan.
3. Secara keseluruhan, pengelompokkan antara klaster hirarki dengan klaster non hirarki memberikan hasil yang berbeda.

DAFTAR PUSTAKA

- Ambarwati, Atika Nurani. (2019). *Latent Class Cluster Analysis untuk Pengelompokkan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Indeks Pembangunan Manusia 2017*. Jurnal Variance: Journal Statistics and Its Applications.
- Badan Pusat Statistik. (2019). *Indeks Pembangunan Manusia 2018*. Jakarta.
- Badan Pusat Statistik. (2019). *Provinsi Banten Dalam Angka*. Banten.
- Badan Pusat Statistik. (2019). *Provinsi Daerah Istimewa Yogyakarta Dalam Angka*. Yogyakarta.
- Badan Pusat Statistik. (2019). *Provinsi DKI Jakarta Dalam Angka*. Jakarta.
- Badan Pusat Statistik. (2019). *Provinsi Jawa Barat Dalam Angka*. Jawa Barat.
- Badan Pusat Statistik. (2019). *Provinsi Jawa Tengah Dalam Angka*. Jawa Tengah.
- Badan Pusat Statistik. (2019). *Provinsi Jawa Timur Dalam Angka*. Jawa Timur.
- Badan Pusat Statistik. (2019). *Tabel Dinamis: Indikator Indeks Pembangunan Manusia*. www.bps.go.id (diakses pada 1 Desember 2020).
- Gordon, A.D. (1999). *Classification (2nd Edition)*. London: Chapman & Hall.
- Gudono. (2011). *Analisis Data Multivariat Edisi Pertama*. Yogyakarta: BPFE.
- Johnson, R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis, Sixth Edition*. Prentice-Hall International: New Jersey.
- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R*. (sthd.com)
- Kusumanintyas, Clara Aulia. (2017). *Analisis Klaster Untuk Pengelompokkan Kabupaten/Kota di Provinsi Papua Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2017*. Jakarta: Skripsi Politeknik Statistika STIS.
- Rachmatin, Dewi dan Kania Sawitri. (2012). *Perbandingan Antara Metode Agglomeratif, Metode Divisif, dan Metode K-Means dalam Analisis Klaster*. Jurnal Pendidikan Matematika FMIPA UPI.
- Wahyuni, Sri. (2018). *Pengelompokkan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering*. Jakarta: Jurnal Aplikasi Statistika dan Komputasi Statistik, V.10.1.2018.