

Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan

Ardiana Fatma Dewi¹, Kurnia Ahadiyah²

¹IAIN Kediri, ardianafatmad@iainkediri.ac.id

²IAIN Kediri, kurniaahadiyah@iainkediri.ac.id

DOI 10.31102/zeta.2022.7.2.57-63

ABSTRACT

One way to improve the condition of education in each district/city in East Java is to use grouping analysis according to the characteristics of education. Cluster analysis is an analysis that belongs to the group of multivariate analysis which aims to group objects based on their characteristics. This study uses agglomerative cluster analysis with five methods. This cluster analysis is called "agglomerative hierarchy clustering". The purpose of this study is to compare the method of analysis of the hierarchical cluster by finding the best model through the search for the largest cophenetic correlation coefficient value. The data that used in this study is secondary data obtained from the Central Statistics Agency (BPS) of East Java and the Ministry of Education of the Republic of Indonesia. In the grouping of districts/cities in East Java Province based on education indicators using the Agglomerative Hierarchical Clustering method, it was found that the optimal algorithm used was the average linkage algorithm with a cophenetic correlation value obtained of 0.807, when the value is close to one then the solution resulting from the clustering process is good. Based on education indicators from 38 regencies/cities in East Java Province, they are divided into two clusters where the first cluster consists of 3 regencies/cities and the second cluster is 35 regencies/cities.

Keywords: *agglomerative, clustering, hierarchical, cophenetic*

ABSTRAK

Salah satu cara agar dapat meningkatkan kondisi pendidikan di setiap kabupaten/kota di Jawa Timur adalah dengan menggunakan analisis pengelompokan sesuai dengan karakteristik pendidikannya. Analisis Klaster merupakan analisis yang termasuk ke dalam golongan analisis multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimilikinya. Penelitian ini menggunakan analisis klaster tipe agglomerative dengan lima metode. Analisis klaster ini disebut dengan "agglomerative hierarchy clustering". Tujuan penelitian ini adalah membandingkan metode analisis klaster herarki tersebut dengan mencari model terbaik melalui pencarian nilai koefisien korelasi cophenetic terbesar. Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Jawa Timur dan Kementerian Pendidikan Republik Indonesia. Pada pengelompokan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan dengan menggunakan metode Agglomerative Hierarchical Clustering dihasilkan bahwa algoritma optimal yang digunakan yaitu pada algoritma average linkage dengan nilai korelasi cophenetic yang diperoleh sebesar 0,807, dimana ketika nilai tersebut mendekati satu maka solusi yang dihasilkan dari proses clustering tersebut cukup baik. Berdasarkan indikator pendidikan dari 38 Kabupaten/Kota di Provinsi Jawa Timur terbagi menjadi dua cluster dimana pada cluster pertama beranggotakan 3 Kabupaten/Kota dan cluster kedua 35 Kabupaten Kota.

Keywords: *agglomerative, clustering, hierarchical, cophenetic*

1. PENDAHULUAN

Di Indonesia, pendidikan menjadi salah satu unsur yang mendukung pembangunan dan kemajuan negara. Pendidikan juga dapat meningkatkan kualitas sumber daya manusia jika sistem pendidikan itu sendiri berjalan dengan baik. Namun, dengan kondisi di Indonesia saat ini yang sedang dihadapkan dengan virus Covid-19, pendidikan menjadi salah satu aspek yang terdampak karena pemerintah membuat kebijakan Pembelajaran Jarak Jauh (PJJ) untuk semua jenjang pendidikan. Jawa Timur merupakan provinsi yang memiliki jumlah penduduk terbesar kedua setelah Jawa Barat. Semakin banyak jumlah penduduk, maka semakin banyak pula faktor yang mempengaruhi kondisi pendidikan.

BPS (2021) menyebutkan terjadinya penurunan Angka Partisipasi Kasar (APK) di jenjang SD dan SMP pada saat pandemi Covid-19. Untuk jenjang SD pada tahun 2019 presentase APK sebesar 105,56% sedangkan pada tahun 2020 menurun sebesar 1,21% menjadi 104,35%. Untuk jenjang SMP pada tahun 2019 presentase APK sebesar 94% dan pada tahun 2020 menurun 2% sehingga menjadi 94%. Namun untuk jenjang SMA tidak mengalami peningkatan maupun penurunan antara tahun 2019 dan 2020 yaitu tetap sebesar 85%. Angka Partisipasi Kasar merupakan proporsi anak sekolah di suatu jenjang pendidikan berdasarkan umur yang sesuai dengan jenjang pendidikannya. Semakin tinggi nilai APK di suatu daerah, maka semakin banyak anak yang bersekolah. Dengan adanya penurunan partisipasi anak dalam bersekolah, maka perlu adanya pemetaan kabupaten/kota di Jawa Timur agar kondisi pendidikan di tiap kabupaten mengalami peningkatan.

Salah satu cara agar dapat meningkatkan kondisi pendidikan di setiap kabupaten/kota di Jawa Timur adalah dengan menggunakan analisis pengelompokan sesuai dengan karakteristik pendidikannya. Analisis Klaster merupakan analisis yang termasuk ke dalam golongan analisis multivariat yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimilikinya (Nafisah, 2017). Penelitian ini menggunakan analisis klaster tipe agglomerative dengan lima metode. Analisis klaster ini disebut dengan “agglomerative hierarchy clustering”. Agglomerative hierarchy clustering dibagi menjadi 4 jenis metode berdasarkan sudut pandang cara mengelompokkan objeknya. Diantaranya adalah *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward*. Kelima metode tersebut memiliki tujuan yang yaitu mengelompokkan objek yang memiliki karakter yang sama (Roux, 2018). Tujuan penelitian ini adalah membandingkan kelima metode analisis klaster herarki tersebut dengan mencari model terbaik melalui pencarian nilai koefisien korelasi *cophenetic* terbesar.

2. TINJAUAN PUSTAKA

2.1. Multikolinieritas

Multikolinieritas adalah kondisi dimana beberapa objek penelitian yang memiliki korelasi antar variabel prediktornya. Multikolinieritas dapat menyebabkan data mempunyai determinan sama dengan nol pada matriks $(X^{-1}X)^{-1}$. Selain itu, ada beberapa masalah yang dapat ditimbulkan akibat multikolinieritas, diantaranya adalah sebagai berikut:

1. Nilai dugaan dari koefisien regresi menjadi sangat sensitif terhadap perubahan data
2. Koefisien regresi dugaan menjadi tidak nyata meskipun nilai R^2 tinggi
3. Pada metode kuadrat terkecil, simpangan baku koefisien regresi dugaan menjadi sangat besar.

Salah satu cara untuk mengetahui terdapat multikolinieritas atau tidak pada variabel, dapat dilakukan dengan menghitung nilai *Variance Inflation Factors* (VIF) dengan rumus sebagai berikut:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Dengan R_i merupakan koefisien determinasi variabel bebas ke- i . Tolak ukur terjadinya multikolinieritas pada variabel prediktor jika nilai VIF nya lebih besar dari 10 (Mu'afa, 2019).

2.2. Analisis Komponen Utama

Analisis komponen utama adalah analisis statistika yang termasuk kedalam golongan analisis multivariat. Analisis komponen utama merupakan analisis yang digunakan untuk mentransformasi variabel yang saling berkorelasi menjadi variabel baru yang tidak saling berkorelasi satu sama lain dengan cara mereduksi beberapa variabel sehingga mempunyai dimensi yang lebih kecil tanpa mengurangi informasi data dan tetap dapat menerangkan sebagian besar dari keragaman variabel aslinya. Variabel baru tersebut dinamakan komponen utama (*principal component*).

Pembentukan komponen utama melalui dua matriks yaitu matriks ragam-peragam (Σ) dan matriks korelasi yang digunakan untuk perhitungan nilai eigen λ_i dan vektor eigen γ_i . Nilai eigen yang diperoleh dari persamaan $|\Sigma - \lambda I| = 0$ adalah $\lambda_1, \lambda_2, \dots, \lambda_p$, sedangkan vektor eigen yang diperoleh dari persamaan $(\Sigma - \lambda_i I)\gamma_i = 0$ adalah $\gamma_1, \gamma_2, \dots, \gamma_p$ dimana $i = 1, 2, \dots, p$ (Mariana, 2013). Komponen utama ke- i dapat didefinisikan sebagai berikut:

$$KU_i = \gamma'_{i1}X_1 + \dots + \gamma'_{ip}X_p$$

dengan ragam sebagai berikut:

$$\begin{aligned}\sigma_{KU_1}^2 &= \text{Var}(KU_i) \\ &= \lambda_i \\ &= \gamma_i' \Sigma \gamma_i \\ &= \sum_{i=1}^p \sum_{j=1}^p \gamma_{1i} \gamma_{1j} \sigma_{ij}\end{aligned}$$

Dimana $j = 1, 2, \dots, p$. Adapun beberapa kriteria pemilihan k komponen utama diantaranya adalah sebagai berikut:

1. Dalam pemilihan jumlah komponen tersebut belum ada aturan tertentu yang disepakati oleh semua ahli statistika. Sebagian ahli statistika ada yang mengambil akar ciri yang lebih besar dari 1 atau mengambil komponen utama tertentu, dimana proporsi keragaman y yang dapat diterangkan oleh komponen tersebut dianggap cukup berarti.
2. Proporsi kumulatif keragaman data asal yang dijelaskan oleh k komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar.
3. Dengan menggunakan scree plot yaitu plot antara i dengan λ_i , pemilihan nilai k berdasarkan scree plot ditentukan dengan melihat letak terjadinya belokan dengan menghapus komponen utama yang menghasilkan beberapa nilai eigen kecil membentuk pola garis lurus.

2.3. Agglomerative Hierarchical Clustering

Analisis *cluster* adalah metode statistika yang digunakan untuk pengelompokan data yang mempunyai banyak variabel dengan tujuan utamanya mengelompokkan objek yang mirip. Salah satu cara mengelompokkan objek yang mirip yaitu dengan cara mengukur jarak antar objek dengan rumus *Euclidean*:

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_{ik} - y_{jk})^2}$$

Analisis *cluster* dibagi menjadi dua metode, yaitu hirarki dan non hirarki. Berdasarkan cara mengelompokkannya, metode hirarki dibagi menjadi dua yaitu Agglomerative Hierarchical Clustering dan Divisive Clustering. Agglomerative Hierarchical Clustering dibagi menjadi 4 metode yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward*. Adapun langkah-langkah pengelompokan dengan metode Agglomerative Hierarchical Clustering dengan objek sebanyak N cluster adalah sebagai berikut (Murtagh, 2012):

1. Dimulai dengan N cluster, setiap cluster mengandung entiti tunggal dan sebuah matriks

simetrik dari jarak (similarities) $D = \{d_{ik}\}$ dengan tipe $N \times N$.

2. Cari matriks jarak untuk pasangan cluster yang terdekat (paling mirip). Misalkan jarak antara cluster U dan V yang paling mirip adalah d_{UV} .
3. Gabungkan cluster U dan V . Label cluster yang baru dibentuk dengan (UV) . Update entries pada matrik jarak dengan cara :
 - a. Hapus baris dan kolom yang bersesuaian dengan cluster U dan V
 - b. Tambahkan baris dan kolom yang memberikan jarak-jarak antara cluster (UV) dan cluster-cluster yang tersisa.
4. Ulangi langkah 2 dan 3 sebanyak $(N - 1)$ kali. (Semua objek akan berada dalam cluster tunggal setelah algoritma berakhir). Catat identitas dari cluster yang digabungkan dan tingkat-tingkat (jarak atau similaritas) dimana penggabungan terjadi.

Single Linkage

Metode *Single Linkage* merupakan metode analisis cluster herarki dengan cara mengelompokkan objek berdasarkan jarak yang paling dekat terlebih dahulu (Goreti, 2016). Pertama, kita terlebih dahulu menentukan jarak terpendek $D = \{d_{ik}\}$ serta menggabungkan objek-objek yang bersesuaian. Berdasarkan langkah ke-3 pada Agglomerative Hierarchical Clustering, jarak antara cluster UV dan W dapat dihitung dengan

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}$$

dimana d_{UW} adalah jarak terpendek antara U dan W , sedangkan d_{VW} adalah jarak terpendek antara V dan W .

Complete Linkage

Metode *Complete Linkage* merupakan metode analisis cluster herarki dengan cara mengelompokkan objek berdasarkan jarak paling jauh atau yang mempunyai kemiripan terkecil antar objek (Ningsih, 2016). Pertama, kita terlebih dahulu menentukan elemen matriks $D = \{d_{ik}\}$ serta menggabungkan objek-objek yang bersesuaian misalnya U dan V untuk menjadi cluster (UV) . Berdasarkan langkah ke-3 pada Agglomerative Hierarchical Clustering, jarak antara cluster UV dan W dapat dihitung dengan

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\}$$

dimana d_{UW} adalah jarak terdekat antara U dan W , sedangkan d_{VW} adalah jarak terdekat antara V dan W .

Average Linkage

Metode *average linkage* merupakan metode analisis cluster herarki dengan cara mengelompokkan objek berdasarkan jarak rata-rata antara semua pasangan objek (Asiska, 2019). Pertama, tentukan elemen matriks $D = \{d_{ik}\}$ untuk mendapatkan objek yang paling mirip (paling dekat) satu sama lain misalnya U dan V . Berdasarkan langkah ke-3 pada

Agglomerative Hierarchical Clustering, jarak antara cluster UV dan W dapat dihitung dengan

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{UV}N_W}$$

dimana d_{ik} adalah jarak antar objek i pada cluster (UV) dan objek k pada cluster W , sedangkan N_{UV} dan N_W adalah banyaknya item pada cluster (UV) dan W .

Ward Method

Metode Ward merupakan metode analisis cluster herarki yang didasari oleh hilangnya informasi karena penggabungan objek-objek menjadi cluster. Metode ini menggunakan perhitungan dengan memaksimalkan homogenitas menggunakan *Error Sum of Squares* (ESS) di dalam satu kelompok (Fathia, 2019).

$$ESS = \sum_{j=1}^k \left(\sum_{i=1}^{n_j} x_{ij}^2 - \frac{1}{n_j} \left(\sum_{i=1}^{n_j} x_{ij} \right)^2 \right)$$

dimana:

X_{ij} = Nilai objek ke- i

$i = 1, 2, 3, \dots$ pada kelompok ke- j

K = Jumlah kelompok setiap stage

n_j = Jumlah kelompok ke- i pada kelompok ke- j

2.4. Model Terbaik

Dalam perbandingan metode dalam statistika, perlu adanya pencarian model terbaik. Setelah mendapatkan hasil dari masing-masing proses analisis cluster hirarki, maka perlu dilakukan uji validitas cluster guna melihat kebaikan model hasil analisis cluster. Salah satu ukuran yang dapat digunakan untuk menguji validitas hasil cluster adalah koefisien korelasi (Widodo, 2018).

Koefisien korelasi cophenetic adalah koefisien korelasi antar elemen-elemen asli matriks yang tidak mirip (matriks jarak Euclidian) dengan elemen-elemen yang dihasilkan oleh matriks cophenetic. Adapun rumus untuk menghitung koefisien korelasi cophenetic sebagai berikut (Saracli, 2013):

$$r_{Coph} = \frac{\sum_{i < k} (d_{ik} - \bar{d})(d_{c_{ik}} - \bar{d}_c)}{\sqrt{\left[\sum_{i < k} (d_{ik} - \bar{d})^2 \right] \left[\sum_{i < k} (d_{c_{ik}} - \bar{d}_c)^2 \right]}}$$

dimana:

r_{Coph} = koefisien korelasi cophenetic

d_{ik} = jarak Euclidean antar objek i dan k

\bar{d} = rata-rata d_{ik}

$D_{c_{ik}}$ = jarak cophenetic antara objek i dan k

\bar{d}_c = rata-rata $d_{c_{ik}}$

3. METODE PENELITIAN

3.1. Data Penelitian

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Jawa Timur dan Kementerian

Pendidikan Republik Indonesia. Terdapat 38 unit pengamatan yang digunakan yaitu sebanyak Kabupaten/ Kota yang ada di Jawa Timur. Data yang akan dianalisis berkaitan dengan indikator pendidikan di Jawa Timur pada Tahun 2021, berikut ini pada Tabel 1 merupakan beberapa variabel yang digunakan.

Tabel 1. Variabel Penelitian

	Variabel
X ₁	Angka Partisipasi Kasar (APK) SD
X ₂	Angka Partisipasi Kasar (APK) SMP
X ₃	Angka Partisipasi Kasar (APK) SMA
X ₄	Jumlah Siswa SD
X ₅	Jumlah Siswa SMP
X ₆	Jumlah Siswa SMA
X ₇	Jumlah Sekolah SD
X ₈	Jumlah Sekolah SMP
X ₉	Jumlah Sekolah SMA
X ₁₀	Rata-rata Lama Sekolah (RLS)

3.2. Metode Analisis Data Penelitian

Terdapat beberapa tahapan pada analisis Agglomerative Hierarchical Clustering diantaranya adalah sebagai berikut ini.

1. Analisis statistika deskriptif untuk mengetahui karakteristik umum pendidikan yang ada di Kabupaten/Kota yang ada di Provinsi Jawa Timur.
2. Uji korelasi untuk mengetahui keeratan hubungan antar variabel, karena pada analisis ini menggunakan variabel yang cukup banyak sehingga dimungkinkan antar variabel terdapat korelasi yang kuat. Ketika terdapat korelasi yang kuat diantara variabel tersebut maka dilakukan Analisis Komponen Utama untuk mereduksi variabel tanpa mengurangi informasi yang ada. Jika tidak terdapat korelasi maka langsung dilanjutkan ke langkah 3.
3. Menghitung jarak *Euclidean* untuk mengetahui ukuran kedekatan antar objek.
4. Analisis cluster untuk mengelompokan Kabupaten/ Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan menggunakan metode *Agglomerative Hierarchical Clustering*.
5. Membandingkan beberapa algoritma pada *Agglomerative Hierarchical Clustering* diantaranya yaitu dengan metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward*.
6. Memilih salah satu diantara metode *Single Linkage*, *Complete Linkage*, *Average Linkage*, dan *Ward* yang memiliki nilai koefisien korelasi cophenetic terbesar.
7. Menentukan banyak cluster dengan metode *internal* dan *stability*.
8. Membuat Kesimpulan.

4. HASIL PENELITIAN

4.1. Analisis Deskriptif

Berdasarkan analisis deskriptif dapat diperoleh informasi bahwa rata-rata Angka Partisipasi Kasar (APK) tingkat SD di Provinsi Jawa Timur tahun 2021 sebesar 100,04, dimana APK tertinggi berada pada Kota Kediri dan APK terendah berada pada Kota Probolinggo. Rata-rata Angka Partisipasi Kasar (APK) tingkat SMP di Provinsi Jawa Timur sebesar 105,33 dimana APK SMP tertinggi berada pada Kabupaten Ponorogo dan terendah berada pada Kabupaten Sampang. Sedangkan rata-rata APK tingkat SMA di Provinsi Jawa Timur sebesar 101,95 dimana APK SMA tertinggi berada pada Kota Kediri dan yang terendah di Kabupaten Sampang.

Selain APK SD, SMP dan SMA digunakan juga variabel Jumlah Siswa SD dengan rata-rata siswa SD di Provinsi Jawa Timur sebesar 71.774 siswa dengan jumlah siswa terbanyak pada Kota Surabaya dan paling sedikit di Kota Mojokerto. Rata-rata siswa SMP di Provinsi Jawa Timur 32.513 siswa dengan siswa terbanyak berada pada Kota Surabaya dan paling sedikit di Kota Batu. Rata-rata siswa SMA di Provinsi Jawa Timur sebesar 33.475 siswa dimana terbanyak ada pada Kota Surabaya dan paling sedikit pada Kabupaten Situbondo.

Pada penelitian ini juga diamati variabel banyak sekolah, dimana rata-rata banyak SD di provinsi Jawa Timur sebanyak 502 sekolah, dimana SD terbanyak ada pada Kabupaten Malang dan paling sedikit ada pada Kota Mojokerto. Rata-rata banyak SMP di Jawa Timur sebanyak 125 sekolah, dimana terbanyak pada Kabupaten Malang dan paling sedikit di Kota Mojokerto. Sedangkan rata-rata banyak SMA di Jawa Timur sebanyak 95 sekolah paling banyak ada di Kota Surabaya dan paling sedikit pada Kota Mojokerto.

Serta Rata-rata Lama Sekolah (RLS) di Jawa Timur sebesar 8 tahun dimana RLS tertinggi pada Kota Madiun dan terendah pada Kabupaten Sampang. Beberapa variabel yang telah disebutkan diatas merupakan beberapa indikator yang akan dijadikan dasar untuk pengelompokan Kabupaten/ Kota di Provinsi Jawa Timur pada Tahun 2021. Pada penelitian ini digunakan Metode *Agglomerative Hierarchical Clustering*.

4.2. Hasil Penerapan Metode *Agglomerative Hierarchical Clustering* pada Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur berdasarkan Indikator Pendidikan

4.2.1 Uji Korelasi dan Analisis Komponen Utama

Langkah awal sebelum dilakukan analisis *cluster* perlu dilakukan analisis uji multikolinearitas karena variabel yang digunakan dalam penelitian ini cukup banyak. Uji untuk mengetahui apakah terdapat

multikolinearitas antar variabel yaitu dengan uji korelasi semua variabel penelitian.

Tabel 2. Uji Korelasi

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
X ₁	1									
X ₂	-0,157	1								
X ₃	-0,025	0,167	1							
X ₄	-0,029	-0,013	-0,279	1						
X ₅	-0,011	-0,037	-0,190	0,979	1					
X ₆	0,026	0,019	0,006	0,858	0,867	1				
X ₇	0,000	0,026	-0,416	0,755	0,703	0,629	1			
X ₈	0,146	-0,129	-0,447	0,829	0,789	0,726	0,794	1		
X ₉	0,069	-0,133	-0,282	0,833	0,812	0,799	0,731	0,946	1	
X ₁₀	0,011	0,079	0,531	-0,096	0,019	0,044	-0,596	-0,436	-0,284	1

Berdasarkan Tabel 2 dapat diketahui bahwa terdapat beberapa variabel memiliki nilai korelasi yang dikategorikan berhubungan kuat. Olehkarena itu dilakukan analisis komponen utama untuk menghasilkan skor komponen utama yang saling bebas dari korelasi.

Sebelum dilakukan Analisis Komponen Utama perlu dilakukan uji untuk mengetahui kecukupan data yaitu nilai *Kaiser Meyer Olkin (KMO)* bertujuan untuk mengetahui apakah semua data yang diambil cukup untuk difaktorkan dan uji pemeriksaan asumsi *barlett sphericity* dilakukan untuk mengetahui apakah terdapat hubungan antar variabel.

Tabel 3. Uji *Kaiser Meyer Olkin (KMO)*

	Nilai
KMO	0,686

Berdasarkan Tabel 3 dapat diketahui bahwa nilai KMO sebesar 0,686 apabila $KMO > 0,5$ maka dapat dikatakan bahwa data yang dianalisis sudah memenuhi asumsi kecukupan data.

Tabel 4. Uji *Bartlett Sphericity*

	Chi-Square	p-value
<i>Bartlett Sphericity</i>	427,13	$2,2 \times 10^{-16}$

Berdasarkan Tabel 4 dapat diketahui bahwa nilai *Chi-Square* sebesar 427,13 dengan taraf signifikansi sebesar 5% dihasilkan *p-value* sebesar $2,2 \times 10^{-16}$, jika *p-value* kurang dari 5% maka dapat disimpulkan bahwa terdapat korelasi antar variabel pada data indikator pendidikan di Provinsi Jawa Timur pada tahun 2021.

Hasil yang didapatkan dalam analisis komponen utama berupa skor komponen utama yang dihitung berdasarkan persamaan dari komponen utama yang terbentuk, dan selanjutnya akan digunakan sebagai inputan untuk menggantikan variabel asal. Berikut dibawah ini merupakan hasil korelasi setelah dilakukan analisis komponen utama dapat dilihat pada Tabel 5.

Tabel 5. Uji Korelasi setelah dilakukan Analisis Komponen Utama

	X ₁	X ₂	X ₃	...	X ₈	X ₉	X ₁₀
X ₁	1						
X ₂	$8,49 \times 10^{-17}$	1					
X ₃	$2,19 \times 10^{-16}$	$3,005 \times 10^{-16}$	1				
.							
.							
X ₈	$4,049 \times 10^{-16}$	$-9,461 \times 10^{-16}$	$1,415 \times 10^{-15}$		1		
X ₉	$1,806 \times 10^{-15}$	$-1,574 \times 10^{-15}$	$6,305 \times 10^{-15}$		$6,054 \times 10^{-16}$	1	
X ₁₀	$8,49 \times 10^{-16}$	$-4,100 \times 10^{-15}$	$3,663 \times 10^{-15}$		$1,176 \times 10^{-14}$	$9,629 \times 10^{-15}$	1

Berdasarkan Tabel 5 dapat dilihat bahwa dengan adanya analisis komponen utama data akan ditransformasikan sehingga antar variabel yang mulanya mengandung korelasi tinggi akan menjadi lebih kecil. Hal tersebut sesuai dengan asumsi pada analisis *cluster*, bahwa antar variabel telah saling bebas.

Setelah pemeriksaan asumsi pada analisis komponen utama selanjutnya dilakukan analisis faktor yang digunakan untuk mengetahui faktor baru yang terbentuk. Berikut ini merupakan acuan yang akan digunakan untuk membentuk analisis faktor pada Tabel 6.

Tabel 6. Nilai Eigen Value

Nilai	Komponen					
	1	2	3	4	...	10
Eigen Value	2,287	1,311	1,083	0,925	...	0,099
% of Variance	0,523	0,172	0,117	0,085	...	0,000
Cumulative %	0,523	0,695	0,812	0,898	...	1,000

Berdasarkan Tabel 6 dapat dilihat bahwa ketika *eigen value* lebih besar dari 1, maka komponen tersebut merupakan banyak faktor yang nantinya akan terbentuk yaitu komponen ke 3. Artinya pada penelitian ini dari sepuluh variabel yang digunakan maka akan disederhanakan sehingga menjadi 3 faktor, dengan total varians sebesar 81,2%. Faktor baru yang terbentuk bebas dari korelasi dan selanjutnya akan dilakukan analisis *cluster*.

4.2.2 Analisis Cluster

Analisis *cluster* pada penelitian ini digunakan untuk mengelompokkan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan yang sudah terbentuk dalam analisis faktor. Pengelompokan didasarkan pada jarak *Euclidean* dengan pembentukan *cluster* menggunakan algoritma yang menghasilkan korelasi *chopenetic* yang paling tinggi diantara algoritma berikut ini *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Ward Linkage*, *Centroid Linkage*, *Median Linkage*. Hasil dari beberapa algoritma *cluster* yang telah disebutkan dapat dilihat pada tabel dibawah.

Tabel 7. Nilai Korelasi *Chopenetic*

Metode	Korelasi <i>Chopenetic</i>
<i>Single Linkage</i>	0,831
<i>Complete Linkage</i>	0,745
<i>Average Linkage</i>	0,833
<i>Ward Linkage</i>	0,608
<i>Centroid Linkage</i>	0,811
<i>Median Linkage</i>	0,537

Pada Tabel 7 dapat dilihat bahwa nilai korelasi *chopenetic* yang tertinggi yaitu 0,833 pada algoritma

Average Linkage sehingga algoritma tersebut digunakan dalam penelitian ini.

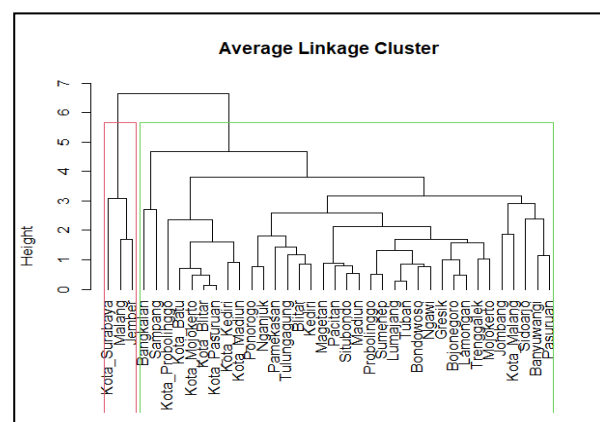
Setelah diketahui algoritma yang optimal pada analisis *cluster*, maka langkah berikutnya yaitu menentukan banyak *cluster* yang optimal berdasarkan validasi internal. Berikut dibawah ini merupakan hasil dari validasi *cluster* internal.

Tabel 8. Validasi *Cluster* Internal

	Nilai	Jumlah Cluster
<i>Connectivity</i>	5,015	2
<i>Dunn</i>	0,388	2
<i>Silhouette</i>	0,519	2

Pada Tabel 8 dapat dilihat bahwa dari 2 sampai 6 *cluster* baik *connectivity index*, *dunn index*, dan *silhouette index* memiliki hasil perolehan yang sama yaitu sebanyak 2 *cluster* yang memiliki nilai optimal. Sehingga pada pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan indikator pendidikan dikelompokkan menjadi 2 *cluster*.

Berikut ini merupakan hasil analisis *cluster* pada data indikator pendidikan di Provinsi Jawa Timur yang dikelompokkan menjadi 2 *cluster* dapat dilihat pada gambar dendrogram dibawah.



Gambar 1. Dendrogram Analisis Average Linkage Cluster

Gambar 1 merupakan dendrogram visualisasi *cluster* optimal dengan metode *average linkage* yang terbentuk berdasarkan indikator pendidikan di Provinsi Jawa Timur pada Tahun 2021. Berdasarkan validasi analisis *cluster* secara internal diketahui bahwa *cluster* yang optimal yaitu sebanyak 2. Pada dendrogram yang terbentuk dapat diketahui juga anggota pada masing-masing *cluster*, dimana *cluster* pertama yaitu ditandai dengan batas merah dan *cluster* kedua ditandai dengan garis hijau. Berikut ini merupakan tabel daftar anggota pada masing-masing *cluster* dapat dilihat pada Tabel 8.

Tabel 9. Daftar Anggota *Cluster*

	Jumlah Anggota	Anggota <i>Cluster</i>
Cluster 1	3	Kota Surabaya, Malang, Jember
Cluster 2	35	Bangkalan, Sampang, Kota Probolinggo, Kota Batu, Kota Mojokerto, Kota Blitar, Kota Pasuruan, Kota Kediri, Kota Madiun, Ponorogo, Nganjuk, Pemakasan, Tulungagung, Blitar, Kediri, Magetan, Pacitan, Situbondo, Madiun, Probolinggo, Sumenep, Lumajang, Tuban, Bondowoso, Ngawi, Gresik, Bojonegoro, Lamongan, Trenggalek, Mojokerto, Jombang, Kota Malang, Sidoarjo, Banyuwangi, Pasuruan

Pada Tabel 9 diperoleh informasi bahwa pengelompokan berdasarkan indikator pendidikan dari 38 Kabupaten/Kota di Provinsi Jawa Timur terbagi menjadi dua *cluster* dimana pada *cluster* pertama beranggotakan 3 Kabupaten/Kota dan *cluster* kedua 35 Kabupaten Kota. *Cluster* pertama yaitu pada Kota Surabaya, Malang, Jember serta *cluster* kedua yaitu pada Bangkalan, Sampang, Kota Probolinggo, Kota Batu, Kota Mojokerto, Kota Blitar, Kota Pasuruan, Kota Kediri, Kota Madiun, Ponorogo, Nganjuk, Pemakasan, Tulungagung, Blitar, Kediri, Magetan, Pacitan, Situbondo, Madiun, Probolinggo, Sumenep, Lumajang, Tuban, Bondowoso, Ngawi, Gresik, Bojonegoro, Lamongan, Trenggalek, Mojokerto, Jombang, Kota Malang, Sidoarjo, Banyuwangi, Pasuruan.

5. KESIMPULAN

Pada pengelompokan Kabupaten/Kota di Provinsi Jawa Timur berdasarkan indikator pendidikan dengan menggunakan metode *Agglomerative Hierarchical Clustering* dihasilkan bahwa algoritma optimal yang digunakan yaitu pada algoritma *average linkage* dengan nilai korelasi *cophenetic* yang diperoleh sebesar 0,807. Berdasarkan validasi *cluster* baik internal maupun *stability* serta didasarkan pada karakteristik yang dimiliki oleh Kabupaten/ Kota di Provinsi Jawa Timur maka jumlah *cluster* yang representative yaitu 2 *cluster*, dimana pada setiap kelompok jumlah anggotanya tidak seimbang dan lebih banyak pada *cluster* yang kedua. Sehingga dapat ditarik kesimpulan juga berdasarkan beberapa indikator pendidikan yang ada pada penelitian ini bahwa mayoritas pendidikan di Kabupaten/Kota di Provinsi Jawa Timur memiliki pendidikan yang merata.

DAFTAR PUSTAKA

- Asiska, N., Satyahadewi, N., dan Perdana, H. (2019) Pencarian *Cluster* Optimum pada Single Linkage, Complete Linkage, dan Average Linkage, Vol 08 No. 3 pp. 393-398, Buletin Ilmiah Math, Stat, dan Terapannya (Bimaster).
- BPS (2021) Statistik Indonesia dalam Infografis (*Statistical Yearbook of Indonesia in Infographics*), Badan Pusat Statistika, Jakarta.
- Fathia, A. N., Rahmawati, R., dan Tarno. (2016) Analisis Klaster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa menggunakan Metode Ward dan Single Linkage, Vol 5 No 4 pp. 801-810, Jurnal Gaussian.
- Goreti, M. Novia, Y. dan Wahyuningsih, S. (2016) Perbandingan Hasil Analisis *Cluster* dengan Menggunakan Metode *Single Linkage* dan Metode *C-Means*, Vol 7 No. 1 pp 9-16, Jurnal EKSPONENSIAL.
- Mariana (2013) Analisis Komponen Utama, Vol 2 No. 2 pp. 2303-0992. Jurnal Matematika dan pembelajarannya.
- Mu'afa, S. F. dan Ulinnuha, N. (2019) Perbandingan Metode Single Linkage, Complete Linkage dan Average Linkage dalam Pengelompokan Kecamatan Berdasarkan Variabel Jenis Ternak Kabupaten Sidoarjo. Vol 4 No. 2 pp. 2502-3470, Jurnal Ilmiah Bidang Teknologi Informasi dan komunikasi (INFORM).
- Murtagh, F. Contreras, P. (2012) *Algorithms for hierarchical clustering: an overview*. Vol 00, WIREs Data Mining and Knowledge Discovery.
- Nafisah Q. dan Chandra N. E. (2017) Analisis *Cluster Average Linkage* Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur. Vol 3 No.2 pp. 31-36, Zeta-Math Journal.
- Ningsih, S. Wahyuningsih, S. dan Nasution, Y. N. (2016) Perbandingan Kinerja Metode *Complete Linkage* dan *Average Linkage* dalam Menentukan Hasil Analisis *Cluster*, Prosiding Seminar Sains dan Teknologi, Vol 1 No. 1, FMIPA Unmul.
- Roux M. (2018) *A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms*. Vol 35 pp. 345-366. Springer.
- Saracli, S. Dogan, N. dan Dogan, I. (2013) *Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation*, Vol 203 pp. 1-8. Journal of Inequalities and Applications. SpringerOpen.
- Widodo, E. Sari, N. N. Hidayati, I, Yubinas, F, Yuniarti, M. Novyantika, D. (2018) Analisis *Cluster* Penderita Disabilitas Mental di Provinsi Daerah Istimewa Yogyakarta Tahun 2016, prosiding KNPMP III 2018, pp. 577-586, Pendidikan Matematika FMIPA UMS.