

Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan Faktor-Faktor yang Mempengaruhi AIDS

Rakhmah Wahyu Maya, Anisa Ramadhan, Ikacipta Mega Ayuputri, Bambang Wijnarko Otok
 Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data
 Institut Teknologi Sepuluh Nopember
 Jl. Arif Rahman Hakim, Surabaya 60111 Indonesia
 e-mail: bambang_wo@statistika.its.ac.id

Abstrak— *Acquired Immune Deficiency Syndrome (AIDS)* merupakan salah satu penyakit mematikan yang sampai saat ini belum ditemukan vaksin pencegahan atau obat untuk menyembuhkannya. *AIDS* disebabkan oleh virus *Human Immunodeficiency Virus (HIV)*. Virus tersebut menyerang sistem kekebalan tubuh manusia. Sebagian besar orang tertular *AIDS* dikarenakan faktor pendidikan, kemiskinan, kesehatan yang didapatkan oleh masyarakat dan tenaga kesehatan. Oleh karena itu, peneliti menganalisis cluster faktor-faktor yang mempengaruhi penyakit *AIDS* di Jawa Timur pada tahun 2008. Analisis Cluster digunakan untuk mengelompokkan wilayah terjadinya penyakit *AIDS*. Penelitian tersebut membandingkan hasil pengujian analisis cluster Hirarki dengan menggunakan metode *single linkage*, *complete linkage* dan *average linkage*. Berdasarkan hasil analisis dapat diketahui bahwa jumlah cluster optimum yang terbentuk adalah 3 cluster. Selanjutnya dilakukan analisis manova. Berdasarkan hasil manova dapat diketahui bahwa faktor Cluster berpengaruh terhadap variabel yang mempengaruhi penyakit *AIDS* di Jawa Timur.

Kata Kunci—*AIDS, Average linkage, Cluster analysis, Complete linkage, HIV, K-means, and Single linkage*

I. PENDAHULUAN

Acquired Immune Deficiency Syndrome (AIDS) merupakan salah satu penyakit mematikan yang sampai saat ini belum ditemukan vaksin pencegahan atau obat untuk menyembuhkannya. *AIDS* disebabkan oleh virus *Human Immunodeficiency Virus (HIV)*. Virus tersebut menyerang sistem kekebalan tubuh manusia [1].

Pada 2012, Kemenkes memperkirakan ada 591.718 orang terinfeksi HIV di Indonesia. Namun pada akhir Maret 2014, hanya ada 134.053 orang diketahui terinfeksi HIV melalui tes sukarela. Pada waktu yang sama, 54.231 orang dilaporkan sudah sampai ke stadium *AIDS* dan 9.615 diketahui sudah meninggal dunia akibatnya. Sebagian besar orang tertular *AIDS* melalui hubungan seks dengan orang yang terinfeksi HIV, penggunaan jarum suntik bergantian dengan orang yang terinfeksi HIV, kelahiran oleh ibu yang terinfeksi [1]. Sebagian besar orang tertular *AIDS* dikarenakan faktor pendidikan, kemiskinan, dan kesehatan yang didapatkan oleh masyarakat.

Oleh karena itu, peneliti menganalisis cluster faktor-faktor yang mempengaruhi penyakit *AIDS* di Jawa Timur pada tahun 2008. Variabel yang digunakan adalah persentase penduduk yang tamat SMA di tiap kabupaten/kota di Jawa Timur, persentase penduduk miskin di tiap kabupaten/kota di Jawa Timur, persentase sarana kesehatan (RS dan puskesmas) di tiap kabupaten/kota di Jawa Timur. Analisis Cluster digunakan untuk mengelompokkan wilayah terjadinya penyakit *AIDS*. Penelitian tersebut membandingkan hasil pengujian analisis

cluster Hirarki dengan menggunakan metode *single linkage* dan *complete linkage* dengan jarak Euclidean. Adanya penelitian ini diharapkan mengetahui untuk mengelompokkan wilayah terjadinya penyakit *AIDS* di Jawa Timur sehingga pemerintah dapat mengambil kebijakan yang tepat untuk mengatasi penyebaran penyakit *AIDS* yang terus meningkat.

II. TINJAUAN PUSTAKA

A. Statistika Deskriptif

Statistika deskriptif adalah serangkaian metode yang berkaitan dengan pengumpulan dan penyajian suatu data sehingga dapat memberikan informasi yang berguna tanpa adanya inferensia atau kesimpulan [2]. Pada penelitian ini akan digunakan metode statistika deskriptif rata-rata dan standart deviasi.

Rata-rata dari suatu data diperoleh dari membandingkan jumlah semua nilai datum dengan banyak nilai datumnya.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \tag{1}$$

dimana:

\bar{x} = Rata-rata

x_i = Nilai data ke-i

n = Ukuran data (Banyak datum yang diamati)

Standar deviasi merupakan akar dari varians, dimana varians merupakan ukuran penyebaran data yang mengukur rata-rata jarak kuadrat semua titik pengamatan terhadap titik pusat (rataaan).

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \tag{2}$$

dimana:

s = Standar Deviasi

x_i = Nilai data ke-i

\bar{x} = Rata-rata

n = Banyak data

B. Analisis Cluster

Analisis kelompok adalah satu metode yang digunakan untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok berdasarkan karakteristik yang dimiliki dengan ciri tertentu relatif homogen, sehingga objek dalam kelompok memiliki kesamaan yang sama sedangkan objek antar kelompok tidak memiliki kesamaan yang sama [3]. Cluster atau pengelompokan yang baik adalah cluster yang menunjukkan ciri sebagai berikut.

1. Homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok.
2. Heterogenitas (perbedaan) yang tinggi antar kelompok yang satu dengan kelompok yang lain.

Dalam analisis *cluster*, terdapat banyak metode untuk mengelompokkan observasi ke dalam *cluster*. Secara umum metode pengelompokan dalam analisis *cluster* dibedakan menjadi metode hierarki dan non hierarki. Metode hierarki digunakan apabila belum ada informasi jumlah *cluster* yang dipilih. Sedangkan metode non hierarki bertujuan mengelompokkan n objek ke dalam k *cluster* ($k < n$) dimana nilai k telah ditentukan sebelumnya [3]. Jarak *Euclidian* merupakan salah satu konsep jarak yang sering dipilih dalam analisis *cluster* dengan rumusan sebagai berikut.

$$d(x, y) = \left[(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2 \right]^{1/2} \quad (3)$$

Terdapat metode penggabungan yang bisa digunakan dalam analisis *cluster* hirarki yaitu sebagai berikut.

1. *Single Linkage*

Single linkage merupakan metode *cluster non* yang mana kelompok-kelompok dibentuk dari *similarities*/kemiripan yang paling besar. Menemukan jarak terpendek dalam dan menggabungkan objek-objek yang bersesuaian misalnya U dan V, untuk mendapatkan *cluster* (UV). Untuk langkah dari algoritma di atas jarak-jarak antara (UV) dan *Cluster* W yang lain dihitung dengan persamaan (4) sebagai berikut [3].

$$d_{(uv)w} = \min(d_{uw}, d_{vw}) \quad (4)$$

Berdasarkan persamaan (3), nilai d_{UW} dan d_{VW} secara berturut-turut adalah jarak terpendek antar *cluster* U dan W dan *cluster* V dan W.

2. *Complete Linkage*

Complete Linkage merupakan metode *cluster non hierarki* yang memastikan bahwa semua item-item dalam satu *Cluster* berada dalam jarak paling jauh satu sama lain. Untuk langkah dari algoritma di atas jarak-jarak antara (UV) dan *Cluster* W yang lain dihitung dengan persamaan (5) sebagai berikut [3].

$$d_{(uv)w} = \max(d_{uw}, d_{vw}) \quad (5)$$

Berdasarkan persamaan (5), nilai d_{UW} dan d_{VW} secara berturut-turut adalah jarak terpendek antar *cluster* U dan W dan *cluster* V dan W.

3. *Average Linkage*

Prosedur pada *average linkage* hampir sama dengan *single linkage* maupun *complete linkage*, namun kriteria yang digunakan adalah rata-rata jarak seluruh individu dalam suatu *Cluster* dengan jarak seluruh individu dalam *Cluster* yang lain dengan rumus sebagai berikut [3].

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(uv)} N_w} \quad (6)$$

C. *Uji Bartlett*

Uji Bartlett digunakan untuk mengetahui ada atau tidaknya hubungan antar variabel dalam kasus multivariat. Hipotesis pengujian yang digunakan pada uji *Bartlett* adalah sebagai berikut [3].

- $H_0 : \rho = I$ (tidak terdapat hubungan antar variabel)
 $H_1 : \rho \neq I$ (terdapat hubungan antar variabel)

Statistik uji yang digunakan dalam uji ini dituliskan dalam Persamaan 7 berikut.

$$\chi_{hitung}^2 = \left[n - 1 - \frac{2p + 5}{6} \right] \ln |R| \quad (7)$$

$|R|$ merupakan determinan dari matriks korelasi, n merupakan banyaknya sampel, dan p banyaknya variabel. Jika gagal tolak H_0 , maka variabel tidak layak untuk dilakukan metode multivariat terutama metode analisis komponen utama dan analisis faktor. Jika ditetapkan taraf signifikan sebesar α maka tolak H_0 yang artinya adanya hubungan antar variabel dengan ketentuan yaitu $\chi_{hitung}^2 > \chi_{\alpha; 1/2 p(p-1)}^2$.

D. *Uji Homogenitas*

Uji homogenitas digunakan untuk menguji homogenitas matriks varians-kovarians. Berikut ini adalah hipotesis yang digunakan.

$$H_0 : \sum_1 = \sum_2 = \dots = \sum_g = \sum$$

$$H_1 : \text{minimal ada satu } \sum_i \neq \sum_j, i = 1, 2, \dots, g$$

Statistik uji:

$$M = \sum_{l=1}^g (n_l - 1) \ln |S| - \sum_{l=1}^g (n_l - 1) \ln |S_l| \quad (8)$$

$$S = \frac{1}{N} \sum_{l=1}^g (n_l - 1) \ln S_l; N = \sum_{l=1}^g (n_l - g)$$

dan,

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \left(\sum_{l=1}^g \frac{1}{(n_l - 1)} - \frac{1}{\sum_{l=1}^g (n_l - 1)} \right) \quad (9)$$

Keterangan:

n_l = banyak data sampel ke- l

S_l = matriks sampel ke- l

$|S|$ = nilai mutlak matriks S

g = banyak populasi

Daerah kritis:

$$\text{Tolak } H_0 \text{ jika } MC^{-1} > \chi_{\left(\frac{1}{2}(g-1)(p+1)\right)}^2(\alpha), \text{ sehingga}$$

dapat disimpulkan bahwa matriks varians-kovarians dari l -populasi adalah heterogen.

E. *MANOVA One-way*

MANOVA One-way merupakan pengembangan dari *ANOVA One-way*. *One-way MANOVA* digunakan untuk menguji apakah ke- g jenis populasi (dari satu populasi yang sama) menghasilkan vektor rata-rata yang sama untuk p variabel perlakuan atau variabel dependen yang diamati dalam penelitian [3].

Tabel 1. *MANOVA One-way*

Source of Variation	Matrix of SSCP	Degrees of freedom
Treatment	$B = \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})(\bar{x}_l - \bar{x})'$	$g-1$
Error	$W = \sum_{l=1}^g \sum_{j=1}^{n_l} n_l (\bar{x}_{lj} - \bar{x}_l)(\bar{x}_{lj} - \bar{x}_l)'$	$\sum_{l=1}^g n_l - g$
Total	$B + W = \sum_{l=1}^g \sum_{j=1}^{n_l} n_l (\bar{x}_{lj} - \bar{x})(\bar{x}_{lj} - \bar{x})'$	$\sum_{l=1}^g n_l - 1$

Prosedur pengujian MANOVA *One-way* adalah sebagai berikut.

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_g$$

$$H_1 : \text{minimal ada satu } \tau_i = \tau_j, i = 1, 2, \dots, g$$

Statistik uji :

dengan menggunakan nilai *F* dari distribusi *Wilks' Lambda*

Tabel 2. Distribusi dari *Wilks' Lambda*

No of Variables	No of groups	Sampling distribution for multivariate normal data
$p = 1$	$g \geq 2$	$\left(\frac{\sum n_l - g}{g - 1} \right) \left(\frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F_{(g-1), \sum n_l - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\sum n_l - g - 1}{g - 1} \right) \left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F_{2(g-1), 2(\sum n_l - g - 1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\sum n_l - p - 1}{p} \right) \left(\frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F_{p, \sum n_l - p - 1}$
$p \geq 1$	$g = 3$	$\left(\frac{\sum n_l - p - 2}{p} \right) \left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}} \right) \sim F_{2p, 2(\sum n_l - p - 2)}$

Dengan,

$$\Lambda^* = \frac{|W|}{|W + B|} = \frac{\left| \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)(x_{lj} - \bar{x}_l)' \right|}{\left| \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})(x_{lj} - \bar{x})' \right|} \quad (10)$$

F. Acquired Immune Deficiency Syndrome (AIDS)

Acquired Immune Deficiency Syndrome (AIDS) adalah sekumpulan gejala dan infeksi yang timbul karena rusaknya sistem kekebalan tubuh manusia akibat infeksi virus HIV. Pada 2012, Kemenkes memperkirakan ada 591.718 orang terinfeksi HIV di Indonesia. Namun pada akhir Maret 2014, hanya ada 134.053 orang diketahui terinfeksi HIV melalui tes sukarela. Pada waktu yang sama, 54.231 orang dilaporkan sudah sampai ke stadium AIDS dan 9.615 diketahui sudah meninggal dunia akibatnya. Sebagian besar orang tertular AIDS melalui hubungan seks dengan orang yang terinfeksi HIV, penggunaan jarum suntik bergantian dengan orang yang terinfeksi HIV, kelahiran oleh ibu yang terinfeksi [1]

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan dalam laporan praktikum mengenai pengelompokan kabupaten atau kota berdasarkan faktor-faktor yang mempengaruhi AIDS di Jawa Timur pada tahun 2008 dari Tugas Akhir yang berjudul "Perbandingan Analisis Regresi Poisson, *Generalized Poisson Regression* dan *Geographically Weighted Poisson Regression* (Studi Kasus: Pemodelan Jumlah Kasus Aids di Jawa Timur Tahun 2008)" oleh Novia Assriyanti (1307100074). Data yang digunakan, diambil pada hari Jumat 6 April 2018 pukul 10.00 di Ruang Baca Jurusan Statistika ITS Surabaya.

B. Variabel Penelitian

Variabel dependen pada penelitian ini adalah faktor-faktor yang mempengaruhi penyakit AIDS di Jawa Timur Berikut variabel yang digunakan dalam penelitian.

Tabel 1 Variabel yang digunakan pada penelitian

Variabel	Keterangan
x_1	Persentase penduduk yang tamat SMA di tiap kabupaten/kota di Jawa Timur.
x_2	Persentase penduduk miskin di tiap kabupaten/kota di Jawa Timur.
x_3	Persentase sarana kesehatan (RS dan puskesmas) di tiap kabupaten/kota di Jawa Timur.
x_4	Persentase tenaga medis di tiap kabupaten/kota di Jawa Timur

C. Metode Analisis Data

Metode analisis yang digunakan dalam penelitian adalah sebagai berikut :

1. Mengumpulkan data sekunder.
2. Mengelompokkan variable ke dalam kelompok sebanyak yang ditentukan untuk analisis *cluster* teknik hierarki.
3. Mengelompokkan variable ke dalam kelompok sebanyak yang ditentukan untuk analisis *cluster* teknik non hierarki.
4. Melakukan analisis Manova berdasarkan kelompok yang telah terbentuk.
4. Menginterpretasi hasil analisis data dan membuat kesimpulan.

IV. HASIL DAN PEMBAHASAN

A. Karakteristik Data

Karakteristik data adalah analisis yang pertama kali dilakukan. Karakteristik data merupakan hasil analisis yang memberikan informasi tentang deskripsi data. Karakteristik dari data jumlah kasus AIDS di tiap Kabupaten/Kota di Jawa Timur pada tahun 2008.

Tabel 1 Karakteristik Data

Variabel	Mean	St.Deviasi
X_1	34,074	7,657
X_2	7,876	4,170
X_3	21,631	7,116
X_4	1,014	0,618

Tabel 1 diatas menunjukkan pada variabel X_1 yaitu persentase jumlah penduduk yang tamat SMA/SMK di provinsi Jawa Timur sebesar 34,074. Nilai tersebut lebih rendah dari 50 dan memiliki variabilitas yang cukup besar. Hal ini menunjukkan masih banyak penduduk di Provinsi Jawa Timur yang berpendidikan rendah. Nilai persentase penduduk miskin dibawah 10%. Nilai tersebut memang dibawah nilai nasional, namun nilai 7,876 tersebut merupakan cukup besar jika dilihat dari jumlah penduduk Provinsi Jawa Timur yang terbesar kedua di Indonesia. Nilai variabel X_3 dan X_4 yang mengindikasikan sarana dan tenaga kesehatan masih rendah.

B. Analisis Cluster Hierarki

Analisis *Cluster* digunakan untuk mengelompokkan wilayah terjadinya penyakit AIDS. Berikut adalah analisis dan pembahasannya.

1. Single Linkage

Langkah pertama yang dilakukan adalah deteksi missing value dengan hasil sebagaimana Tabel 2 berikut.

Tabel 2. Hasil Proses *Missing Value* Metode *Single Linkage*

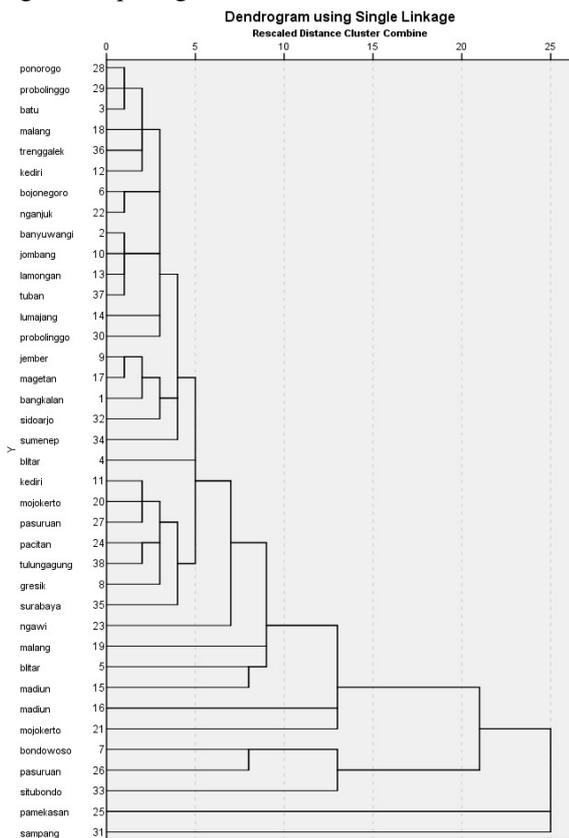
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
38	100%	0	0%	38	100%

Berdasarkan Tabel 2 dapat diketahui bahwa tidak terdapat missing value dengan persentase valid sebesar 100%. Selanjutnya dilakukan analisis *cluster* dengan hasil sebagai berikut.

Tabel 4. *Agglomeration Schedule* Metode *Single Linkage*

Stage	Cluster Combined		Coefficient	Stage Cluster First Appears		Next Stage
	Cluste r 1	Cluste r 2		Cluste r 1	Cluste r 2	
1	28	29	2.518	0	0	4
2	6	22	2.713	0	0	15
3	2	10	4.225	0	0	7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	1	7	96.786	34	33	36
36	1	25	118.623	35	0	37
37	1	31	119.27	36	0	0

Berdasarkan Tabel 4 dapat diketahui bahwa kabupaten ke-28 dan ke-29 digabungkan menjadi satu *cluster* karena memiliki jarak terdekat, kemudian dilanjutkan kepada *stage* 2. Anggota *cluster* pada *stage* pertama ditambahkan kabupaten ke-6 dan ke-22, begitu pula sampai *stage* 37. Pengelompokan juga dapat dilihat pada dendrogram sebagaimana pada gambar 1 berikut.



Gambar 1 Hasil *cluster* metode *single linkage*

Gambar 1 menunjukkan dari 38 kabupaten/kota dapat terbentuk 37 hingga 2 *cluster*. Jika terbentuk dua *cluster* maka satu *cluster* beranggotakan satu anggota yaitu hanya kabupaten sampang. Hal ini menunjukkan menurut metode *single linkage* keempat variabel yang mempengaruhi jumlah kasus AIDS tersebut berbeda dari *cluster* lainnya yang beranggota 37 kabupaten/kota di Provinsi Jawa Timur.

2. Complete Linkage

Langkah pertama yang dilakukan adalah deteksi missing

value dengan hasil sebagaimana Tabel 2 berikut.

Tabel 5. Hasil Proses *Missing Value* Metode *Complete Linkage*

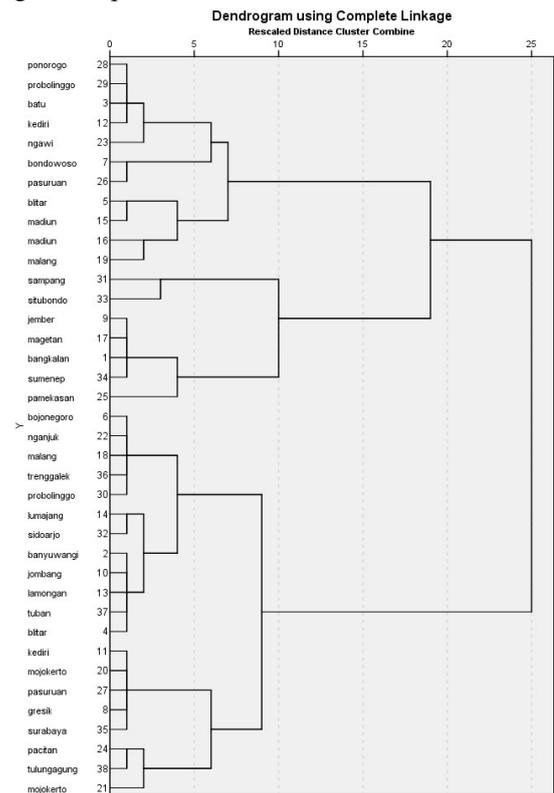
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
38	100%	0	0%	38	100%

Berdasarkan Tabel 5 dapat diketahui bahwa tidak terdapat missing value dengan persentase valid sebesar 100%. Selanjutnya dilakukan analisis *cluster* dengan hasil sebagai berikut.

Tabel 6. *Agglomeration Schedule* Metode *Complete Linkage*

Stage	Cluster Combined		Coefficient	Stage Cluster First Appears		Next Stage
	Cluste r 1	Cluste r 2		Cluste r 1	Cluste r 2	
1	28	29	2.518	0	0	6
2	6	22	2.713	0	0	11
3	2	10	4.225	0	0	10
⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	1	31	481.745	30	27	36
36	1	3	987.383	35	33	37
37	1	2	1326.484	36	34	0

Berdasarkan Tabel 4 dapat diketahui bahwa kabupaten ke-28 dan ke-29 digabungkan menjadi satu *cluster* karena memiliki jarak terdekat, kemudian dilanjutkan kepada *stage* 2. Anggota *cluster* pada *stage* pertama ditambahkan kabupaten ke-6 dan ke-22, begitu pula sampai *stage* 37. Pengelompokan juga dapat dilihat pada dendrogram sebagaimana pada Gambar 2 berikut.



Gambar 2 Hasil *cluster* metode *complete linkage*

Gambar 2 menunjukkan dengan menggunakan jarak yang sama tapi dengan metode berbeda menghasilkan pengelompokan yang berbeda. Jika terbentuk dua *cluster* maka satu *cluster* beranggotakan 18 kabupaten/kota dan *cluster* yang lain beranggotakan 20 kabupaten/kota.

3. Average Linkage

Langkah pertama yang dilakukan adalah deteksi missing value dengan hasil sebagaimana Tabel 3 berikut.

Tabel 7. Hasil Proses *Missing Value* Metode *Average Linkage*

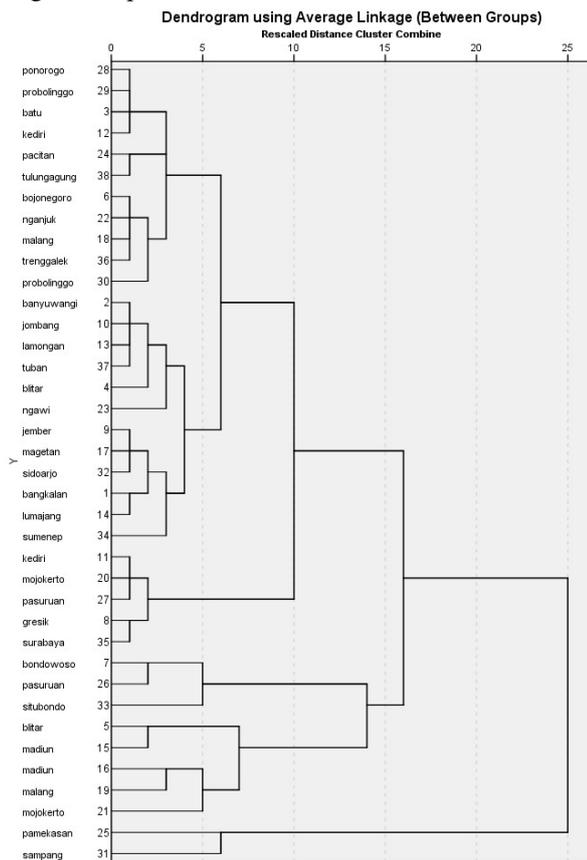
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
38	100%	0	0%	38	100%

Berdasarkan Tabel 7 dapat diketahui bahwa tidak terdapat missing value dengan persentase valid sebesar 100%. Selanjutnya dilakukan analisis *cluster* dengan hasil sebagai berikut.

Tabel 8. *Agglomeration Schedule* Metode *Average Linkage*

Stage	Cluster Combined		Coefficient	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
	r 1	r 2		r 1	r 2	
1	28	29	2.518	0	0	5
2	6	22	2.713	0	0	12
3	2	10	4.225	0	0	7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	5	7	302.769	33	30	36
36	1	5	363.520	34	35	37
37	1	25	577.179	36	31	0

Berdasarkan Tabel 8 dapat diketahui bahwa kabupaten ke-28 dan ke-29 digabungkan menjadi satu *cluster* karena memiliki jarak terdekat, kemudian dilanjutkan kepada *stage* 2. Anggota *cluster* pada *stage* pertama ditambahkan kabupaten ke-6 dan ke-22, begitu pula sampai *stage* 37. Pengelompokan juga dapat dilihat pada dendrogram sebagaimana pada Gambar 3 berikut.



Gambar 3 Hasil *cluster* metode *average linkage*

C. Pemilihan Metode Terbaik

Pemilihan metode terbaik dilakukan berdasarkan nilai pseudo F dan icdrate berdasarkan Tabel 9 berikut.

Tabel 9 Pemilihan Metode Terbaik

Metode	Banyaknya Cluster	Pseudo-F	Icdrate
Single Linkage	2 Cluster	4,337	0,336
	3 Cluster	4,124	0,267
	4 Cluster	6,891	0,285
Complete Linkage	2 Cluster	9,651	1,429
	3 Cluster	15,599	0,889
	4 Cluster	14,258	0,609
Average Linkage	2 Cluster	3,027	0,915
	3 Cluster	18,113	0,885
	4 Cluster	9,663	0,905

Berdasarkan pada Tabel 9 dapat diketahui bahwa pada metode complete linkage dan average linkage menghasilkan kesimpulan yg sama yaitu terbentuk 3 *cluster* sedangkan pada metode single linkage terbentuk 4 *cluster*. Sehingga dapat diketahui jumlah *cluster* yang optimal adalah 3 *cluster*.

D. Analisis Manova

Analisis MANOVA menggunakan kelompok yang telah diketahui melalui nilai *pseudo-F*. Kelompok pertama memiliki anggota Kabupaten Sampang, kelompok kedua adalah Kabupaten Pamekasan, kelompok ketiga adalah Kabupaten Situbondo, Bondowoso dan Pasuruan, sedangkan sisanya masuk kedalam kelompok keempat. Hasil analisis MANOVA terdiri dari uji asumsi distribusi normal, uji Bartlett, dan analisis MANOVA *one way*. Berikut adalah hasil analisis uji asumsi distribusi normal.

Tabel 10. Hasil Uji Distribusi Normal

Korelasi Pearson	Nilai kritis	Keputusan
0,865	0,809	tolak H ₀

Taraf signifikan yang digunakan sebesar 0,05. Tabel 10 dapat diketahui bahwa nilai korelasi Pearson lebih besar dari nilai kritis, artinya data berdistribusi normal multivariat. Selanjutnya dilakukan pengujian Bartlett ntuk mengetahui hubungan antar variabel yang diperoleh hasil sebagai berikut.

Tabel 11. Hasil Uji Bartlett

P-value	Taraf Signifikan	Keputusan
0	0,05	Tolak H ₀

Tabel 11 menunjukkan bahwa *p-value* kurang dari taraf signifikan, artinya data ada hubungan yang signifikan antar variabel yang mempengaruhi penyakit AIDS di Jawa Timur. Selanjutnya adalah analisis manova yang diketahui hasil analisis adalah seperti berikut.

Tabel 12. Hasil MANOVA

F	F _{tabel}	P-value	Taraf Signifikansi	Keputusan
8,167	1,827	0	0,05	Tolak H ₀

Tabel 12 menjelaskan bahwa nilai F lebih besar dari F_{tabel} dan *p-value* kurang dari 0,05. Hal tersebut dapat diartikan faktor *Cluster* berpengaruh terhadap variabel yang mempengaruhi penyakit AIDS di Jawa Timur, sehingga dilanjutkan analisis individu.

Tabel 13. Hasil Uji Individu

Variabel	P-value	Taraf Signifikansi	Keputusan
X1	0,020	0,05	Tolak H ₀
X2	0,000	0,05	Tolak H ₀
X3	0,000	0,05	Tolak H ₀
X4	0,0113	0,05	Tolak H ₀

Informasi pada Tabel 13 dapat disimpulkan bahwa hasil pengelompokan *cluster* berpengaruh pada variabel yang mempengaruhi penyakit AIDS di Jawa Timur.

V KESIMPULAN DAN SARAN

Berdasarkan analisis diatas dapat disimpulkan jumlah *cluster* optimum yang terbentuk adalah 3 *cluster*. Selanjutnya dilakukan analisis manova. Berdasarkan hasil manova dapat diketahui bahwa faktor *cluster* berpengaruh terhadap variabel yang mempengaruhi penyakit AIDS di Jawa Timur.

Dalam penelitian ini diharapkan agar lebih memahami materi yang dibahas, agar tidak terdapat kesalahan dalam pengolahan data ataupun dalam menginterpretasikannya. Panduan yang jelas dalam melakukan penelitian ini sangat penting serta dalam menganalisis data hasil pengamatan diperlukan kecermatan dan ketelitian

DAFTAR PUSTAKA

- [1] Pasal 1 Peraturan Menteri Kesehatan Republik Indonesia Nomor 21 Tentang Penanggulangan HIV dan AIDS. Tahun 2013
- [2] R. Walpole, Intoduction to Statistics, New York: Macmillan Publishing Co. Inc, 1995.
- [3] R. Johnson and D. Wichern, Applied Multivariate Statistical Analysis, New Jersey: Pearson Education, 2007.