

Analisis Diskriminan untuk Validasi *Cluster* pada Studi Kasus Pengelompokan Kecamatan di Kabupaten Jember Berdasarkan Status Kemiskinan

Discriminant Analysis for Cluster Validation in A Case Study of District Grouping in Jember Regency Based on Poverty

Fikriana Nur Istiqomah, I Made Tirta*), Dian Anggraeni
 Jurusan Matematika FMIPA Universitas Jember
 *E-mail: itirta.fmipa@unej.ac.id

ABSTRACT

Cluster validation is a procedure to evaluate the results of cluster analysis quantitatively and objectively on a data. The validation process is very important to get the results of a good and appropriate grouping. In the validation process, the author uses internal validation, stability, and discriminant analysis test. This study aims to obtain validation results from the hierarchy and kmeans method. This data grouping uses "iris" simulation data, which results from the grouping method used can be applied to the original data to see which validation method is used for all data and produce an optimal grouping. The result of the study show that in the "iris" data, a single linkage link is an appropriate grouping method because the result of the grouping are optimal for all validations and classification of group members whose groups are significant. In District poverty data in Jember Regency with a single linkage link optimal grouping was obtained and complete linkage links were also used as a method that resulted in optimal grouping for all validation. Cluster validation discriminant analysis test is appropriate for various types of data in general and shows that single linkage methods are better than other methods for grouping and validation methods for "iris" data and District data in Jember Regency based on variables of poverty status.

Keywords: Cluster Analysis, Diskriminant Analysis, Multivariate Analysis, Validation Cluster.

PENDAHULUAN

Analisis Multivariat merupakan salah satu jenis analisis statistika yang dapat digunakan untuk menganalisis data dengan variabel independent dan/atau variabel dependent-nya lebih dari satu. Analisis cluster bertujuan untuk mengelompokkan objek ke dalam suatu kelompok sehingga objek yang berada dalam satu kelompok akan memiliki kesamaan yang tinggi daripada objek yang berada di kelompok lain (Sharma, 1996).

Menurut Shidqon (2012) pada analisis cluster, tidak ada metode-metode pengelompokan yang secara khusus dapat menjelaskan tentang ketepatan sebuah metode digunakan dalam proses pengelompokan data. Masing-masing metode pengelompokan memiliki kelebihan dan kelemahan, diantaranya adalah sensitif terhadap adanya *outlier*, kesulitan menangani variasi bentuk dan ukuran, serta memisahkan kelompok yang besar. Pengelompokan yang baik yaitu memiliki kesamaan individu dalam satu kelompok yang tinggi dan kesamaan antar

kelompok yang rendah serta tidak adanya *outlier*. Metode pengelompokan yang tepat penggunaannya dalam pengelompokan data dan memenuhi kriteria pengelompokan yang baik, membutuhkan satu proses terakhir yaitu proses validasi *cluster* atau validasi pengelompokan, sehingga nantinya dari pada pengelompokan data tersebut akan didapatkan hasil yang lebih baik. Menurut Brock, *et.al* (2018) validasi *cluster* ada 3 macam metode validasi yang digunakan untuk menentukan jumlah *cluster* optimal yaitu validasi internal, validasi stabilitas, dan validasi biologis. Pada penelitian ini digunakan validasi internal, validasi stabilitas dan juga uji analisis diskriminan sebagai salah satu metode untuk memvalidasi hasil solusi *cluster*.

Analisis diskriminan adalah teknik multivariate yang termasuk *dependence method*, yakni adanya variabel dependen dan variabel independen, sehingga ada variabel yang hasilnya tergantung dari data variabel independen (Santosa, 2002). Pada situasi ini, solusi *cluster* menjadi variabel dependen yang digunakan untuk mendapatkan pengelompokan

yang optimal. Adapun pengolahan data dilakukan dengan menggunakan akses *web online Virtual Statistics Laboratory* Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember dengan alamat <http://statslab-rshiny.fmipa.unej.ac.id/RProg/Multiv/>.

Penelitian analisis *cluster* telah dilakukan oleh beberapa peneliti diantaranya Sitepu (2011) yang menggunakan analisis *cluster* terhadap tingkat pencemaran udara pada sektor industri, dengan hasil validasi diuji menggunakan parsial F yang akan menjadi variabel pembeda dalam peng-cluster-an, Rachmatin (2014) yang menggunakan aplikasi metode-metode agglomerative dalam analisis cluster pada data tingkat polusi udara untuk hasil validasi dengan plot RMSSTD (*Root Mean Square Standart Deviation*) dan CD (*Cluster Distance*) sehingga hasil plot dari jumlah cluster yang digunakan menghasilkan jumlah *cluster* yang valid (dapat dipercaya) dan dapat dipilih 5 sampai 6 *cluster* dan Shidqon (2012) yang menggunakan metode validasi internal untuk validasi *cluster* pada pengelompokan Kecamatan di Kabupaten Jember berdasarkan status kemiskinan. Akan tetapi setiap data dan metode-metode yg digunakan pada hasil validasi memiliki karakteristik tersendiri sehingga memungkinkan adanya pemilihan data dan penentuan metode hasil validasi yang berbeda.

Analisis diskriminan untuk validasi *cluster* dapat digunakan sebagai pengklasifikasian suatu individu atau objek ke dalam suatu kelompok yang telah ditentukan sebelumnya berdasarkan variabel-variabel bebasnya^[2]. Penelitian ini, menggunakan data simulasi yang ada di web Pusat Belajar dan Laboratorium Statistika Virtual yaitu data “iris” dan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya. Data “iris” digunakan sebagai salah satu acuan pengelompokan dalam analisis data ini, karena data “iris” mempunyai klasifikasi variabel kelompok yang asli sehingga dengan lebih mudah dilihat ketepatan pengelompokan analisis data yang dihasilkan dari beberapa metode yang digunakan dengan kelompok yang telah terdefinisi sebelumnya. Berdasarkan uraian tersebut, pengelompokan dengan metode analisis data yang tepat digunakan dalam data “iris” sebelumnya dapat diaplikasikan pada pengelompokan data Kecamatan di Kabupaten Jember berdasarkan

peubah status kemiskinannya yang juga dapat di kelompokkan ke dalam beberapa *cluster* kecamatan yang mempunyai sifat relatif sama atau kemiripan sifat berdasarkan peubah status kemiskinannya. Kemiskinan didefinisikan sebagai konsekuensi dari suatu struktur masyarakat dengan penduduk yang padat, terbatasnya sumber daya, tingkat kesehatan yang rendah dan kesempatan pendidikan yang tidak merata (Ernayati dan Ita, 1996).

Tujuan penelitian ini adalah mengetahui prosedur penentuan metode analisis *Cluster* terbaik dari metode hirarki dan *kmeans*; mengetahui hasil yang didapatkan dari validasi analisis *cluster* dan hasil validasinya terhadap pengelompokan data “iris” dan data kemiskinan Kecamatan di Kabupaten Jember; serta mengetahui bahwa ujian alisis diskriminan tepat digunakan sebagai salah satu metode untuk memvalidasi hasil solusi dari analisis *cluster* untuk data secara umum.

METODE

Data yang digunakan dalam penelitian ini adalah data simulasi yang ada di web Pusat Belajar dan Laboratorium Statistika Virtual yaitu data “iris” (Data dapat dilihat pada lampira C) dan data sekunder skripsi Muhda Shidqon (2012). Sumber data berasal dari hasil sensus Badan Pusat Statistik (BPS) Kabupaten Jember tahun 2010. Pengelompokan ini dilakukan berdasarkan faktor-faktor yang mempengaruhi kemiskinan pada suatu kecamatan. Dalam penelitian ini, peneliti memilih objek pengamatan Kecamatan di Kabupaten Jember yang terdiri dari 31 kecamatan, yaitu :

X_1 = Kepadatan Penduduk (Jiwa/km²).

X_2 = Banyaknya sekolah (SD/MI (Negeri & Swasta), SLTP/MTs (Negeri & Swasta), SMA/MA (Negeri & Swasta).

X_3 = Banyaknya fasilitas kesehatan (Rumah sakit, Puskesmas).

X_4 = Banyaknya Rumah Tangga Pengguna PLN.

X_5 = Banyaknya Rumah Tangga yang Mempunyai Telepon.

X_6 = Banyaknya Industri berbadan hukum (Industri Kecil, Industri Sedang, dan Industri Besar).

X_7 = Produksi Pertanian (Ton).

Pengolahan data menggunakan akses web online *Virtual Statistics Laboratory* Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember. Adapun langkah-langkah dalam penelitian menggunakan Analisis diskriminan untuk validasi *cluster* yaitu :

- Input data pada web Pusat Belajar dan Laboratorium Statistika Virtual bagian Analisis Multivariat secara online.
- Menentukan prosedur pengelompokan Pada langkah menentukan prosedur

pengelompokan ini menggunakan metode hirarki dan metode *kmeans*.

1. Metode hirarki
Pengelompokan data dengan metode hirarki ini menggunakan tautan *single linkage*, *complete linkage* dan *average linkage*, sehingga masing-masing objek dalam data akan membentuk kelompok dan jumlah kelompok.
2. Metode *kmeans*
Pengelompokan data dengan metode *kmeans* ini menggunakan algoritma *Mac Queen*.

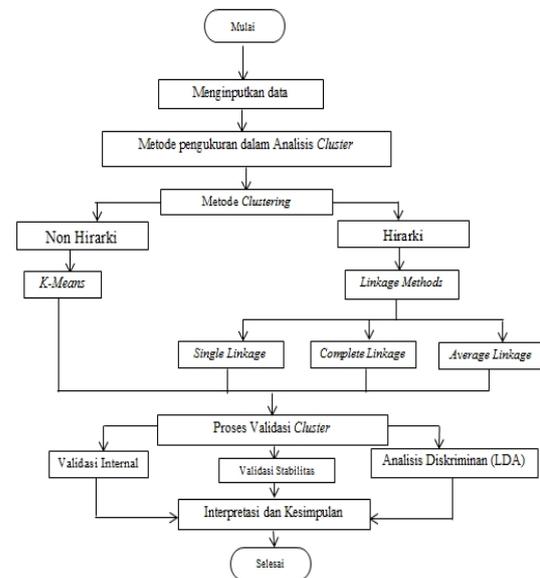
c. Melakukan validasi kelompok
Pada langkah ini, akan dilakukan validasi terhadap hasil pengelompokan untuk mendapatkan hasil pengelompokan yang optimal.

1. Validasi internal
Pada langkah sebelumnya, dipilih metode hirarki dengan tautan *single linkage*, *complete linkage* dan *average linkage* yang dilakukan denganurut. Namun, untuk setiap tautan yang dipilih langsung dilanjutkan pada langkah melakukan validasi kelompok dengan validasi internal karena setiap tautan yang dipilih dan dilakukan validasi kemungkinan didapatkan nilai pada validasi internal yaitu nilai *connectivity*, *silhouette*, dan *dunn index* yang berbeda-beda atau mungkin ada yang sama. Begitu juga pada metode *Kmeans* menggunakan algoritma *Mac Queen*.
2. Validasi Stabilitas
Pada langkah sebelumnya, dipilih metode hirarki dengan tautan *single linkage*, *complete linkage* dan *average linkage* yang dilakukan denganurut. Namun, untuk setiap tautan yang dipilih langsung dilanjutkan pada langkah melakukan validasi kelompok dengan validasi stabilitas karena setiap tautan yang dipilih dan dilakukan validasi kemungkinan didapatkan nilai pada validasi stabilitas yaitu nilai *APN*, dan *FOM* yang berbeda-beda atau mungkin ada yang sama. Begitu juga pada metode *Kmeans* menggunakan algoritma *Mac Queen*.
3. Uji analisis diskriminan.
Pada tahapan validasi *cluster* dengan uji analisis diskriminan, ada beberapa tahapan yang dilakukan; dimulai dengan melakukan pengelompokan dengan metode hirarki tautan *single linkage*, *complete linkage* dan *average linkage* secara berurutan. Selanjutnya, dilanjutkan dengan pengelompokan metode *kmeans*. Hasil yang diperoleh dari pengelompokan metode hirarki dan

kmeans digunakan sebagai variabel *dependent* dalam uji analisis diskriminan untuk validasi *cluster*. Hasil pengelompokan metode hirarki dan *kmeans* pada validasi *cluster* menggunakan uji analisis diskriminan masing-masing adalah tidak sama (berbeda).

d. Interpretasi hasil dan mengambil kesimpulan.

Skema dari langkah-langkah penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

HASIL DAN PEMBAHASAN

Data “iris” digunakan sebagai salah satu acuan pengelompokan dalam analisis data ini, karena data “iris” mempunyai klasifikasi variabel kelompok yang asli sehingga dengan lebih mudah dilihat ketepatan pengelompokan analisis data yang dihasilkan dari beberapa metode yang digunakan dengan kelompok yang telah terdefinisi sebelumnya. Berdasarkan pernyataan tersebut, pengelompokan dengan metode analisis data yang tepat digunakan dalam data “iris” sebelumnya dapat diaplikasikan pada pengelompokan data Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya yang juga dapat di kelompokkan ke dalam beberapa *cluster* kecamatan yang mempunyai sifat relatif sama atau kemiripan sifat berdasarkan peubah status kemiskinannya.

Validasi Cluster Data “iris”

Tabel 1 hasil pengelompokan data “iris” menggunakan metode hirarki tautan *single*

linkage dan *average linkage* adalah pengelompokan dengan hasil yang optimal untuk metode validasi internal karena terlihat dari nilai *connectivity* dan nilai *dunn index*-nya yang lebih kecil daripada nilai *connectivity* untuk tautan metode hirarki yang lainnya dan metode *kmeans* serta angka pada nilai *silhouette* yang mendekati 1. Validasi stabilitas, metode hirarki tautan *average linkage* dan metode *Kmeans* menghasilkan pengelompokan yang baik (stabil), karena nilai daripada APN dan FOM-nya menghasilkan nilai yang lebih kecil dari nilai APN dan FOM yang dihasilkan oleh metode hirarki tautan yang lainnya. Secara umum, Jumlah kelompok optimal yang didapatkan pada validasi internal dan stabilitas adalah sebanyak 2 kelompok.

Tabel 1, didapatkan hasil pengelompokan untuk data “iris” ini mempunyai 3 variabel *dependent* yang dapat digunakan dalam uji analisis diskriminan sebagai validasinya yaitu variabel hirarki, *kmeans*, dan *species*. Variabel *species* adalah variabel yang dimiliki oleh data “iris” pada awalnya dengan pengelompokan keanggotaannya telah diketahui, memiliki 150 objek data dan memiliki 3 kelompok yaitu setosa, versi color dan virginica. Berdasarkan hal tersebut, pada validasi internal dan stabilitas yang telah digunakan sebelumnya mendapatkan hasil jumlah kelompok optimalnya sebanyak 2 kelompok, sehingga pada validasi *cluster* dengan menggunakan uji analisis diskriminan pengelompokan yang dilakukan harus membentuk kelompok optimal sebanyak 3 kelompok sesuai dengan jumlah kelompok pada variabel *species*. Pengelompokan pada tautan *single linkage* digunakan untuk membuat model diskriminannya, karena merupakan metode pengelompokan yang optimal untuk semua metode validasi. Analisis diskriminan linier membentuk 2 model persamaan untuk membedakan 3 kelompok berdasarkan banyaknya objek 150 dan banyaknya kelompok adalah 3 yaitu kelompok setosa, versicolor dan virginica. Nilai peluang awal (*prior probabilities*) pengklasifikasian objek ke dalam 3 kelompok tautan *single linkage* adalah sama yaitu sebesar 0,33; 0,65; dan 0,013. Nilai peluang tersebut diperoleh dari perbandingan jumlah sampel total dari ketiga kelompok. Berdasarkan dugaan koefisien yang dapat dilihat pada lampiran D.1, didapatkan 2 bentuk model diskriminan yang dinotasikan sebagai D_1 dan D_2 adalah sebagai berikut:

$$D_1 = 0.4801072X_1 + 1.6957809X_2 \\ - 1.6804854X_3 \\ - 0.3374967X_4$$

$$D_2 = -0.0217617X_1 - 2.5309705X_2 \\ - 1.5715342X_3 \\ + 2.8343891X_4$$

Tabel 2 menunjukkan banyaknya objek pada data “iris” yang terklasifikasikan secara tepat oleh model diskriminan linier variabel *species* dengan pengelompokan menggunakan tautan *single linkage* adalah 150 objek (*hit ratio* = 100%). Pada validasi menggunakan uji analisis diskriminan data “iris”, didapatkan untuk semua tautan metode hirarki, *kmeans* dan variabel *species* pengelompokannya adalah optimal karena nilai pada *hit ratio* lebih besar daripada nilai C_{max} dan nilai *Press's Q* yang dihasilkan lebih besar dari nilai titik kritis Chi-Square tabel berderajat bebas 1, sehingga prediksi dalam klasifikasinya dianggap signifikan secara statistik. Namun, pada metode hirarki hanya tautan *single linkage* yang tidak terjadi kesalahan dalam klasifikasi (*missclassification*) dalam proses pengelompokannya. Tautan *complete linkage*, *average linkage*, metode *kmeans* dan variabel *species*, pengelompokannya terjadi kesalahan klasifikasi (ada beberapa data dari total jumlah data yang salah memasuki sebuah kelompok) yang terlihat dari hasil uji analisis diskriminannya. Berdasarkan hal tersebut, pengelompokan metode hirarki dengan menggunakan tautan *single linkage* adalah pengelompokan yang optimal dan paling baik untuk seluruh metode validasi yang digunakan pada data “iris” ini, serta untuk semua variabel yang digunakan dalam pengelompokan ini tidak ada yang hasil pengelompokan keanggotaannya sesuai atau sama dengan pengelompokan keanggotaan yang asli pada variabel *species*, tetapi hanya pada metode hirarki tautan *average linkage* hasil pengelompokan keanggotaannya hampir sesuai dengan pengelompokan keanggotaan daripada variabel *species* yang asli.

Validasi Cluster Data Kemiskinan Kecamatan di Kabupaten Jember

Tabel 3 menjelaskan hasil validasi *cluster* pada data Kecamatan Kabupaten Jember berdasarkan status kemiskinannya menghasilkan jumlah kelompok optimal sebanyak 3 kelompok. Pengelompokan dengan

menggunakan metode hirarki tautan *single linkage* dan *complete linkage* merupakan pengelompokan dengan hasil yang optimal pada validasi internal karena terlihat dari nilai *connectivity* dan nilai *dunn index*-nya yang lebih kecil daripada nilai *connectivity* untuk tautan metode hirarki yang lainnya dan metode *kmeans* serta angka pada nilai *silhouette* yang

mendekati 1. Pada validasi stabilitas, metode hirarki dengan tautannya *complete linkage* dan *average linkage* menghasilkan pengelompokan yang baik (stabil), karena nilai APN dan FOM nya menghasilkan nilai yang lebih kecil dari yang lain.

Tabel 1. Perbandingan Hasil Validasi Keseluruhan Data “iris

Metode	Validasi								
	Internal			Stabilitas		Analisis Diskriminan			
	<i>Connectivity</i>	<i>Dunn Index</i>	<i>Silhouette</i>	<i>APN</i>	<i>FOM</i>	<i>Hit Ratio (%)</i>	<i>Cmax (%)</i>	<i>Press's Q</i>	
Hirarki	<i>Single linkage</i>	0	0,435	0,95	$3,2 \times 10^{-3}$	$4,6 \times 10^{-1}$	100*)	65,33*)	300*)
	\sum kelompok	2	2	2	2	8	3	3	3
	<i>Complete linkage</i>	0	0,435	0,95	$3,2 \times 10^{-3}$	$4,15 \times 10^{-1}$	97,33*)	45,33*)	276,48*)
	\sum kelompok	2	2	2	2	7	3	3	3
	<i>Average linkage</i>	0	0,435	0,95	0	$4,16 \times 10^{-1}$	91,33*)	34*)	227,07*)
	\sum kelompok	2	2	2	2	8	3	3	3
Kmeans	<i>Mac-Queen</i>	6,15	0,14	0,68	$1,3 \times 10^{-4}$	$3,13 \times 10^{-4}$	98*)	40,67*)	282,27*)
	\sum kelompok	2	4	2	2	8	3	3	3
<i>Species</i>	\sum kelompok	-	-	-	-	-	98*)	33,33*)	282,27*)
		-	-	-	-	-	3	3	3

Tabel 2. Kebaikan model Analisis Diskriminan Untuk Metode Analisis Diskriminan linier Data “iris”

	Kelompok	Klasifikasi model			Banyak Objek
		Setosa	Versicolor	Virginica	
Klasifikasi sebenarnya	Setosa	50	0	0	50
	Versi Color	0	98	0	98
	Virginica	0	0	2	2
Banyak objek		50	98	2	150

Tabel 3. Perbandingan Hasil Validasi Keseluruhan Data Kemiskinan Kecamatan di Kabupaten Jember

Metode	Validasi								
	Internal			Stabilitas		Analisis Diskriminan			
	Connectivity	Dunn Index	Silhouette	APN	FOM	Hit Ratio (%)	Cmax (%)	Press's Q	
Hirarki	Single linkage	9.67	0.12	0.85	7.57×10^{-2}	2.97×10^3	100%*	90.32%*	62*
	Complete linkage	9.67	0.12	0.85	2.28×10^{-1}	2.85×10^3	100%*	51.61%*	62*
	Average linkage	11.38	0.03	0.81	4.32×10^{-2}	2.92×10^3	100%*	51.61%*	62*
Kmeans	Mac-Queen	14.78	0.15	0.46	1.31×10^{-1}	3×10^3	90.32%*	45.16%*	45.3*

Tabel 4. Ketepatan Klasifikasi Masing-masing kelompok Pada Metode *Single Linkage* Data Kemiskinan Kecamatan Kabupaten Jember

	Kelompok	Klasifikasi model			Banyak Objek
		I	II	III	
Klasifikasi sebenarnya	I	28	0	0	28
	II	0	1	0	1
	III	0	0	2	2
Banyak objek		28	1	2	31

Tabel 5. Ketepatan Klasifikasi Masing-masing kelompok Pada Metode *Complete Linkage* Data Kemiskinan Kecamatan Kabupaten Jember

	Kelompok	Klasifikasi model			Banyak Objek
		I	II	III	
Klasifikasi sebenarnya	I	16	0	0	16
	II	0	12	0	12
	III	0	0	3	3
Banyak objek		16	12	3	31

Tautan *complete linkage* dapat digunakan sebagai tautan yang memenuhi kriteria pengelompokan yang optimal untuk semua metode validasi yang digunakan pada pembuatan model analisis diskriminannya. Nilai peluang awal (prior probabilities) pengklasifikasian objek kedalam kelompok tersebut yaitu sebesar 0,52; 0,38 dan 0,09. Nilai peluang awal tersebut diperoleh dari perbandingan jumlah sampel total dari ketiga kelompok. Berdasarkan dugaan koefisien, didapatkan 2 bentuk model diskriminan yang dinotasikan sebagai D1 dan D2 adalah sebagai berikut:

$$D_1 = 0.0016762628X_1 - 0.0112894866X_2 + 0.4273345188X_3 - 0.0048378699X_4 + 0.0001456018X_5 + 0.0010246224X_6 + 0.0001842413X_7$$

$$D_2 = 1.486650e - 03X_1 - 2.830492e - 02X_2 + 8.145533e - 01X_3 + 9.294488e - 0X_4 - 3.343122e - 04X_5 + 1.303436e - 03X_6 + 8.107912e - 06X_7$$

Berdasarkan penjelasan analisis diskriminan sebelumnya pada data "iris", selanjutnya dapat kita aplikasikan pada data

Kecamatan Kabupaten Jember berdasarkan peubah status kemiskinannya. Analisis diskriminan linier merupakan analisis yang digunakan untuk mengklasifikasikan objek dari data kemiskinan kecamatan ini ke dalam kelompok kecamatan dengan status ekonomi yang menengah ke bawah. Tautan *single linkage* dapat kita gunakan pada data kemiskinan Kecamatan Kabupaten Jember sebagai tautan yang digunakan pada data “iris” sebelumnya yang dijadikan acuan. Nilai peluang awal (*prior probabilities*) pengklasifikasian objek kedalam kelompok tersebut yaitu sebesar 0,90; 0,03 dan 0,06. Nilai peluang awal tersebut diperoleh dari perbandingan jumlah sampel total dari ketiga kelompok. Berdasarkan dugaan koefisien, didapatkan 2 bentuk model diskriminan yang dinotasikan sebagai D_1 dan D_2 adalah sebagai berikut:

$$D_1 = 0.0 X_1 - 0.0 X_2 + 0.4 X_3 - 0.0 X_4 + 0.0 X_5 + 0.0 X_6 + 0.0 X_7.$$

$$D_2 = 1.486650e - 03X_1 - 2.830492e - 02X_2 + 8.145533e - 01X_3 + 9.294488e - 04X_4 - 3.343122e - 04X_5 + 1.303436e - 03X_6 + 8.107912e - 06X_7$$

Kebaikan model diskriminan untuk metode analisis diskriminan linier dapat dilihat dari ketepatan klasifikasi masing-masing kelompok pada metode *single linkage* dan *complete linkage* yang dapat dilihat pada tabel 4 dan tabel 5. Banyaknya objek pada data Kemiskinan Kecamatan Kabupaten Jember yang terklasifikasikan secara tepat oleh model diskriminan linier dengan pengelompokan menggunakan *single linkage* adalah 31 objek (*hit ratio* = 100%).

Banyaknya objek pada data Kemiskinan Kecamatan Kabupaten Jember yang terklasifikasikan secara tepat oleh model diskriminan linier dengan pengelompokan menggunakan *single linkage* adalah 31 objek (*hit ratio* = 100%). Berdasarkan hal tersebut, pengelompokan metode hirarki dengan

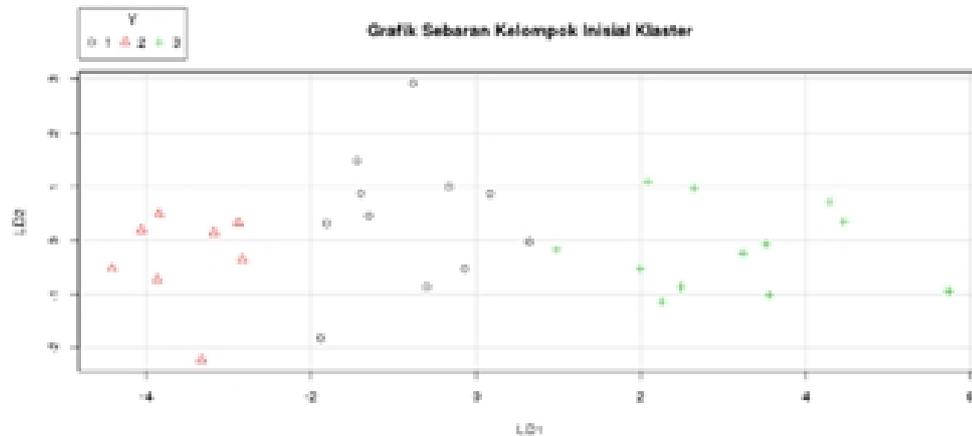
menggunakan tautan *single linkage* yang digunakan dalam data “iris” sebelumnya yang dijadikan acuan juga termasuk dalam metode pengelompokan yang optimal dalam data Kemiskinan Kabupaten Jember, seperti halnya dengan *complete linkage* yang merupakan pengelompokan yang optimal dan paling baik untuk seluruh metode validasi yang digunakan pada data pengelompokan Kecamatan di Kabupaten Jember berdasarkan peubah status kemiskinannya. Tautan *single linkage* dan *complete linkage* dalam data kemiskinan kecamatan ini, dapat disimpulkan keoptimalannya dengan melihat hasil dari *proportion of trace* yang dihasilkan oleh keduanya sehingga diketahui mana yang lebih optimal selain melihat dari pada nilai *hit ratio* nya.

Proportion of trace:	
LD1	LD2
0.9447	0.0553

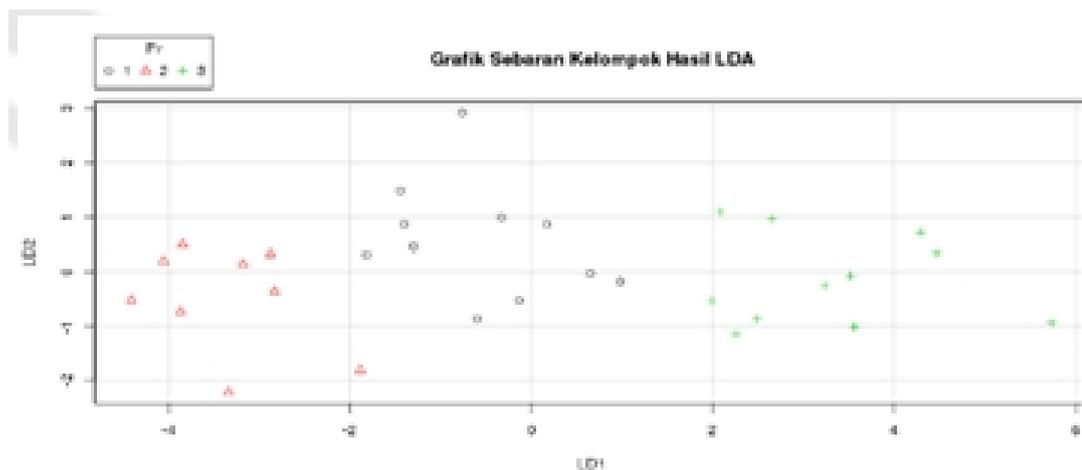
Berdasarkan *proportion of trace* yang dihasilkan dari tautan *single linkage* untuk uji analisis diskriminan sebagai validasi *cluster* nya, menunjukkan bagian LD1 peluangnya sebanyak 94%. Namun, pada *proportion of trace* yang dihasilkan dari tautan *complete linkage* menunjukkan hasil bagian pada LD1 peluangnya lebih banyak dari tautan *single linkage* yaitu sebesar 97%. Sehingga didapatkan bahwa tautan *complete linkage* adalah pengelompokan yang optimal.

Proportion of trace:	
LD1	LD2
0.9769	0.0231

Hasil pengelompokan data kemiskinan Kecamatan di Kabupaten Jember dari metode *Kmeans* menggunakan analisis diskriminan sebagai validasinya adalah adanya beberapa data yang terjadi *missclassification* dengan *hit ratio*=90,32% seperti yang ditunjukkan pada Gambar 2 dan Gambar 3.



Gambar 2. Grafik Sebaran Kelompok Inisial Kluster



Gambar 3. Grafik Sebaran Kelompok Hasil LDA

KESIMPULAN

Penelitian ini mendapatkan bahwa Tautan *Single Linkage* metode hirarki menghasilkan pengelompokan yang optimal untuk semua metode validasi yang digunakan pada data simulasi “iris” sebagai data acuan untuk melakukan pengelompokan data kemiskinan Kecamatan di Kabupaten Jember. Uji analisis diskriminan sebagai salah satu metode untuk memvalidasi hasil solusi dari analisis *cluster* adalah tepat digunakan untuk analisis data pengelompokan secara umum. Validasi *cluster* menggunakan uji analisis diskriminan lebih mudah dilakukan dalam proses perhitungannya dan lebih spesifik pada pengklasifikasian anggota dalam setiap kelompoknya serta dapat juga melihat kesalahan klasifikasi

(*missclassification*) dari nilai APER dan persentase nilai akurasi secara langsung tanpa mempertimbangkan nilai lainnya seperti pada validasi internal dan validasi stabilitas sehingga diketahui kelompok yang dihasilkan tersebut adalah tepat pengklasifikasian keanggotaan pengelompokannya.

DAFTAR PUSTAKA

- Brock, G., Vasyi, P., Susmita, D., dan Somnath, D. 2008. *cValid: An R Package for Cluster Validation*. *Journal of Statistical Software*, **25**(4): 1-22.
- Ernayanti, dan Ita, N. 1996. *Budaya Kemiskinan di Desa Tertinggal di Yogyakarta (Kasus Desa Karang Tengah, Kecamatan Imogiri Kabupaten Bantul)*.

- Direktorat Sejarah dan Nilai Tradisional Bagian Proyek Pengkajian dan Pembinaan Kebudayaan Masa Kini, Jakarta.
- Rachmatin, D dan Sawitri, K. 2014. *Perbandingan Antara Metode Agglomeratif, Metode Divisif, dan Metode K-Means dalam Analisis Klaster*. Seminar Nasional Matematika UNPAR.
- Santosa, S. 2002. *Buku Latihan SPSS Statistik Multivariat*. Jakarta: PT Elex Media Komptindo.
- Sharma, S. (1996). *Applied Multivariate Technique*. Canada: John Wiley & Sons.
- Shidqon, M. 2012. Validasi Metode Kelompok Pada Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Peubah Status Kemiskinan. *Skripsi*. Jember: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.
- Sitepu, R., Irmeilyana, dan Gultom, B. 2011. Analisis *Cluster* terhadap Tingkat Pencemaran Udara Pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. 4(34): 12-17.

