

BAB 14

ANALISIS KLASTER

Sekilas Analisis Klaster

Berikut pemaparan singkat mengenai analisis klaster menurut Hair dkk. (2010:477).

“Cluster analysis groups individuals or objects into clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters. The attempt is to maximize the homogeneity of objects within the clusters while also maximizing the heterogeneity between clusters.”

Malhotra dan Birks (2006:597) menyatakan sebagai berikut.

“Cluster analysis is a class of techniques used to classify objects or cases into relatively homogeneous groups called clusters. Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters. Cluster analysis is also called classification analysis or numerical taxonomy³. Both cluster analysis and discriminant analysis are concerned with classification. Discriminant analysis, however, requires prior knowledge of the cluster or group membership for each object or case included, to develop the classification rule. In contrast, in cluster analysis there is no a priori information about the group or cluster membership for any of the objects. Groups or clusters are suggested by the data, not defined a priori⁵.

Janssens dkk. (2008:317) menyatakan sebagai berikut.

“The objective of cluster analysis is to take a sample of n individuals or objects, each of which is measured for p variables, and group it into g classes, where g is less than n . In other words, the goal is to sort cases (individuals, products, brands, stimuli) into groups so that a high degree of similarity exists between cases in the same group, and a low degree of similarity between cases belonging to different groups. This similarity is evaluated on the basis of the value of each case (individual, product, etc.) for the variables (characteristics, attributes) upon which the cluster analysis is performed.”

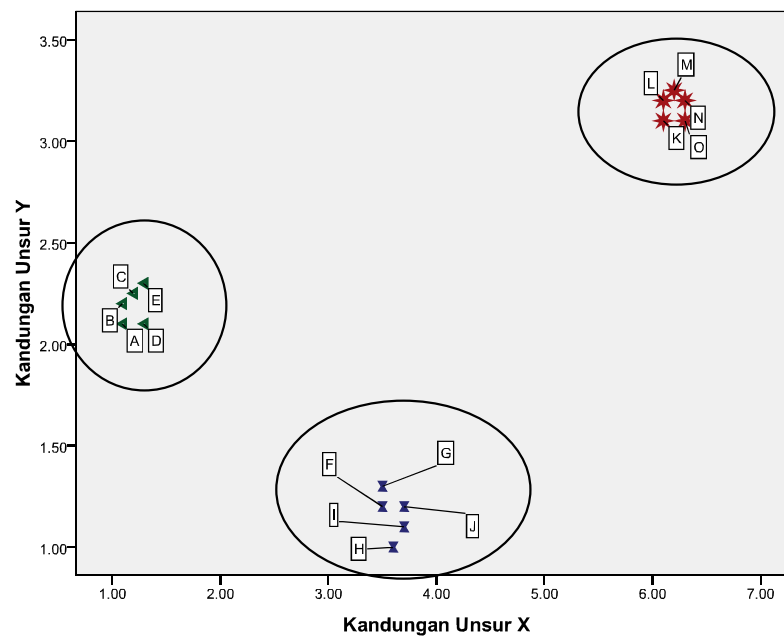
Berdasarkan uraian di atas, analisis klaster (*cluster analysis*) merupakan suatu teknik statistika yang digunakan untuk mengelompokkan (*group*) sekumpulan objek (manusia, produk, tanaman, dan sebagainya) ke dalam beberapa klaster. Perhatikan bahwa suatu objek hanya bisa masuk atau tergabung dalam satu klaster. Beberapa objek yang berada dalam satu klaster cenderung saling mirip, namun cenderung berbeda terhadap objek-objek yang berada dalam klaster lainnya. Sebagai contoh perhatikan data pada Tabel 14.1. Berdasarkan data pada Tabel 14.1, objek yang diteliti adalah batu, sebanyak 15 batu. Masing-masing batu memiliki kadar X dan kadar Y. Gambar 14.1 memberikan gambaran yang cukup jelas untuk pengelompokkan (*cluster*). Berdasarkan Gambar 14.2, jika dibentuk klaster sebanyak 3, maka:

- ⇒ Batu A, B, C, D, dan E berada dalam satu klaster, misalkan klaster pertama.
- ⇒ Batu F, G, H, I, dan J berada dalam satu klaster, misalkan klaster kedua.
- ⇒ Batu K, L, M, N, dan O berada dalam satu klaster, misalkan klaster ketiga.

Perhatikan bahwa batu A, B, C, D, dan E cenderung mirip, karena berada di dalam satu kluster, yakni kluster pertama, namun cenderung berbeda terhadap batu-batu yang berada dalam kluster yang berbeda. Tiga kluster yang tersaji dalam Gambar 14.1 melibatkan dua variabel kluster, yakni variabel **kadar X** (sumbu horizontal) dan **kadar Y** (sumbu vertikal).

Tabel 14.1

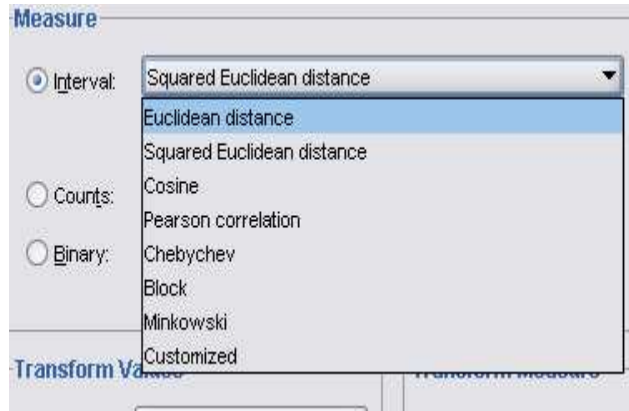
Batu	kadar X	kadar Y
A	1.1	2.1
B	1.1	2.2
C	1.2	2.25
D	1.3	2.1
E	1.3	2.3
F	3.5	1.2
G	3.5	1.3
H	3.6	1
I	3.7	1.1
J	3.7	1.2
K	6.1	3.1
L	6.1	3.2
M	6.2	3.25
N	6.3	3.2
O	6.3	3.1



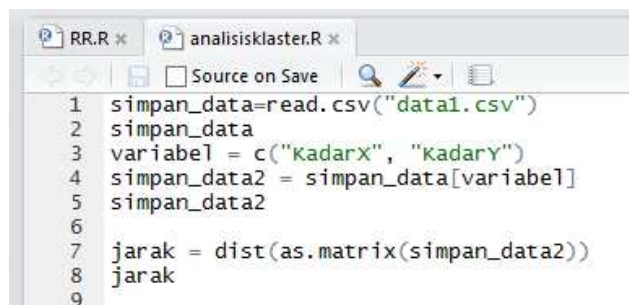
Gambar 14.1

Ukuran Kemiripan (Measure of Similarity)

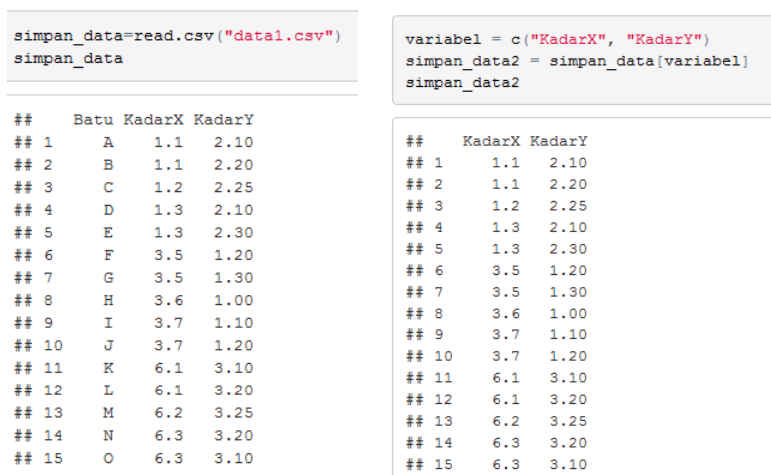
Gambar 14.2 menyajikan beberapa ukuran kemiripan, yakni di antaranya adalah *Euclidean distance* dan *Squared Euclidean distance*. Gambar 14.5 menyajikan *Euclidean distance* (jarak *Euclidean*) untuk tiap-tiap pasang objek (batu). Berdasarkan Gambar 14.1, suatu objek akan semakin mirip dengan objek yang lain, jika posisinya semakin berdekatan. Dengan kata lain, jarak di antara objek tersebut semakin kecil (nilai *Euclidean distance* semakin kecil).



Gambar 14.2 Berbagai Ukuran Kemiripan



Gambar 14.3



Gambar 14.4

```

jarak = dist(as.matrix(simpan_data2))
jarak

##           1           2           3           4           5           6           7
## 2  0.1000000
## 3  0.1802776  0.1118034
## 4  0.2000000  0.2236068  0.1802776
## 5  0.2828427  0.2236068  0.1118034  0.2000000
## 6  2.5632011  2.6000000  2.5283394  2.3769729  2.45967
## 7  2.5298221  2.5632011  2.4884734  2.3409408  2.41660
## 8  2.7313001  2.7730849  2.7060118  2.5495098  2.64114
## 9  2.7856777  2.8231188  2.7518176  2.6000000  2.68328
## 10 2.7513633  2.7856777  2.7115494  2.5632011  2.64007
## 11 5.0990195  5.0803543  4.9731781  4.9030603  4.86621
## 12 5.1195703  5.0990195  4.9912423  4.9244289  4.88364
## 13 5.2280493  5.2069655  5.0990195  5.0331402  4.99124
## 14 5.3150729  5.2952809  5.1877259  5.1195703  5.08035
## 15 5.2952809  5.2773099  5.1703482  5.0990195  5.0635956  3.3837849  3.3286634
##           8           9           10          11          12          13          14
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9  0.1414214
## 10 0.2236068  0.1000000
## 11 3.2649655  3.1240999  3.0610456
## 12 3.3301652  3.1890437  3.1240999  0.1000000
## 13 3.4383862  3.2973474  3.2330326  0.1802776  0.1118034
## 14 3.4828150  3.3421550  3.2802439  0.2236068  0.2000000  0.1118034
## 15 3.4205263  3.2802439  3.2202484  0.2000000  0.2236068  0.1802776  0.1000000

```

Euclidean distance untuk objek A (nomor 1) dan objek O (nomor 15) adalah 5,295.

Euclidean distance untuk objek A (nomor 1) dan objek C (nomor 3) adalah 0,180.

Gambar 14.5 *Euclidean Distance* untuk Tiap-Tiap Pasang Objek (Batu)

Berdasarkan Gambar 14.5, diketahui *Euclidean distance* untuk objek A (nomor 1) dan objek C (nomor 3) adalah 0,180. Nilai tersebut dihitung sebagai berikut.

$$\sqrt{(1,2 - 1,1)^2 + (2,25 - 2,1)^2} = 0,180277 \text{ atau dibulatkan } 0,180.$$

Diketahui *Euclidean distance* untuk objek A (nomor 1) dan objek O (nomor 15) adalah 5,295. Nilai tersebut dihitung sebagai berikut.

$$\sqrt{(6,3 - 1,1)^2 + (3,1 - 2,1)^2} = 5,295280 \text{ atau dibulatkan } 5,295.$$

Gambar 14.7 menyajikan *Squared Euclidean distance* (jarak *Euclidean* yang dikuadratkan) untuk tiap-tiap pasang objek (batu). Berdasarkan Gambar 14.7, diketahui *Squared Euclidean distance* untuk objek A dan objek C adalah 0,032. Nilai tersebut dihitung sebagai berikut.

$$(1,2 - 1,1)^2 + (2,25 - 2,1)^2 = 0,0325.$$

Diketahui *Squared Euclidean distance* untuk objek A dan objek O adalah 28,040. Nilai tersebut dihitung sebagai berikut.

$$(6,3 - 1,1)^2 + (3,1 - 2,1)^2 = 28,04.$$

Diketahui *Squared Euclidean distance* untuk objek C dan objek D adalah 0,0325. Nilai tersebut dihitung sebagai berikut.

$$(1,3 - 1,2)^2 + (2,1 - 2,25)^2 = 0,0325.$$

```

RR.R x analisisklaster.R x
Source on Save
1 simpan_data=read.csv("data1.csv")
2 simpan_data
3 variabel = c("KadarX", "KadarY")
4 simpan_data2 = simpan_data[variabel]
5 simpan_data2
6
7 jarak = dist(as.matrix(simpan_data2)) #euclidean distance
8 jarak
9
10 jarak_pangkat_2=jarak*jarak #squared euclidean distance
11 jarak_pangkat_2
12
13

```

Gambar 14.6

```

jarak_pangkat_2=jarak*jarak
jarak_pangkat_2

##      1      2      3      4      5      6      7      8      9
## 2  0.0100
## 3  0.0325  0.0125
## 4  0.0400  0.0500  0.0325
## 5  0.0800  0.0500  0.0125  0.0400
## 6  6.5700  6.7600  6.3925  5.6500  6.0500
## 7  6.4000  6.5700  6.1925  5.4800  5.8400  0.0100
## 8  7.4600  7.6900  7.3225  6.5000  6.9800  0.0500  0.1000
## 9  7.7600  7.9700  7.5725  6.7600  7.2000  0.0500  0.0800  0.0200
## 10 7.5700  7.7600  7.3525  6.5700  6.9700  0.0400  0.0500  0.0500  0.0100
## 11 26.0000 25.8100 24.7325 24.0400 23.6800 10.3700 10.0000 10.6600  9.7600
## 12 26.2100 26.0000 24.9125 24.2500 23.8500 10.7600 10.3700 11.0900 10.1700
## 13 27.3325 27.1125 26.0000 25.3325 24.9125 11.4925 11.0925 11.8225 10.8725
## 14 28.2500 28.0400 26.9125 26.2100 25.8100 11.8400 11.4500 12.1300 11.1700
## 15 28.0400 27.8500 26.7325 26.0000 25.6400 11.4500 11.0800 11.7000 10.7600
##      10      11      12      13      14
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11  9.3700
## 12  9.7600  0.0100
## 13 10.4525  0.0325  0.0125
## 14 10.7600  0.0500  0.0400  0.0125
## 15 10.3700  0.0400  0.0500  0.0325  0.0100

```

Gambar 14.7 Squared Euclidean Distance untuk Tiap-Tiap Pasang Objek (Batu)

Malhotra dan Birks (2006:600) menyatakan sebagai berikut

“Because the objective of clustering is to group similar objects together, some measure is needed to assess how similar or different the objects are. The most common approach is to measure similarity in terms of distance between pairs of objects. Objects with smaller distances between them are more similar to each other than are those at larger distances. There are several ways to compute the distance between two objects⁹. The most commonly used measure of similarity is the euclidean distance or its square¹⁰. The euclidean distance is the square root of the sum of the squared differences in values for each variable. Other distance measures are also available. The city-block or Manhattan distance between two objects is the sum of the absolute differences in values for each variable. The Chebychev distance between two objects is the maximum absolute difference in values for any variable. For our example, we use the squared euclidean distance.”

Berdasarkan uraian tersebut, secara umum, ukuran kemiripan yang umum digunakan adalah *Euclidean distance* atau *Squared Euclidean distance*. Lebih lanjut Malhotra dan Birks (2006:600) dan Hair dkk. (2010:496-497) menganjurkan untuk melakukan standarisasi data (data ditransormasi ke dalam bentuk normal, dengan rata-rata 0, dan standar deviasi 1) untuk tiap-tiap variabel kluster, apabila data pada variabel-variabel kluster memiliki satuan yang berbeda-beda. Di sisi lain, data yang termasuk *outlier* juga dianjurkan untuk dihapus (Malhotra dan Birks, 2006:601).

Selanjutnya Malhotra dan Birks (2006:601) menyatakan penggunaan ukuran kemiripan (*measure of similarity*) yang berbeda-beda, dapat mempengaruhi hasil kluster, sehingga disarankan untuk menggunakan berbagai ukuran kemiripan dan hasil kluster tersebut diperbandingkan.

Sejalan dengan Malhotra dan Birks, Hair dkk. (2010:496) menyatakan sebagai berikut.

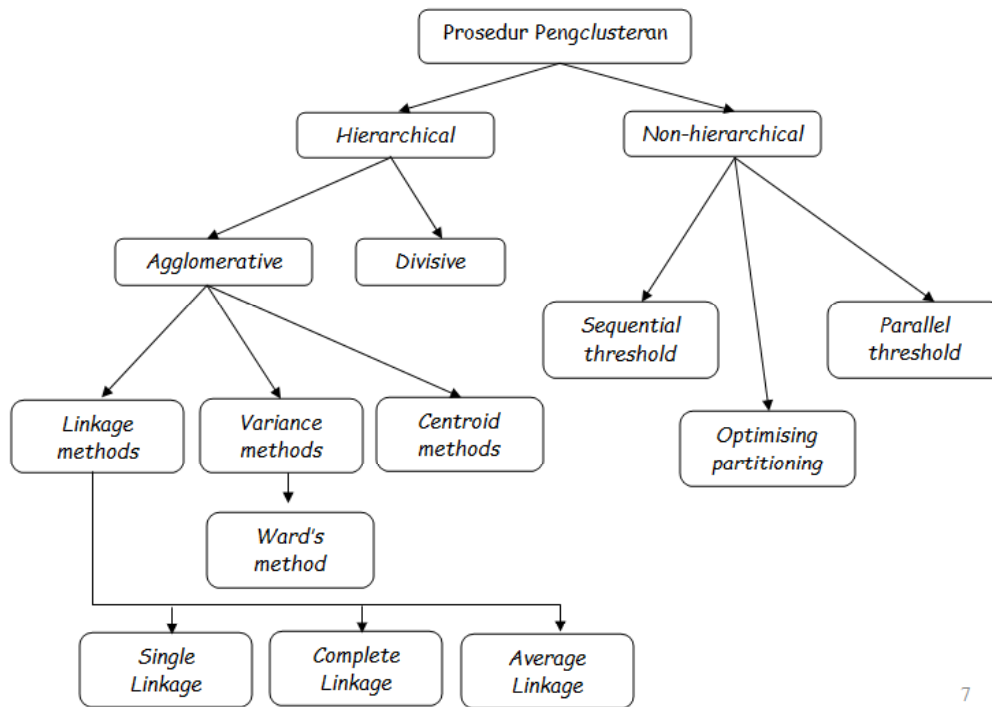
“Which Distance Measures is Best? In attempting to select a particular distance measure, the researcher should remember the following caveats: Difference distance measures or a change in the scales of the variables may lead to different cluster solutions. Thus, it is advisable to use several measures and compare the results with theoretical or know patterns. When the variables are correlated (either positively or negatively) the Mahalanobis distance measure is likely to be the most appropriate because it adjusts for correlations and weights all variable equally. Alternatively, the researcher may wish to avoid using highly redundant variables as input to cluster analysis.”

Prosedur Pengklasteran

Gambar 14.8 menyajikan prosedur pengklasteran dalam analisis kluster (Malhotra dan Birks, 2006:601). Berdasarkan Gambar 14.8, prosedur pengklasteran dapat menggunakan metode *hierarchical* atau metode *non-hierarchical*. Pada metode *hierarchical*, jumlah kluster belum atau tidak diketahui sebelumnya, sementara pada metode *non-hierarchical* jumlah kluster ditetapkan terlebih dahulu, sebelum melakukan pengklasteran objek. Dengan kata lain, pada metode *non-hierarchical*, tahap awal ialah menentukan jumlah kluster yang diinginkan, kemudian tiap-tiap objek pengamatan digabungkan ke dalam salah satu kluster yang telah ditetapkan.

Selanjutnya, dalam metode *hierarchical* terdiri dari dua metode, yakni metode *agglomerative* dan metode *divisive*. Metode *agglomerative* dimulai dengan menganggap tiap-tiap objek sebagai kluster-kluster yang berbeda atau terpisah. Kemudian dua kluster atau objek paling dekat digabung menjadi satu kluster. Proses ini terus berlanjut, sampai seluruh objek bergabung menjadi satu kluster. Sementara pada metode *divisive* merupakan kebalikan dari metode *agglomerative*, yakni dimulai dengan menganggap tiap-tiap objek berasal dalam satu kluster, kemudian dipecah atau dipisahkan sampai setiap objek berada dalam kluster-kluster yang terpisah (Malhotra dan Birks, 2006:601).

Metode *agglomerative* terdiri dari 3 metode, yakni metode *linkage*, *variance*, dan *centroid*. Metode *linkage* terdiri dari metode *single linkage*, *complete linkage*, dan *average linkage*, sementara pada metode *variance* terdiri dari metode *ward*. Pada metode *non-hierarchical* terdiri dari metode *sequential threshold*, *optimising partitioning*, dan *parallel threshold*. Metode *non-hierarchical* sering disebut dengan istilah *k-means clustering*.



Gambar 14.8 Prosedur Pengklasteran (Malhotra dan Birks, 2006:601)

Analisis Klaster dengan Metode Average Linkage

Berikut diberikan contoh penggunaan analisis kluster metode *average linkage*. Diberikan data seperti pada Gambar 14.9. Data pada Gambar 14.9 disajikan dalam grafik seperti pada Gambar 14.10. Berikut akan digunakan analisis kluster metode *average linkage* untuk pengklasteran. Gambar 14.11 menyajikan *Squared Euclidean distance* (matriks jarak/*distance matrix*).

	Batu	A	B
1	A	1.10	1.10
2	B	1.20	.85
3	C	1.30	.97
4	D	1.40	.90
5	E	1.35	1.00
6	F	1.20	1.00

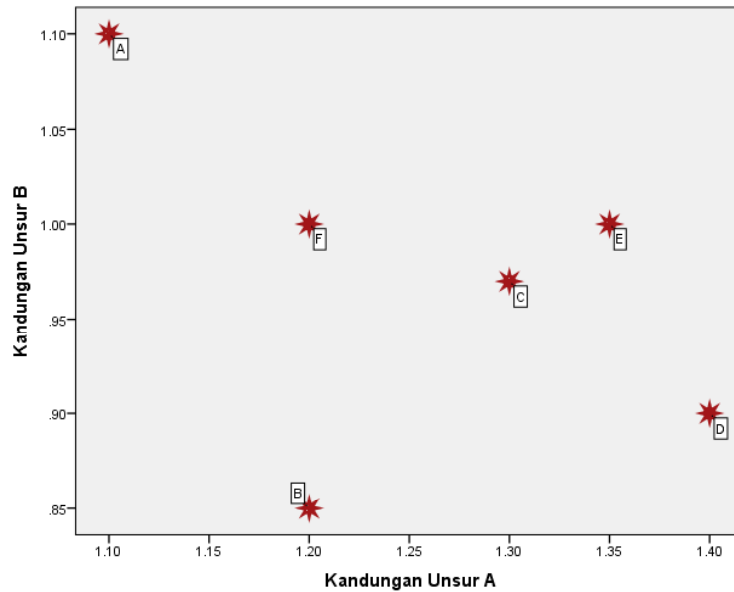
Gambar 14.9

Berdasarkan Gambar 14.11, diketahui *Squared Euclidean distance* untuk objek A dan objek C adalah 0,057. Nilai tersebut dihitung sebagai berikut.

$$(1,1 - 0,97)^2 + (1,1 - 1,3)^2 = 0,0569 \text{ atau dibulatkan } 0,057.$$

Diketahui *Squared Euclidean distance* untuk objek B dan objek F adalah 0,023. Nilai tersebut dihitung sebagai berikut.

$$(0,85 - 1)^2 + (1,2 - 1,2)^2 = 0,0225 \text{ atau dibulatkan } 0,023.$$



Gambar 14.10

Proximity Matrix

Case	Squared Euclidean Distance					
	1:A	2:B	3:C	4:D	5:E	6:F
1:A	.000	.073	.057	.130	.073	.020
2:B	.073	.000	.024	.042	.045	.023
3:C	.057	.024	.000	.015	.003	.011
4:D	.130	.042	.015	.000	.012	.050
5:E	.073	.045	.003	.012	.000	.023
6:F	.020	.023	.011	.050	.023	.000

This is a dissimilarity matrix

Gambar 14.11 Squared Euclidean Distance (Matriks Jarak)

Berdasarkan Gambar 14.11, diketahui nilai *Squared Euclidean distance* **paling kecil** berada pada pasangan objek C dan objek E (pasangan objek yang berbeda), yakni bernilai 0,003 (Perhatikan Gambar 14.11). **Maka objek C dan objek E bergabung menjadi cluster (C,E)**. Pada Gambar 14.12, terlihat bahwa pada *Stage 1*, objek C (3) dan objek E (5) bergabung menjadi *cluster (C,E)*. Perhatikan juga bahwa **nilai coefficient 0,003**, yang merupakan jarak antara objek C dan objek E.

Average Linkage (Between Groups)

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	.003	0	0	2
2	3	4	.014	1	0	4
3	1	6	.020	0	0	5
4	2	3	.037	0	2	5
5	1	2	.055	3	4	0

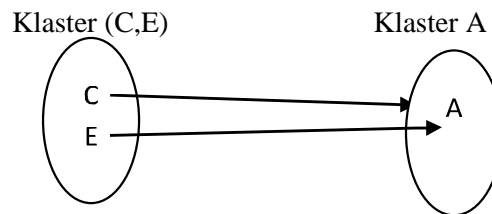
Gambar 14.12 Output SPSS untuk Analisis Kluster Metode Average Linkage

Selanjutnya menghitung jarak antara *cluster (C,E)* terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E) terhadap objek A.

$$d_{(C,E)A} = \frac{d_{(C,A)} + d_{(E,A)}}{N_{(C,E)} \times N_{(A)}} = \frac{0,0569 + 0,0725}{2 \times 1} = 0,0647.$$

Perhatikan bahwa $N_{(C,E)}$ dan N_A masing-masing menyatakan jumlah objek dalam kluster (C,E) dan A.



Menentukan jarak antara kluster (C,E) dan kluster A adalah hitung jarak dari C ke A, dan jarak dari E ke A. Kemudian jumlahkan dan bagi 2. 2 dalam hal ini $2 \times 1 = 2$. 2 menyatakan jumlah anggota kluster (C,E) dan 1 menyatakan jumlah anggota kluster A.

⇒ Menghitung jarak antara *cluster* (C,E) terhadap objek B.

$$d_{(C,E)B} = \frac{d_{(C,B)} + d_{(E,B)}}{N_{(C,E)} \times N_{(B)}} = \frac{0,0244 + 0,045}{2 \times 1} = 0,0347.$$

Gambar 14.13 menyajikan jarak antara *cluster* (C,E) terhadap masing-masing objek.

	A	B	C	D	E	F
A	0.000000	0.072500	0.056900	0.130000	0.072500	0.020000
B	0.072500	0.000000	0.024400	0.042500	0.045000	0.022500
C	0.056900	0.024400	0.000000	0.014900	0.003400	0.010900
D	0.130000	0.042500	0.014900	0.000000	0.012500	0.050000
E	0.072500	0.045000	0.003400	0.012500	0.000000	0.022500
F	0.020000	0.022500	0.010900	0.050000	0.022500	0.000000
Jarak	0.0647	0.0347		0.0137		0.0167

Gambar 14.13

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.14.

	C,E	A	B	D	F
C,E	0	0.0647	0.0347	0.0137	0.0167
A	0.0647	0	0.072500	0.130000	0.020000
B	0.0347	0.072500	0	0.042500	0.022500
D	0.0137	0.130000	0.042500	0	0.050000
F	0.0167	0.020000	0.022500	0.050000	0

Gambar 14.14 Matriks Jarak

Berdasarkan Gambar 14.14, diketahui nilai **jarak paling kecil** berada pada pasangan (C,E) dan D (pasangan objek yang berbeda), yakni bernilai 0,0137, **maka (C,E) dan D bergabung menjadi *cluster* (C,E,D)**. Pada Gambar 14.12, terlihat bahwa pada *Stage 2*, objek C (3) dan

objek D (4) bergabung. Perhatikan juga bahwa **nilai coefficient 0,014 (pembulatan dari 0,0137)**.

Selanjutnya menghitung jarak antara *cluster* (C,E,D) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E,D) terhadap objek A.

$$d_{(C,E,D)A} = \frac{d_{(C,A)} + d_{(E,A)} + d_{(D,A)}}{N_{(C,E,D)} \times N_{(A)}} = \frac{0,0647 + 0,0647 + 0,13}{3 \times 1} = 0,086467.$$

⇒ Menghitung jarak antara *cluster* (C,E,D) terhadap objek B.

$$d_{(C,E,D)B} = \frac{d_{(C,B)} + d_{(E,B)} + d_{(D,B)}}{N_{(C,E,D)} \times N_{(B)}} = \frac{0,0347 + 0,0347 + 0,0425}{3 \times 1} = 0,0373.$$

⇒ Menghitung jarak antara *cluster* (C,E,D) terhadap objek F.

$$d_{(C,E,D)F} = \frac{d_{(C,F)} + d_{(E,F)} + d_{(D,F)}}{N_{(C,E,D)} \times N_{(F)}} = \frac{0,0167 + 0,0167 + 0,05}{3 \times 1} = 0,0278.$$

Gambar 14.15 menyajikan jarak antara *cluster* (C,E,D) terhadap masing-masing objek.

	C,E	A	B	D	F
C,E	0	0.0647	0.0347	0.0137	0.0167
A	0.0647	0	0.0725	0.13	0.02
B	0.0347	0.0725	0	0.0425	0.0225
D	0.0137	0.13	0.0425	0	0.05
F	0.0167	0.02	0.0225	0.05	0
Jarak		0.086467	0.0373		0.0278

Gambar 14.15

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.16.

	C,E,D	A	B	F
C,E,D	0	0.086467	0.0373	0.0278
A	0.086467	0	0.0725	0.02
B	0.0373	0.0725	0	0.0225
F	0.0278	0.02	0.0225	0

Gambar 14.16

Berdasarkan Gambar 14.16, diketahui nilai **jarak paling kecil** berada pada pasangan objek A dan objek F, yakni bernilai 0,02, **maka objek A dan objek F bergabung menjadi cluster (A,F)**. Pada Gambar 14.12, terlihat bahwa pada *Stage 3*, objek 1 (A) dan objek 6 (F). Perhatikan juga bahwa **nilai coefficient 0,02**.

Selanjutnya menghitung jarak antara *cluster* (A,F) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (A,F) terhadap objek B.

$$d_{(A,F)B} = \frac{d_{(A,B)} + d_{(F,B)}}{N_{(A,F)} \times N_{(B)}} = \frac{0,0725 + 0,0225}{2 \times 1} = 0,0475.$$

⇒ Menghitung jarak antara *cluster* (A,F) terhadap *cluster* (C,E,D).

$$d_{(A,F)(C,E,D)} = \frac{d_{(A,C)} + d_{(A,E)} + d_{(A,D)} + d_{(F,C)} + d_{(F,E)} + d_{(F,D)}}{6}$$

$$d_{(A,F)(C,E,D)} = \frac{(3 \times 0,086467) + (3 \times 0,0278)}{6} = 0,057133.$$

Gambar 14.17 menyajikan jarak antara *cluster* (A,F) terhadap masing-masing objek.

	C,E,D	A	B	F
C,E,D	0	0.086467	0.0373	0.0278
A	0.086467	0	0.0725	0.02
B	0.0373	0.0725	0	0.0225
F	0.0278	0.02	0.0225	0
Jarak	0.057133		0.0475	

Gambar 14.17

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.18.

	A,F	C,E,D	B
A,F	0	0.057133	0.0475
C,E,D	0.057133	0	0.0373
B	0.0475	0.0373	0

Gambar 14.18

Berdasarkan Gambar 14.18, diketahui nilai **jarak paling kecil** berada pada pasangan (C,E,D) dan B, yakni bernilai 0,0373, **maka (C,E,D) dan B bergabung menjadi cluster (C,E,D,B)**. Pada Gambar 14.12, terlihat bahwa pada *Stage 4*, objek 2 dan objek 3. Perhatikan juga bahwa **nilai coefficient 0,037**.

Selanjutnya menghitung jarak antara *cluster* (C,E,D,B) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E,D,B) terhadap *cluster* (A,F).

$$d_{(C,E,D,B)(A,F)} = \frac{d_{(A,C)} + d_{(A,E)} + d_{(A,D)} + d_{(A,B)} + d_{(F,C)} + d_{(F,E)} + d_{(F,D)} + d_{(F,B)}}{N_{(C,E,D,B)} \times N_{(A,F)}}$$

$$d_{(C,E,D,B)(A,F)} = \frac{(6 \times 0,057133) + (2 \times 0,0475)}{8} = 0,054725.$$

Gambar 14.19 menyajikan jarak antara *cluster* (A,F) terhadap *cluster* (A,F).

	A,F	C,E,D	B
A,F	0	0.057133	0.0475
C,E,D	0.057133	0	0.0373
B	0.0475	0.0373	0
Jarak	0.054725		

Gambar 14.19

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.20.

	A,F	C,E,D,B
A,F	0	0.054725
C,E,D,B	0.054725	0

Gambar 14.20

Gambar 14.20 menyajikan jarak antara *cluster* (C,E,F,D,B) terhadap *cluster* (A,F). Diketahui jarak antara *cluster* (C,E,F,D,B) dan *cluster* (A,F) adalah 0,054725. **Pada Gambar 14.12**, yakni *Stage 5 (objek 1 dan objek 2 bergabung)*. Diketahui **nilai coefficient** adalah 0,054725. Berdasarkan hasil perhitungan diketahui:

- ⇒ Berdasarkan Gambar 14.20, jika dibentuk dua kluster, maka kluster-kluster tersebut adalah {A,F} dan {C,E,D,B}.
- ⇒ Berdasarkan Gambar 14.18, jika dibentuk tiga kluster, maka kluster-kluster tersebut adalah {A,F}, {C,E,D}, dan {B}.

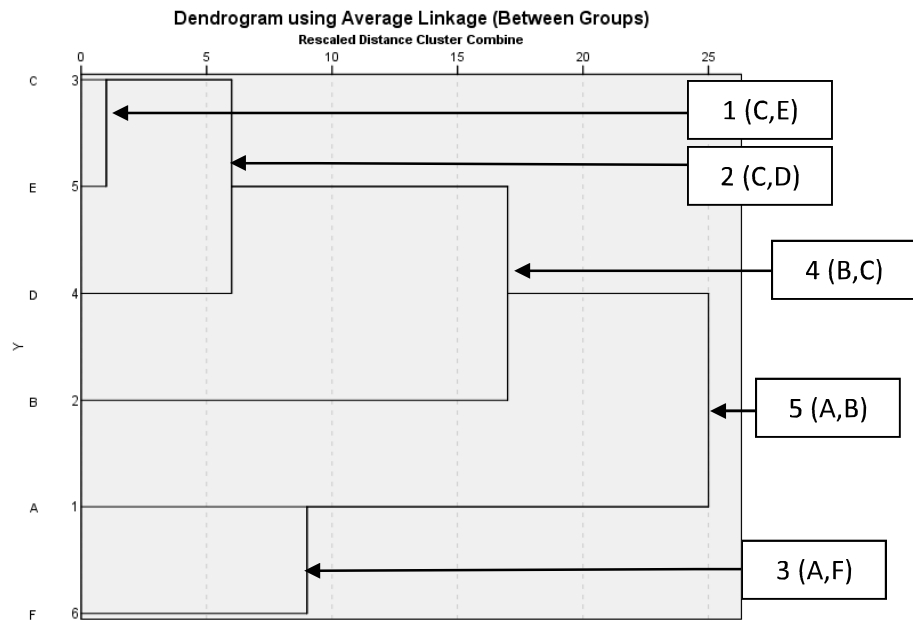
Hasil tersebut sesuai dengan hasil SPSS seperti pada Gambar 14.21. Gambar 14.21 merupakan hasil berdasarkan SPSS untuk analisis kluster metode *average linkage*.

Case	3 Clusters	2 Clusters
1:A	1	1
2:B	2	2
3:C	3	2
4:D	3	2
5:E	3	2
6:F	1	1

Gambar 14.21

Jadi, pada metode *average linkage* memperlakukan jarak di antara dua kluster sebagai jarak rata-rata antara seluruh objek dalam kluster pertama terhadap seluruh objek dalam kluster kedua. Gambar 14.22 menyajikan dendogram. Dendogram menyajikan proses pengklasteran mulai dari *Stage 1* hingga *Stage 6*. Gambar 14.23 disajikan kode R. Sedangkan Gambar 14.24 hingga Gambar 14.26 merupakan hasil eksekusi dari kode R pada Gambar 14.23.

- ⇒ Berdasarkan Gambar 14.25, jika dibentuk dua kluster, maka kluster-kluster tersebut adalah {1A,6F} dan {3C,5E,4D,2B}.
- ⇒ Berdasarkan Gambar 14.18, jika dibentuk tiga kluster, maka kluster-kluster tersebut adalah {A,F}, {C,E,D}, dan {B}.



Gambar 14.22

```

1  simpan_data=read.csv("data2.csv")
2  simpan_data
3  variabel = c("A", "B")
4  simpan_data2 = simpan_data[variabel]
5  simpan_data2
6
7  jarak = dist(as.matrix(simpan_data2)) #euclidean distance
8  jarak
9
10 jarak_pangkat_2=jarak*jarak #squared euclidean distance
11 jarak_pangkat_2
12
13 klaster1 = hclust(jarak_pangkat_2, method="average")
14 plot(klaster1) # display dendrogram
15 dua_klaster = cutree(klaster1, k=2) # cut tree into 5 clusters
16 rect.hclust(klaster1, k=2, border="red")
17
18 klaster2 = hclust(jarak_pangkat_2, method="average")
19 plot(klaster2) # display dendrogram
20 tiga_klaster = cutree(klaster2, k=3) # cut tree into 5 clusters
21 rect.hclust(klaster2, k=3, border="red")
22

```

Gambar 14.23

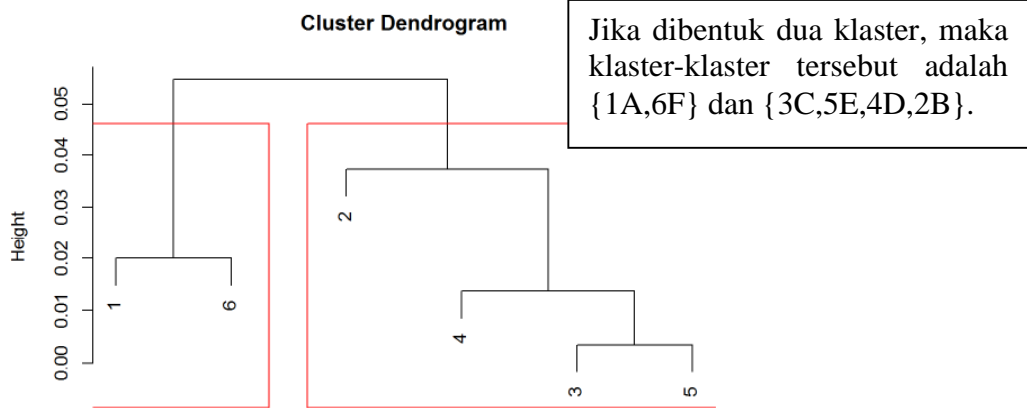
<pre> simpan_data=read.csv("data2.csv") simpan_data </pre>	<pre> jarak = dist(as.matrix(simpan_data2)) #euclidean distance jarak </pre>
<pre> ## Batu A B ## 1 A 1.10 1.10 ## 2 B 1.20 0.85 ## 3 C 1.30 0.97 ## 4 D 1.40 0.90 ## 5 E 1.35 1.00 ## 6 F 1.20 1.00 </pre>	<pre> ## 1 2 3 4 5 ## 2 0.26925824 ## 3 0.23853721 0.15620499 ## 4 0.36055513 0.20615528 0.12206556 ## 5 0.26925824 0.21213203 0.05830952 0.11180340 ## 6 0.14142136 0.15000000 0.10440307 0.22360680 0.15000000 </pre>
<pre> variabel = c("A", "B") simpan_data2 = simpan_data[variabel] simpan_data2 </pre>	<pre> jarak_pangkat_2=jarak*jarak #squared euclidean distance jarak_pangkat_2 </pre>
<pre> ## A B ## 1 1.10 1.10 ## 2 1.20 0.85 ## 3 1.30 0.97 ## 4 1.40 0.90 ## 5 1.35 1.00 ## 6 1.20 1.00 </pre>	<pre> ## 1 2 3 4 5 ## 2 0.0725 ## 3 0.0569 0.0244 ## 4 0.1300 0.0425 0.0149 ## 5 0.0725 0.0450 0.0034 0.0125 ## 6 0.0200 0.0225 0.0109 0.0500 0.0225 </pre>

Gambar 14.24

```

klaster1 = hclust(jarak_pangkat_2, method="average")
plot(klaster1) # display dendrogram
dua_klaster = cutree(klaster1, k=2) # cut tree into 5 clusters
rect.hclust(klaster1, k=2, border="red")

```

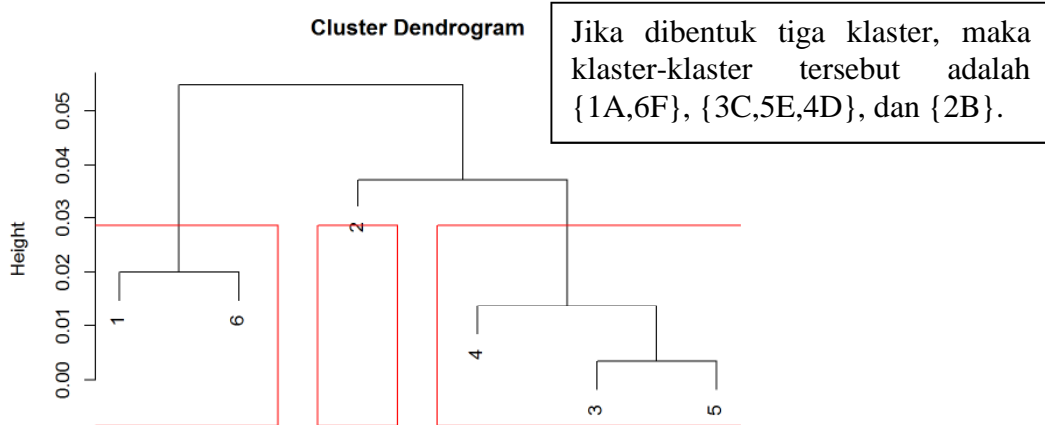


Gambar 14.25

```

klaster2 = hclust(jarak_pangkat_2, method="average")
plot(klaster2) # display dendrogram
tiga_klaster = cutree(klaster2, k=3) # cut tree into 5 clusters
rect.hclust(klaster2, k=3, border="red")

```



Gambar 14.26

Analisis Klaster dengan Metode Single Linkage

Berdasarkan Gambar 14.11, diketahui nilai *Squared Euclidean distance* **paling kecil** berada pada pasangan objek C dan objek E (pasangan objek yang berbeda), yakni bernilai 0,003 (Perhatikan Gambar 14.27). **Maka objek C dan objek E bergabung menjadi cluster (C,E).** Pada Gambar 14.28, terlihat bahwa pada *Stage 1*, objek C dan objek E bergabung menjadi *cluster* (C,E). Perhatikan juga bahwa **nilai coefficient 0,003**, yang merupakan jarak antara objek C dan objek E.

	A	B	C	D	E	F
A	0	0.073	0.057	0.13	0.073	0.02
B	0.073	0	0.024	0.042	0.045	0.023
C	0.057	0.024	0	0.015	0.003	0.011
D	0.13	0.042	0.015	0	0.012	0.05
E	0.073	0.045	0.003	0.012	0	0.023
F	0.02	0.023	0.011	0.05	0.023	0

Gambar 14.27 Squared Euclidean Distance (Matriks Jarak)

Single Linkage

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	5	.003	0	0	2
2	3	6	.011	1	0	3
3	3	4	.012	2	0	4
4	1	3	.020	0	3	5
5	1	2	.023	4	0	0

3 dalam hal ini adalah objek C, dan 5 dalam hal ini adalah objek E.

Gambar 14.28

Selanjutnya menghitung jarak antara cluster (C,E) terhadap objek lainnya.

⇒ Menghitung jarak antara cluster (C,E) terhadap objek A.

$$d_{(C,E)A} = \min\{d_{(C,A)}; d_{(E,A)}\} = \min\{0,057; 0,073\} = 0,057.$$

Dapat diartikan bahwa jarak antara objek C ke objek A lebih dekat, dibandingkan jarak antara objek E ke objek A.

⇒ Menghitung jarak antara cluster (C,E) terhadap objek B.

$$d_{(C,E)B} = \min\{d_{(C,B)}; d_{(E,B)}\} = \min\{0,024; 0,045\} = 0,024.$$

Dapat diartikan bahwa jarak antara objek C ke objek B lebih dekat, dibandingkan jarak antara objek E ke objek B.

⇒ Menghitung jarak antara cluster (C,E) terhadap objek D.

$$d_{(C,E)D} = \min\{d_{(C,D)}; d_{(E,D)}\} = \min\{0,015; 0,012\} = 0,012.$$

⇒ Menghitung jarak antara cluster (C,E) terhadap objek F.

$$d_{(C,E)F} = \min\{d_{(C,F)}; d_{(E,F)}\} = \min\{0,011; 0,023\} = 0,011.$$

Gambar 14.29 menyajikan jarak antara cluster (C,E) terhadap masing-masing objek.

	A	B	C	D	E	F
A	0	0.073	0.057	0.13	0.073	0.02
B	0.073	0	0.024	0.042	0.045	0.023
C	0.057	0.024	0	0.015	0.003	0.011
D	0.13	0.042	0.015	0	0.012	0.05
E	0.073	0.045	0.003	0.012	0	0.023
F	0.02	0.023	0.011	0.05	0.023	0
Minimum	0.057	0.024		0.012		0.011

Gambar 14.29

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.30.

	C,E	A	B	D	F
C,E	0	0.057	0.024	0.012	0.011
A	0.057	0	0.073	0.13	0.02
B	0.024	0.073	0	0.042	0.023
D	0.012	0.13	0.042	0	0.05
F	0.011	0.02	0.023	0.05	0

Gambar 14.30 Matriks Jarak

Sampai pada tahap ini, telah terbentuk 5 kluster, yakni {C,E}, {A}, {B}, {D}, dan {F}. Gambar 14.31 disajikan *output* SPSS.

Jika dibentuk 5 kluster, maka diperoleh kluster {C,E}, {A}, {B}, {D}, dan {F}.

Cluster Membership					
Case	5 Clusters	4 Clusters	3 Clusters	2 Clusters	
1:A	}	1	1	1	1
2:B		2	2	2	2
3:C		3	3	3	1
4:D		4	4	3	1
5:E		3	3	3	1
6:F	5	3	3	3	1

Gambar 14.31 *Output* SPSS untuk Analisis Kluster Metode *Single Linkage*

Berdasarkan Gambar 14.30, diketahui nilai **jarak paling kecil** berada pada pasangan (C,E) dan F, yakni bernilai 0,011. Maka (C,E) dan F bergabung menjadi *cluster* (C,E,F). Pada Gambar 14.28, yakni *Stage 2* (objek 3 dan objek 6 bergabung). Diketahui nilai *coefficient* adalah 0,011 (lihat juga nilai *coefficient* pada Gambar 14.28, *Stage 2*).

Selanjutnya menghitung jarak antara *cluster* (C,E,F) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E,F) terhadap objek A.

$$d_{(C,E,F)A} = \min\{d_{(C,E)A}; d_{(F,A)}\} = \min\{0,057; 0,02\} = 0,02.$$

⇒ Menghitung jarak antara *cluster* (C,E,F) terhadap objek B.

$$d_{(C,E,F)B} = \min\{d_{(C,E)B}; d_{(F,B)}\} = \min\{0,024; 0,023\} = 0,023.$$

⇒ Menghitung jarak antara *cluster* (C,E,F) terhadap objek D.

$$d_{(C,E,F)D} = \min\{d_{(C,E)D}; d_{(F,D)}\} = \min\{0,012; 0,05\} = 0,012.$$

Gambar 14.32 menyajikan jarak antara *cluster* (C,E,F) terhadap masing-masing objek.

	C,E	A	B	D	F
C,E	0	0.057	0.024	0.012	0.011
A	0.057	0	0.073	0.13	0.02
B	0.024	0.073	0	0.042	0.023
D	0.012	0.13	0.042	0	0.05
F	0.011	0.02	0.023	0.05	0
Minimum		0.02	0.023	0.012	

Gambar 14.32

Sehingga diperoleh matriks matriks jarak yang baru seperti pada Gambar 14.33.

	C,E,F	A	B	D
C,E,F	0	0.02	0.023	0.012
A	0.02	0	0.073	0.13
B	0.023	0.073	0	0.042
D	0.012	0.13	0.042	0

Gambar 14.33

Berdasarkan Gambar 14.33, diketahui nilai **jarak paling kecil** berada pada pasangan (C,E,F) dan D, yakni bernilai 0,012. **Maka (C,E,F) dan D bergabung menjadi *cluster* (C,E,F,D).** Pada Gambar 14.28, yakni *Stage 3* (**objek 3 dan objek 4 bergabung**). Diketahui **nilai *coefficient*** adalah 0,012 (lihat juga nilai *coefficient* pada Gambar 14.28, *Stage 3*).

Selanjutnya menghitung jarak antara *cluster* (C,E,F,D) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E,F,D) terhadap objek A.

$$d_{(C,E,F,D)A} = \min\{d_{(C,E,F)A}; d_{(D,A)}\} = \min\{0,02; 0,13\} = 0,02.$$

⇒ Menghitung jarak antara *cluster* (C,E,F,D) terhadap objek B.

$$d_{(C,E,F,D)B} = \min\{d_{(C,E,F)B}; d_{(D,B)}\} = \min\{0,023; 0,042\} = 0,023.$$

Gambar 14.34 menyajikan jarak antara *cluster* (C,E,F,D) terhadap masing-masing objek.

	C,E,F	A	B	D
C,E,F	0	0.02	0.023	0.012
A	0.02	0	0.073	0.13
B	0.023	0.073	0	0.042
D	0.012	0.13	0.042	0
Minimum		0.02	0.023	

Gambar 14.34

Sehingga diperoleh matriks jarak yang baru seperti pada Gambar 14.35.

	C,E,F,D	A	B
C,E,F,D	0	0.02	0.023
A	0.02	0	0.073
B	0.023	0.073	0

Gambar 14.35

Berdasarkan Gambar 14.35, diketahui **nilai jarak paling kecil** berada pada pasangan (C,E,F,D) dan A, yakni bernilai 0,002. **Maka (C,E,F,D) dan A bergabung menjadi cluster (C,E,F,D,A).** Pada Gambar 14.28, yakni *Stage 4 (objek 3 dan objek 1 bergabung)*. Diketahui **nilai coefficient** adalah 0,02 (lihat juga nilai *coefficient* pada Gambar 14.28, *Stage 4*).

Selanjutnya menghitung jarak antara *cluster* (C,E,F,D,A) terhadap objek lainnya.

⇒ Menghitung jarak antara *cluster* (C,E,F,D,A) terhadap objek B.

$$d_{(C,E,F,D,A)B} = \min\{d_{(C,E,F,D)B}; d_{(A,B)}\} = \min\{0,023; 0,073\} = 0,023.$$

Gambar 14.36 menyajikan jarak antara *cluster* (C,E,F,D,A) terhadap objek B. diketahui jarak antara *cluster* (C,E,F,D,A) dan B adalah 0,023. **Pada Gambar 14.28, yakni Stage 5 (objek 1 dan objek 2 bergabung).** Diketahui **nilai coefficient** adalah 0,023 (lihat juga nilai *coefficient* pada Gambar 14.28, *Stage 5*).

	C,D,E,F,A	B
C,D,E,F,A	0	0.023
B	0.023	0

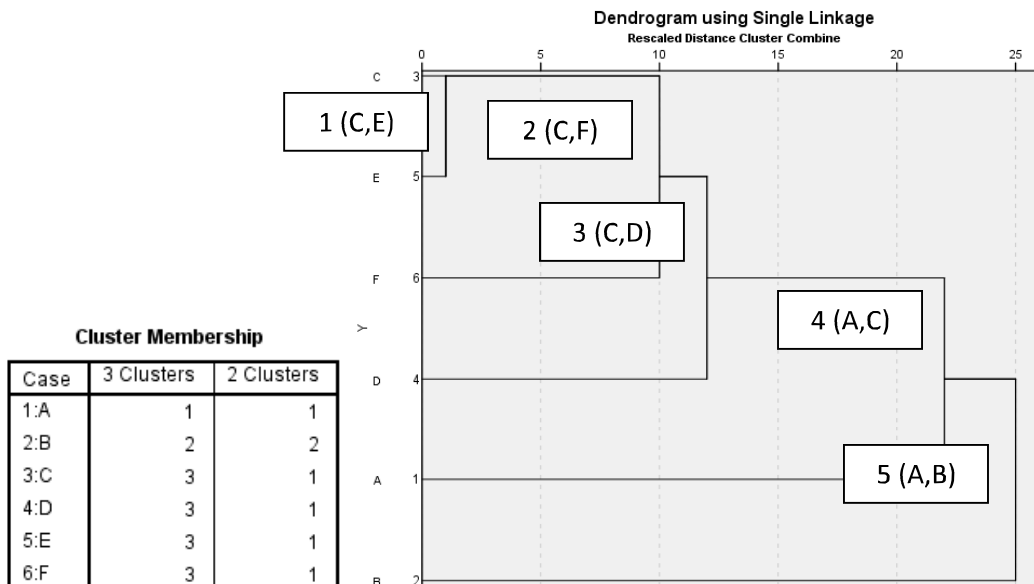
Gambar 14.36

Dari hasil yang telah diperoleh, dapat disimpulkan bahwa:

- ⇒ Jika dibentuk 3 klaster, maka {C,E,F,D} , {A}, dan {B} (lihat Gambar 14.35).
- ⇒ Jika dibentuk 2 klaster, maka {C,E,F,D,A} dan {B} (lihat Gambar 14.36).

Hasil pengklasteran tersebut sesuai dengan hasil yang diperoleh dengan SPSS, seperti pada Gambar 14.37. Berdasarkan Gambar 14.37, jika dibentuk 3 klaster, maka diperoleh klaster {A}, {B}, dan {C,D,E,F}, dan jika dibentuk dua klaster, maka diperoleh klaster {A,C,D,E,F} dan {B}. Gambar 14.34 menyajikan *dendogram* dengan menggunakan metode *single linkage*. Berdasarkan *dendogram* tersebut, dapat ditarik informasi:

- ⇒ Pertama, objek 3 (C) dan objek 5 (E) bergabung menjadi *cluster* (3,5).
- ⇒ Selanjutnya, *cluster* (3,5) bergabung dengan objek 6 (F) membentuk *cluster* (3,5,6).
- ⇒ Kemudian *cluster* (3,5,6) bergabung dengan objek 4 (D) membentuk *cluster* (3,5,6,4).
- ⇒ *Cluster* (3,5,6,4) bergabung dengan objek 1 (A) membentuk *cluster* (3,5,6,4,1).
- ⇒ Dan terakhir *cluster* (3,5,6,4,1) bergabung dengan objek 2 (B) membentuk *cluster* (3,5,6,4,1,2).



Gambar 14.37

Gambar 14.38

Jadi, pada metode *average linkage*, pertama menentukan jarak paling minimum antara dua objek. Misalkan objek i dan objek k memiliki jarak yang paling minimum, maka objek i dan objek k bergabung menjadi suatu kluster (i, k) . Langkah selanjutnya menghitung jarak antara kluster (i, k) terhadap kluster/objek lainnya (misalkan kluster l), dengan rumus sebagai berikut.

$$d_{(i,k),l} = \min(d_{(i,k)}; d_{(i,l)}).$$

Gambar 14.39 disajikan kode R. Sedangkan Gambar 14.40 hingga Gambar 14.42 merupakan hasil eksekusi dari kode R pada Gambar 14.39.

```

1  simpan_data=read.csv("data2.csv")
2  simpan_data
3  variabel = c("A", "B")
4  simpan_data2 = simpan_data[variabel]
5  simpan_data2
6
7  jarak = dist(as.matrix(simpan_data2)) #euclidean distance
8  jarak
9
10 jarak_pangkat_2=jarak*jarak #squared euclidean distance
11 jarak_pangkat_2
12
13 klaster1 = hclust(jarak_pangkat_2, method="single")
14 plot(klaster1) # display dendrogram
15 dua_klaster = cutree(klaster1, k=2) # cut tree into 5 clusters
16 rect.hclust(klaster1, k=2, border="red")
17
18 klaster2 = hclust(jarak_pangkat_2, method="single")
19 plot(klaster2) # display dendrogram
20 tiga_klaster = cutree(klaster2, k=3) # cut tree into 5 clusters
21 rect.hclust(klaster2, k=3, border="red")
22
23 klaster3 = hclust(jarak_pangkat_2, method="single")
24 plot(klaster3) # display dendrogram
25 empat_klaster = cutree(klaster3, k=4) # cut tree into 5 clusters
26 rect.hclust(klaster3, k=4, border="red")
27

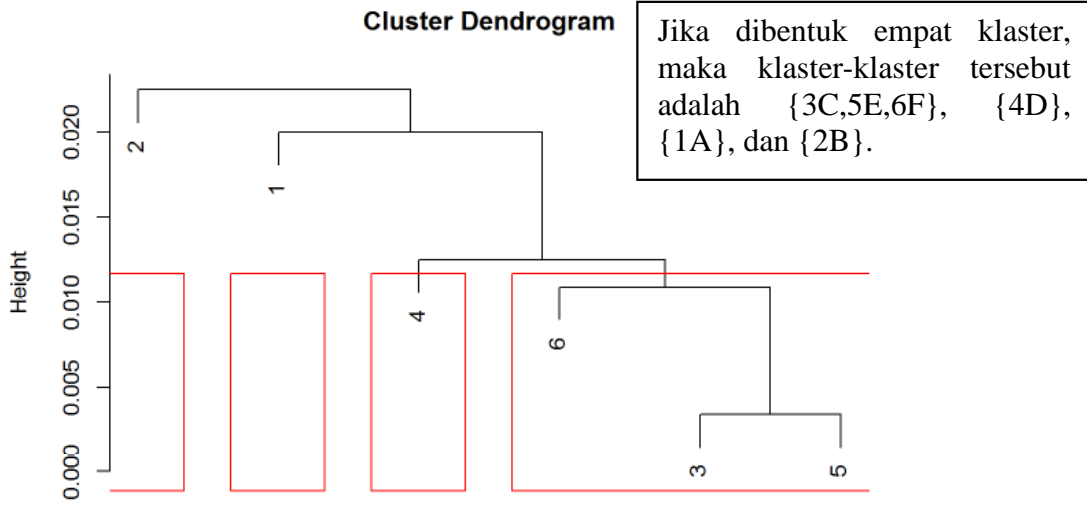
```

Gambar 14.39

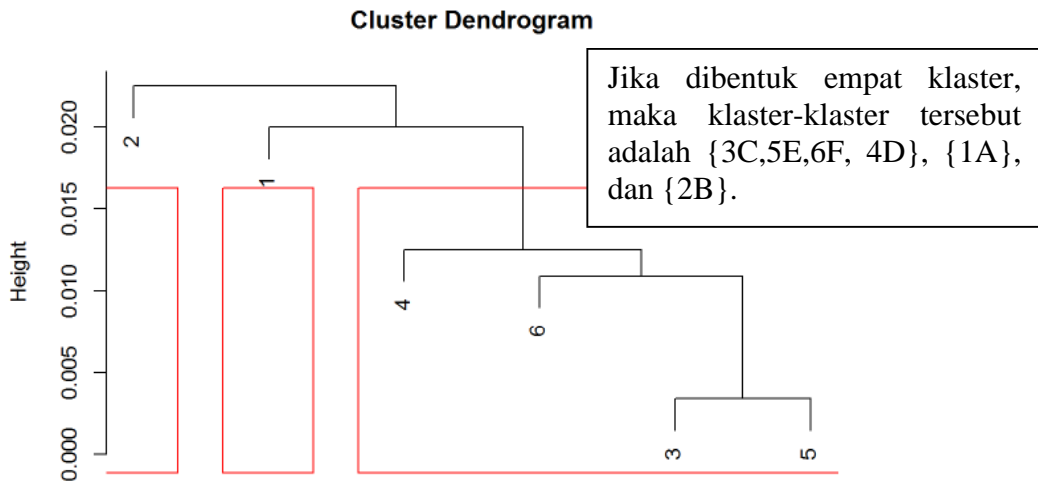
```

klaster3 = hclust(jarak_pangkat_2, method="single")
plot(klaster3) # display dendrogram
empat_klaster = cutree(klaster3, k=4) # cut tree into 5 clusters
rect.hclust(klaster3, k=4, border="red")

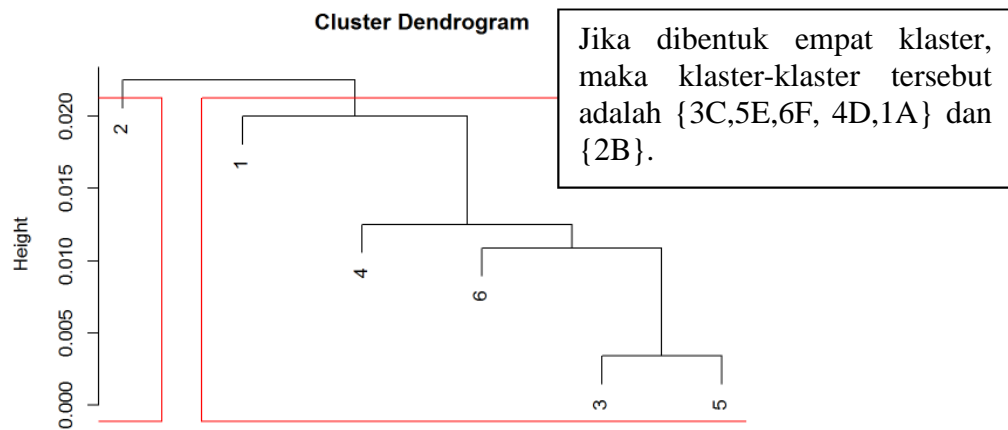
```



Gambar 14.40



Gambar 14.41



Gambar 14.42

Referensi

1. Field, A. 2009. *Discovering Statistics Using SPSS, 3rd Edition*. London: Sage.
2. Gio, P.U. dan E. Rosmaini, 2015. Belajar Olah Data dengan SPSS, Minitab, R, Microsoft Excel, EViews, LISREL, AMOS, dan SmartPLS. USUpres.
3. Hair, J.F Jr., R.E. Anderson, B.J. Babin, dan W.C. Black. 2010. *Multivariate Data Analysis, 7th Edition*. Pearson Prentice Hall.
4. Janssens, W., K. Wijnen, P.D. Pelsmacker, dan P.V. Kenhove. 2008. *Marketing Research with SPSS*. Pearson Prentice Hall.
5. Johnson, R.A. dan D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis, 6th Edition*. Pearson Prentice Hall.
6. Malhotra, N.K. dan D.F. Birks. 2006. *Marketing Research, An Applied Approach, 2nd European Edition*. London: Prentice Hall.
7. Stevens, J.P. 2009. *Applied Multivariate Statistics For The Social Science, 5th Edition*. New York: Routledge.
8. <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/hclust.html>
9. <http://www.r-tutor.com/gpu-computing/clustering/hierarchical-cluster-analysis>
10. <http://cc.oulu.fi/~jarioksa/opetus/metodi/sessio3.pdf>