



# Pengenalan Pola

K-Means Clustering

PTIIK - 2014



# Course Contents

1

Definisi k-Means

2

Algoritma k-Means

3

Studi Kasus

4

Latihan dan Diskusi

# K-Means Clustering

- K-Means merupakan salah satu metode pengelompokan data non hirarki yang digunakan untuk mempartisi  $N$  objek data ke dalam  $K$  kelompok.
- Setiap kelompok data memiliki jarak terdekat dengan centroidnya masing-masing.
- Algoritma K-Means :
  1. Tentukan jumlah cluster/kelompok.
  2. Alokasikan data ke dalam kelompok secara acak.
  3. Hitung pusat cluster dari data yang ada di masing-masing cluster.
  4. Alokasikan masing-masing data ke centroid terdekat.
  5. Kembali ke langkah 3, jika masih ada data yang berpindah cluster atau ada perubahan nilai centroid di atas nilai ambang/Threshold yang ditentukan, atau jika perubahan nilai pada fungsi objektif masih di atas nilai ambang atau dengan batasan iterasi max.
  6. Selesai.

# K-Means Clustering

- Menentukan Centroid (Titik Pusat) setiap kelompok diambil dari nilai rata-rata (Means) semua nilai data pada setiap fiturnya. Jika  $M$  menyatakan jumlah data pada suatu kelompok,  $i$  menyatakan fitur ke- $i$  dalam sebuah kelompok, berikut rumus untuk menghitung centroid :

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

- Hitung jarak titik terdekat (Euclidean Distance):

$$D(x_2, x_1) = \sqrt{\sum_{j=1}^p |x_{2j} - x_{1j}|^2}$$

- Pengalokasian keanggotaan titik :

$$a_{ji} = \begin{cases} 1 & , d = \min(D(x_j, C_i)) \\ 0 & , \text{lainnya} \end{cases}$$

- Fungsi Objektif :

$$F = \sum_{j=1}^N \sum_{i=1}^K a_{ji} D(x_j, C_i)$$



# Studi Kasus

- Perhatikan dataset berikut :

Data	Fitur x	Fitur y	Kelompok 1	Kelompok 2	Kelompok 3
1	1	1	*		
2	4	1		*	
3	6	1		*	
4	1	2	*		
5	2	3			*
6	5	3		*	
7	2	5			*
8	3	5		*	
9	2	6			*
10	3	8		*	

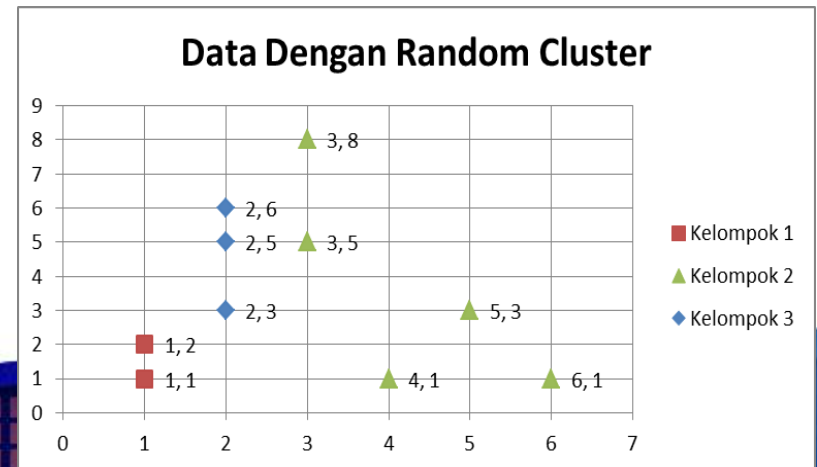
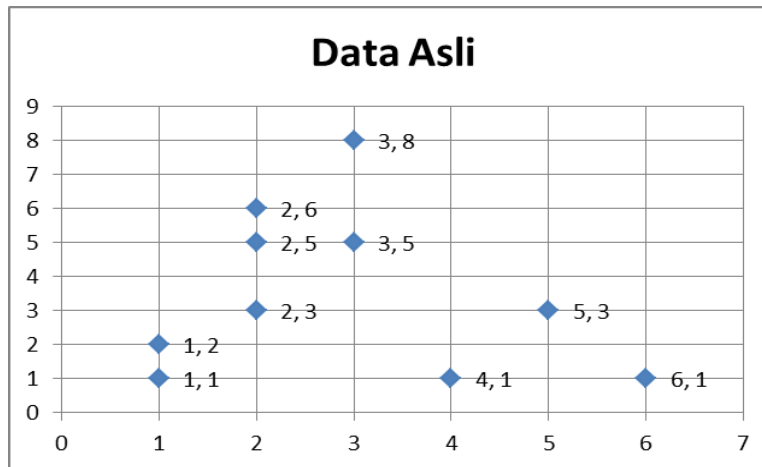
Inisialisasi :

$K = 3$ ,

Fungsi Objektif ( $F$ ) = 0,  
Threshold ( $T$ ) = 0.8, dan  
Data dicluster sebanyak  
 $K$  secara random.

Tentukan Hasil Akhir  
Clusteringnya !

- Bentuk Visualisasi data :







- Menghitung Centroid Setiap Cluster :

Data	F x	F y	K 1	K 2	K 3	K 1 F x	K 1 F y	K 2 F x	K 2 F y	K 3 F x	K 3 F y
1	1	1	*			1	1				
2	4	1		*				4	1		
3	6	1		*				6	1		
4	1	2	*			1	2				
5	2	3			*					2	3
6	5	3		*				5	3		
7	2	5			*					2	5
8	3	5		*				3	5		
9	2	6			*					2	6
10	3	8		*				3	8		
Total			2	5	3	2	3	21	18	6	14

- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = $2 / 2 = 1$	Total K1Fy / Total K1 = $3 / 2 = 1.5$
2	Total K2Fx / Total K2 = $21 / 5 = 4.2$	Total K2Fy / Total K2 = $18 / 5 = 3.6$
3	Total K3Fx / Total K3 = $6 / 3 = 2$	Total K3Fy / Total K3 = $14 / 3 = 4.6667$



- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = 2 / 2 = 1	Total K1Fy / Total K1 = 3 / 2 = 1.5
2	Total K2Fx / Total K2 = 21 / 5 = 4.2	Total K2Fy / Total K2 = 18 / 5 = 3.6
3	Total K3Fx / Total K3 = 6 / 3 = 2	Total K3Fy / Total K3 = 14 / 3 = 4.6667

- Menghitung Jarak Data Ke Centroid :

Data	F x	F y	Jarak Ke C 1	Jarak Ke C 2	Jarak Ke C 3	Min	Kelompok Baru	Kelompok Sebelumnya
1	1	1	0.5000	4.1231	3.8006	0.5000	1	1
2	4	1	3.0414	2.6077	4.1767	2.6077	2	2
3	6	1	5.0249	3.1623	5.4263	3.1623	2	2
4	1	2	0.5000	3.5777	2.8480	0.5000	1	1
5	2	3	1.8028	2.2804	1.6667	1.6667	3	3
6	5	3	4.2720	1.0000	3.4319	1.0000	2	2
7	2	5	3.6401	2.6077	0.3333	0.3333	3	3
8	3	5	4.0311	1.8439	1.0541	1.0541	3	2
9	2	6	4.6098	3.2558	1.3333	1.3333	3	3
10	3	8	6.8007	4.5607	3.4801	3.4801	3	2
Total			1.0000	13.1746	3.3333	(Total berdasarkan kelompok sebelumnya)		

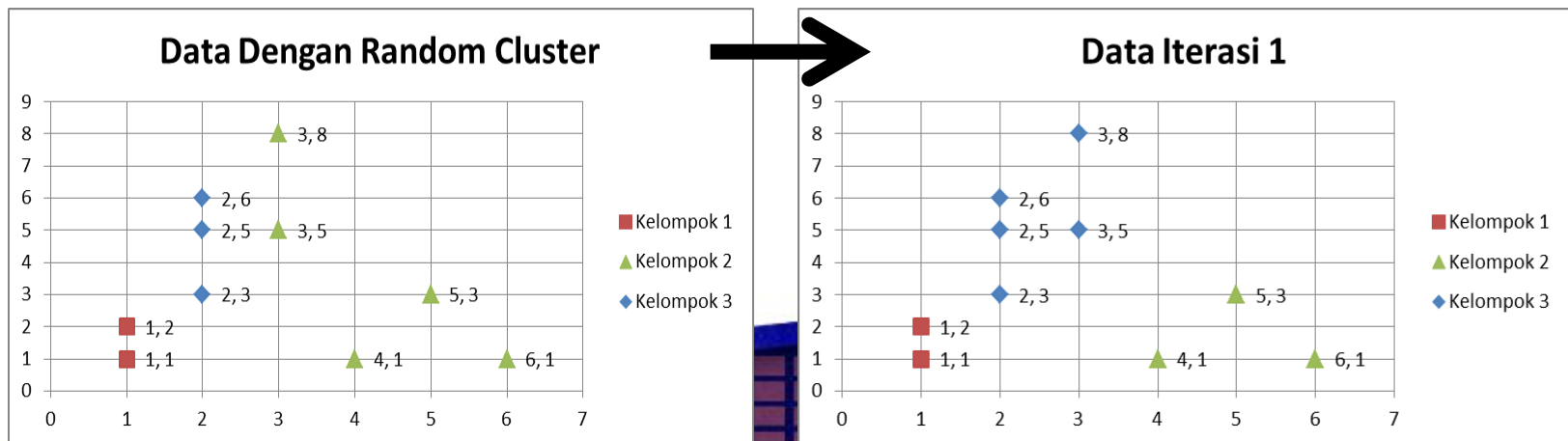
Sehingga, F baru = 1.0000 + 13.1746 + 3.3333 = 17.5079

Delta = | F baru – F lama | = | 17.5079 – 0 | = 17.5079 ( > T ) , Lanjutkan !



Iterasi 1 : (Mengalokasikan Setiap Data Pada Centroid Terdekat)

Data	F x	F y	K 1	K 2	K 3	Jarak Ke C 1	Jarak Ke C 2	Jarak Ke C 3	Min	Kelompok Baru
1	1	1	*			0.5000	4.1231	3.8006	0.5000	1
2	4	1		*		3.0414	2.6077	4.1767	2.6077	2
3	6	1		*		5.0249	3.1623	5.4263	3.1623	2
4	1	2	*			0.5000	3.5777	2.8480	0.5000	1
5	2	3			*	1.8028	2.2804	1.6667	1.6667	3
6	5	3		*		4.2720	1.0000	3.4319	1.0000	2
7	2	5			*	3.6401	2.6077	0.3333	0.3333	3
8	3	5			*	4.0311	1.8439	1.0541	1.0541	3
9	2	6			*	4.6098	3.2558	1.3333	1.3333	3
10	3	8			*	6.8007	4.5607	3.4801	3.4801	3
Total			2	3	5	1.0000	13.1746	3.3333		







- Menghitung Centroid Setiap Cluster :

Data	F x	F y	K 1	K 2	K 3	K 1 F x	K 1 F y	K 2 F x	K 2 F y	K 3 F x	K 3 F y
1	1	1	*			1	1				
2	4	1		*				4	1		
3	6	1		*				6	1		
4	1	2	*			1	2				
5	2	3			*					2	3
6	5	3		*				5	3		
7	2	5			*					2	5
8	3	5			*					3	5
9	2	6			*					2	6
10	3	8			*					3	8
Total			2	3	5	2	3	15	5	12	27

- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = $2 / 2 = 1$	Total K1Fy / Total K1 = $3 / 2 = 1.5$
2	Total K2Fx / Total K2 = $15 / 3 = 5$	Total K2Fy / Total K2 = $5 / 3 = 1.6667$
3	Total K3Fx / Total K3 = $12 / 5 = 2.4$	Total K3Fy / Total K3 = $27 / 5 = 5.4$



- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = 2 / 2 = 1	Total K1Fy / Total K1 = 3 / 2 = 1.5
2	Total K2Fx / Total K2 = 15 / 3 = 5	Total K2Fy / Total K2 = 5 / 3 = 1.6667
3	Total K3Fx / Total K3 = 12 / 5 = 2.4	Total K3Fy / Total K3 = 27 / 5 = 5.4

- Menghitung Jarak Data Ke Centroid :

Data	F x	F y	Jarak Ke C 1	Jarak Ke C 2	Jarak Ke C 3	Min	Kelompok Baru	Kelompok Sebelumnya
1	1	1	0.5000	4.0552	4.6174	0.5000	1	1
2	4	1	3.0414	1.2019	4.6819	1.2019	2	2
3	6	1	5.0249	1.2019	5.6851	1.2019	2	2
4	1	2	0.5000	4.0139	3.6770	0.5000	1	1
5	2	3	1.8028	3.2830	2.4331	1.8028	1	3
6	5	3	4.2720	1.3333	3.5384	1.3333	2	2
7	2	5	3.6401	4.4845	0.5657	0.5657	3	3
8	3	5	4.0311	3.8873	0.7211	0.7211	3	3
9	2	6	4.6098	5.2705	0.7211	0.7211	3	3
10	3	8	6.8007	6.6416	2.6683	2.6683	3	3
Total			1.0000	3.7370	7.1093	(Total berdasarkan kelompok sebelumnya)		

Sehingga, F baru = 1.0000 + 3.7370 + 7.1093 = 11.8464

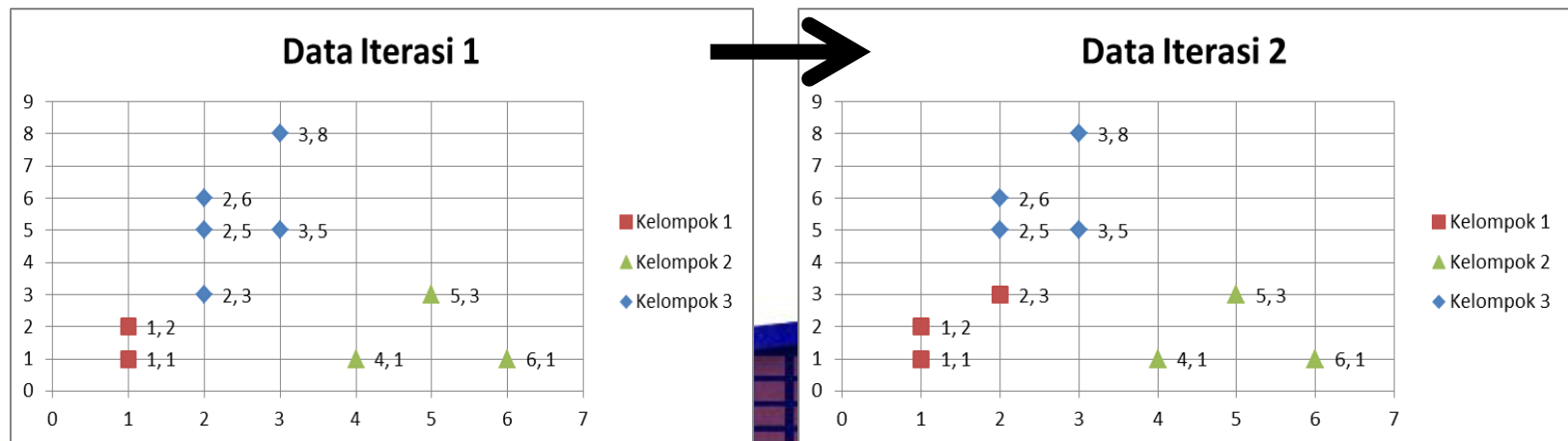
Delta = | F baru – F lama | = | 11.8464 – 17.5079 | = 5.6615 ( > T ) ,

Lanjutkan !



Iterasi 2 : (Mengalokasikan Setiap Data Pada Centroid Terdekat)

Data	F x	F y	K 1	K 2	K 3	Jarak Ke C 1	Jarak Ke C 2	Jarak Ke C 3	Min	Kelompok Baru
1	1	1	*			0.5000	4.0552	4.6174	0.5000	1
2	4	1		*		3.0414	1.2019	4.6819	1.2019	2
3	6	1		*		5.0249	1.2019	5.6851	1.2019	2
4	1	2	*			0.5000	4.0139	3.6770	0.5000	1
5	2	3			*	1.8028	3.2830	2.4331	1.8028	1
6	5	3		*		4.2720	1.3333	3.5384	1.3333	2
7	2	5			*	3.6401	4.4845	0.5657	0.5657	3
8	3	5			*	4.0311	3.8873	0.7211	0.7211	3
9	2	6			*	4.6098	5.2705	0.7211	0.7211	3
10	3	8			*	6.8007	6.6416	2.6683	2.6683	3
Total			2	3	5	1.0000	3.7370	7.1093		





- Menghitung Centroid Setiap Cluster :

Data	F x	F y	K 1	K 2	K 3	K 1 F x	K 1 F y	K 2 F x	K 2 F y	K 3 F x	K 3 F y
1	1	1	*			1	1				
2	4	1		*				4	1		
3	6	1		*				6	1		
4	1	2	*			1	2				
5	2	3	*			2	3				
6	5	3		*				5	3		
7	2	5			*					2	5
8	3	5			*					3	5
9	2	6			*					2	6
10	3	8			*					3	8
Total			3	3	4	4	6	15	5	10	24

- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = $4 / 3 = 1.3333$	Total K1Fy / Total K1 = $6 / 3 = 2$
2	Total K2Fx / Total K2 = $15 / 3 = 5$	Total K2Fy / Total K2 = $5 / 3 = 1.6667$
3	Total K3Fx / Total K3 = $10 / 4 = 2.5$	Total K3Fy / Total K3 = $24 / 4 = 6$



- Hasil Centroid Setiap Cluster :

Kelompok	Centroid Fitur x	Centroid Fitur y
1	Total K1Fx / Total K1 = 4 / 3 = 1.3333	Total K1Fy / Total K1 = 6 / 3 = 2
2	Total K2Fx / Total K2 = 15 / 3 = 5	Total K2Fy / Total K2 = 5 / 3 = 1.6667
3	Total K3Fx / Total K3 = 10 / 4 = 2.5	Total K3Fy / Total K3 = 24 / 4 = 6

- Menghitung Jarak Data Ke Centroid :

Data	F x	F y	Jarak Ke C 1	Jarak Ke C 2	Jarak Ke C 3	Min	Kelompok Baru	Kelompok Sebelumnya
1	1	1	1.0541	4.0552	5.2202	1.0541	1	1
2	4	1	2.8480	1.2019	5.2202	1.2019	2	2
3	6	1	4.7726	1.2019	6.1033	1.2019	2	2
4	1	2	0.3333	4.0139	4.2720	0.3333	1	1
5	2	3	1.2019	3.2830	3.0414	1.2019	1	1
6	5	3	3.8006	1.3333	3.9051	1.3333	2	2
7	2	5	3.0732	4.4845	1.1180	1.1180	3	3
8	3	5	3.4319	3.8873	1.1180	1.1180	3	3
9	2	6	4.0552	5.2705	0.5000	0.5000	3	3
10	3	8	6.2272	6.6416	2.0616	2.0616	3	3
Total			2.5893	3.7370	4.7976	(Total berdasarkan kelompok sebelumnya)		

Sehingga, F baru = 2.5893 + 3.7370 + 4.7976 = 11.1239

Delta = | F baru – F lama | = | 11.1239 – 11.8464 | = 0.7224 ( < T ) ,

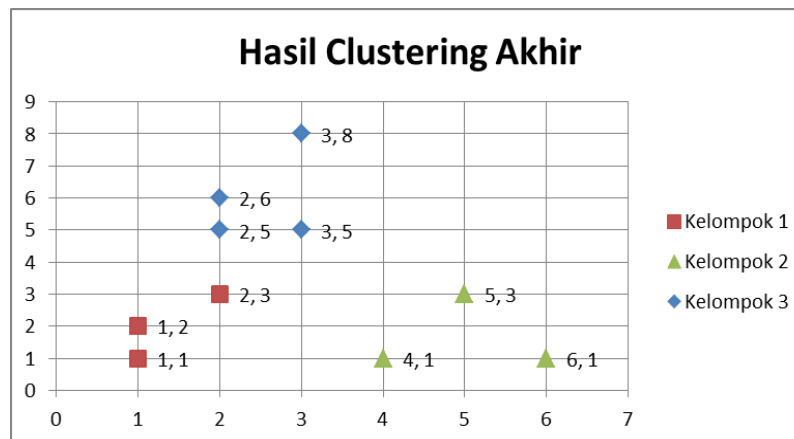
Stop Iterasi !



- Hasil Akhir Clustering Data :

Data	F x	F y	Kelompok Baru
1	1	1	1
2	4	1	2
3	6	1	2
4	1	2	1
5	2	3	1
6	5	3	2
7	2	5	3
8	3	5	3
9	2	6	3
10	3	8	3

- Visualisasi Hasil Akhir Clustering :





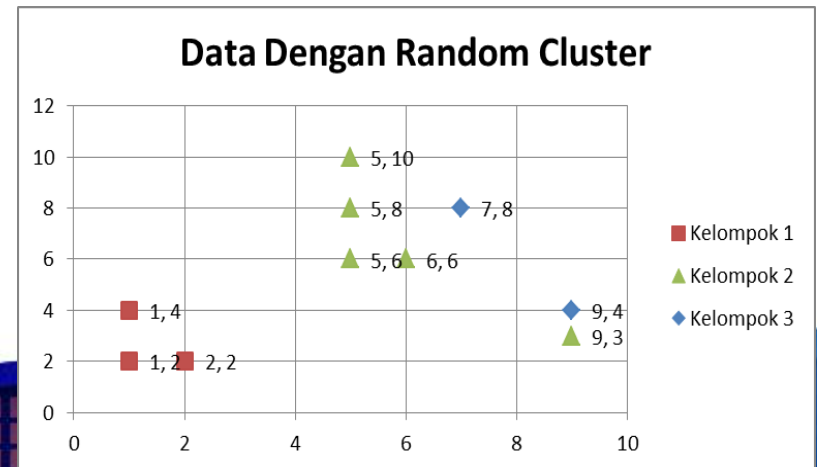
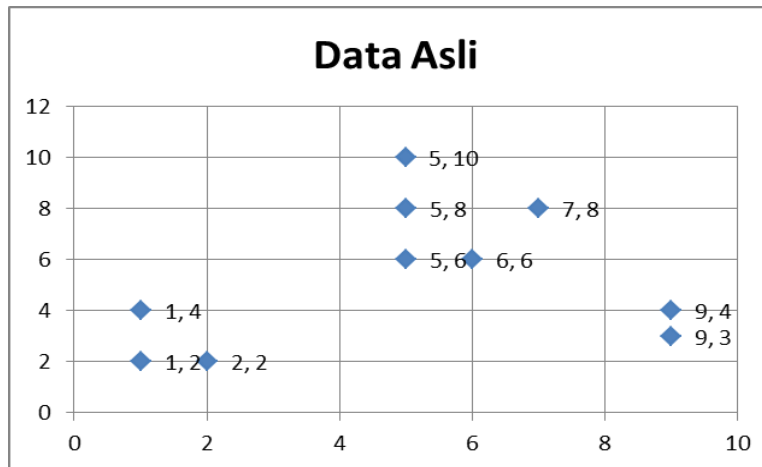
# Latihan Kelompok

- Perhatikan dataset berikut :

Data	Fitur x	Fitur y	Kelompok 1	Kelompok 2	Kelompok 3
1	5	8		*	
2	5	6		*	
3	9	3		*	
4	1	4	*		
5	7	8			*
6	1	2	*		
7	2	2	*		
8	9	4			*
9	5	10		*	
10	6	6		*	

Inisialisasi :  
 $K = 3$ ,  
IterasiMax = 3.  
Tentukan Hasil Akhir  
Clusteringnya !

- Bentuk Visualisasi data :





# Thank You !

afif.supianto@ub.ac.id  
081 331 834 734 / 088 160 127 40