

Nama : David Chandra

NIM : 42419023

a.

Nama : David Chandra  
NIM : 42419023

A. Ada 6 buah data yang akan dikelompokkan menjadi 2 cluster.

Sampel data	X	Y	Kelompok / cluster
1	100	50	
2	40	60	
3	30	70	
4	90	10	
5	65	40	
6	25	35	

Pertama kita akan menghitung centroid. ambil data ke 1 dan ke 2 sebagai pertama. menggunakan rumus Euclidean Distance untuk mendapatkan jarak minimum data terhadap centroid.

cluster	X	Y	Rumus
K1	100	50	
K2	40	60	

Berikut ini adalah rumus dari Euclidian Distance.

$$[(x,y), (a,b)] = \sqrt{(x-a)^2 + (y-b)^2}$$

Penghitungan pertama

mula menghitung centroid pertama. menentukan jarak dari data ke 1 ke data 1, data 1 ke data 2, data 2 ke data 1 dan data 2 ke data 2.

$$\text{cluster 1 } (100, 50) = \sqrt{(100-100)^2 + (50-50)^2} = 0$$

(jarak cluster 1 ke cluster 1)

Jarak dari cluster 2 ke cluster 1  $(40,60) \leftrightarrow (100,50)$ :

$$\sqrt{(40-100)^2 + (60-50)^2} = \sqrt{(-60)^2 + (10)^2} = \sqrt{3600 + 100} = \sqrt{3700} = 60.83$$

Jarak dari cluster 1 ke cluster 2  $(100,50) \leftrightarrow (40,60)$ :

$$\sqrt{(100-40)^2 + (50-60)^2} = \sqrt{(60)^2 + (-10)^2} = \sqrt{3600 + 100} = \sqrt{3700} = 60.83$$

Jarak cluster 2 ke cluster 2  $(40,60) \cdot \sqrt{(40-40)^2 + (60-60)^2} = 0$

cluster	centroid		kelompok cluster
	X	Y	
K1 $(100,50)$	0	60.83	1
K2 $(40,60)$	60.83	0	2

Dari tabel diatas, kita lihat jarak minimum dari data 1 ke data 1 adalah 0; dan jarak minimum dari data 2 ke data 2 yaitu 0. Sehingga yang menjadi centroid K1 adalah data 1 dan data 2 menjadi centroid K2.

### Perhitungan kedua

Setelah mendapatkan centroid, kita berulah ke data ke-3 yaitu  $(30,70)$ . mencari hitung jarak data ke-3 terhadap centroid 1 dan centroid 2. sehingga hasilnya nanti kita mengetahui data ke-3 akan masuk ke cluster K1 atau K2.

Berikut adalah perhitungannya.

Jarak dataset terhadap cluster 1:

$$(100,50) \leftrightarrow (30,70) = \sqrt{(30-100)^2 + (70-50)^2} = \sqrt{(-70)^2 + (20)^2} = \sqrt{4900 + 400} = \sqrt{5300} = 72.80$$

hitung jarak dataset terhadap cluster 2:

$$(40,60) \leftrightarrow (30,70)$$

$$\sqrt{(30-40)^2 + (70-60)^2} = \sqrt{(-10)^2 + (10)^2} = \sqrt{100 + 100} = \sqrt{200} = 14.14$$

Dilihat dari jarak minimum data terhadap centroid, data ke-3 lebih dekat ke K2 dengan nilai 14.14.

selanjutnya kita mengUPDATE nilai centroid. karena data masuk ke K2, maka centroid K2 diupdate dengan cara:

$$x\text{-centroid-bar} = (x\text{-K2} + x\text{-data3})/2$$

$$y\text{-centroid-bar} = (y\text{-K2} + y\text{-data3})/2$$

cluster	x	y
K1	100	50
K2	$\frac{40+30}{2} = 35$	$\frac{60+70}{2} = 65$

Jadi cluster centroid yang baru adalah:

cluster	x	y
K1	100	50
K2	35	65

Perhitungan ketiga

berlanjut ke data ke-4 yaitu (90,10).

hitung jarak dataset terhadap cluster 1.

$$(100,50) \leftrightarrow (90,10) = \sqrt{(90-100)^2 + (10-50)^2} = \sqrt{100+1600} = \sqrt{1700} = 41.23$$

hitung jarak dataset terhadap cluster 2.

karena K2 telah diupdate, jadi kita harus menggunakan nilai centroid K2 yang baru yaitu : (35,65). Hal ini juga berpengaruh apabila centroid K1 diupdate.

$$(35,65) \leftrightarrow (90,10)$$

$$\sqrt{(90-35)^2 + (10-65)^2} = \sqrt{(55)^2 + (-55)^2} = \sqrt{3025+3025} = \sqrt{6050} = 77.78$$

Dari hasil diatas, kita mendapatkan jika dataset ke-4 masuk dalam cluster 1.

seperti langkah sebelumnya, kita update kembali centroid K1 dengan dataset ke-4.

cluster	x	y
K1	<del>100+90</del> 2	<del>50+10</del> 2
K2	35	65

Jadi cluster centroid yang baru adalah:

cluster	x	y
K1	95	30
K2	35	65

Perhitungan ke empat berlanjut ke data ke-5 yaitu (65,40)

hitung jarak dataset terhadap cluster 1  
 $(95,30) \leftrightarrow (65,40)$

$$\sqrt{(65-95)^2 + (40-30)^2} = \sqrt{(-30)^2 + (10)^2} = \sqrt{900 + 100} = \sqrt{1000} = 31.62$$

hitung jarak dataset terhadap cluster 2

$$(35,65) \leftrightarrow (65,40)$$

$$\sqrt{(65-35)^2 + (40-65)^2} = \sqrt{(30)^2 + (-25)^2} = \sqrt{900 + 625} = \sqrt{1525} = 39.06$$

Dari hasil diatas kita mendapatkan jika dataset ke-5 masuk dalam cluster 1.

cluster	x	y
K1	$\frac{95+65}{2}, 80$	$\frac{30+40}{2}, 35$
K2	35	65

Jadi cluster centroid yang baru adalah :

cluster	x	y
K1	80	35
K2	35	65

Perhitungan kelima  
berlanjut ke data ke-6 yaitu (25, 35)

hitung jarak dataset terhadap cluster 1  
 $(80, 35) \leftrightarrow (25, 35)$

$$\sqrt{(25-80)^2 + (35-35)^2} = \sqrt{(-55)^2 + 0^2} = \sqrt{3025} = 55$$

hitung jarak dataset terhadap cluster 2

$$(35, 65) \leftrightarrow (25, 35)$$

$$\sqrt{(25-35)^2 + (35-65)^2} = \sqrt{(-10)^2 + (-30)^2} = \sqrt{100+900} = \sqrt{1000} = 31.62$$

Semua data sudah terkompakkan berdasarkan cluster 1 dan cluster 2. berikut hasil pengelompokannya pada tabel.

sampel data	x	y	kelompok / cluster
1	100	50	1
2	40	60	2
3	30	70	2
4	90	10	1
5	65	40	1
6	25	35	2

b.

Nama: David Chandra  
NIM: 19419023

B. Dari data dengan nilai  $k = 2$ , menghasilkan:

1. 2 kuster berdasarkan usia dan penghasilan.
2. a. Kuster ke 1 adalah kelompok usia antara  $40 \leq x \leq 70$   
dengan penghasilan antara  $2000 \leq y \leq 8000$ .
- b. Kuster ke 2 adalah kelompok usia antara  $20 \leq x \leq 70$   
dengan pendapatan antara  $10000 \leq y \leq 18000$ .

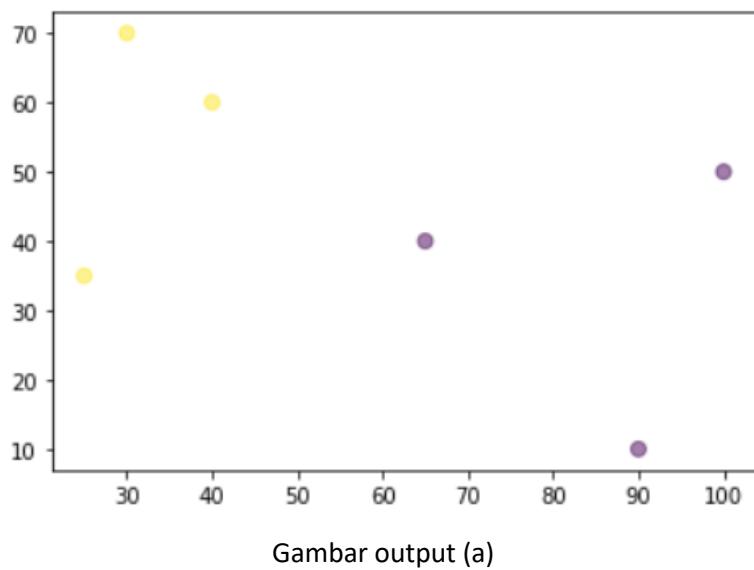
Dari data dengan nilai  $k = 4$ , menghasilkan:

1. 4 kuster berdasarkan usia dan penghasilan.
2. a. Kuster ke 1 adalah kelompok usia antara  $40 \leq x \leq 70$   
dengan penghasilan antara  $2000 \leq y \leq 4000$ .
- b. Kuster ke 2 adalah kelompok usia antara  $40 \leq x \leq 70$   
dengan penghasilan antara  $6000 \leq y \leq 8000$ .
- c. Kuster ke 3 adalah kelompok usia antara  $30 \leq x \leq 70$   
dengan penghasilan antara  $8000 \leq y \leq 14000$ .
- d. Kuster ke 4 adalah kelompok usia antara  $20 \leq x \leq 50$   
dengan penghasilan antara  $14000 \leq y \leq 18000$ .

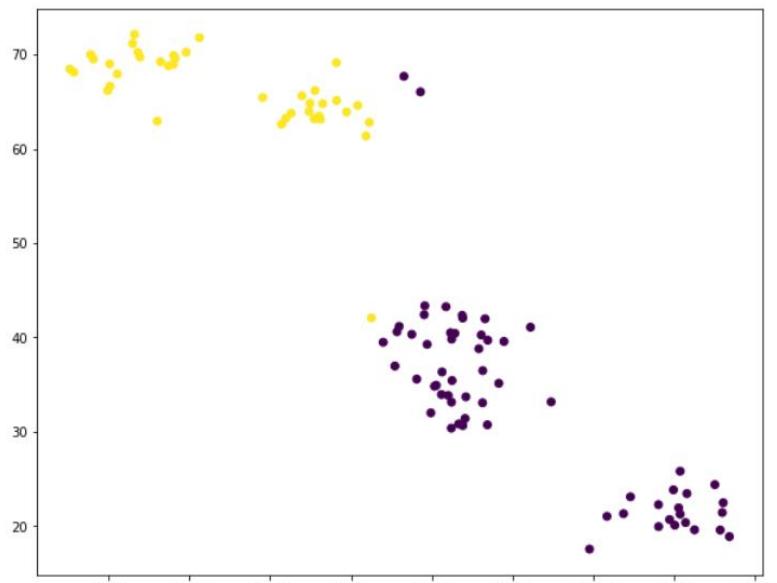
Dari data dengan nilai  $k = 6$ , menghasilkan:

1. 6 kuster berdasarkan usia dan penghasilan.
2. a. Kuster ke 1 adalah kelompok usia antara  $60 \leq x \leq 70$   
dengan penghasilan antara  $2000 \leq y \leq 4000$ .
- b. Kuster ke 2 adalah kelompok usia antara  $60 \leq x \leq 70$   
dengan penghasilan antara  $6000 \leq y \leq 8000$ .
- c. Kuster ke 3 adalah kelompok usia antara  $30 \leq x \leq 70$   
dengan penghasilan antara  $8000 \leq y \leq 18000$ .
- d. Kuster ke 4 adalah kelompok usia antara  $30 \leq x \leq 40$   
dengan penghasilan antara  $10000 \leq y \leq 12000$ .
- e. Kuster ke 5 adalah kelompok usia antara  $20 \leq x \leq 40$   
dengan penghasilan antara  $12000 \leq y \leq 16000$ .
- f. Kuster ke 6 adalah kelompok usia antara  $20 \leq x \leq 30$   
dengan penghasilan antara  $16000 \leq y \leq 18000$ .

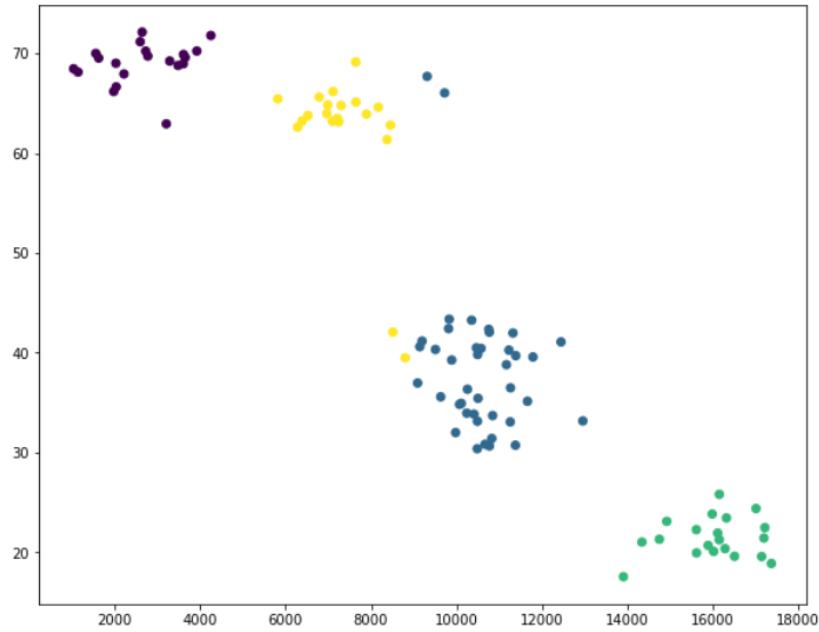
C.



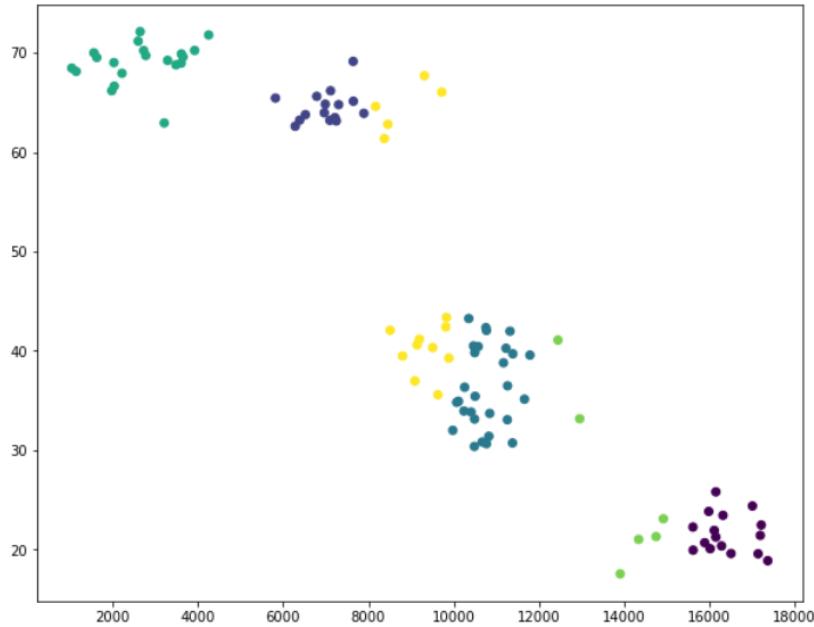
Gambar output (a)



Gambar output menggunakan nilai K-2



Gambar output menggunakan nilai K-4



Gambar output menggunakan nilai K-6