

☐ Nama: Angga Saputra

☐ NIM = 42419001

☐ Matkul = Data Mining dan Wewahosting.

☐ A. Diketahui 6 data yang akan dikelompokkan menjadi 2 cluster yaitu K_1 dan K_2

Data	x	y	kelompok/cluster
1	100	50	
2	40	60	
3	30	70	
4	80	10	
5	65	40	
6	25	35	

☐ Langkah 1. mengambil data ke-1 dan ke-2 untuk menghitung nilai minimum jarak antar centroid. menggunakan rumus: Euclidean Distance.

cluster	x	y
K_1	100	50
K_2	40	60

☐ menghitung jarak data 1 ke data 1, data 1 ke data 2, data 2 ke data 1 dan data 2 ke data 2.

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

(Euclidean Distance)

☐ Cluster 1 (100, 50) = $\sqrt{(100-100)^2 + (50-50)^2} = 0$ (jarak C_1 ke C_1)

☐ Jarak C_2 ke C_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 C_1 ke C_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 C_2 ke C_2 = (40, 60) = $\sqrt{(40-40)^2 + (60-60)^2} = 0$

(7)

Cluster	Centroid		Kelompok
	x	y	
$K_1(100,50)$	0	60,83	1
$K_2(40,60)$	60,83	0	2

Langkah 2. menghitung jarak antar dataset terhadap cluster

$$(100,50) \leftrightarrow (30,70) = \sqrt{(30-100)^2 + (70-50)^2}$$

$$= \sqrt{(-70)^2 + (20)^2} = \sqrt{4900 + 400} = \sqrt{5300} = 72,80$$

Jarak dataset terhadap cluster ke 2

$$(40,60) \leftrightarrow (30,70) = \sqrt{(30-40)^2 + (70-60)^2}$$

$$= \sqrt{(-10)^2 + (10)^2} = \sqrt{100 + 100} = \sqrt{200} = 14,14$$

Jarak dataset 3 terhadap cluster 2 dan cluster 1 lebih

$$(-30,70) \leftrightarrow \text{terkat pada cluster 2 14,14}$$

Langkah 3. memperbarui nilai Centroid karena data masuk ke K_2

$$x\text{-centroid-baru} = (x_{K_2} + x_{\text{data 3}}) / 2$$

$$y\text{-centroid-baru} = (y_{K_2} + y_{\text{data 3}}) / 2$$

$$x_{K_2} = 40 + 30 / 2 = 35$$

$$y_{K_2} = 60 + 70 / 2 = 65$$

cluster	x	y
K_1	100	50
K_2	35	65

Langkah 4. menghitung dataset ke-4 (90,10)

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

$$= \sqrt{(-10)^2 + (-40)^2} = \sqrt{100 + 1600} = \sqrt{1700} = 41,23$$

$$(35,65) \leftrightarrow (90,10) = \sqrt{(90-35)^2 + (10-65)^2}$$

$$= \sqrt{(55)^2 + (-55)^2} = \sqrt{3025 + 3025} = \sqrt{6050} = 77,78$$

masuk ke cluster 1

(3)

Langkah 5. Memperbarui nilai centroid

$$X_{-k_1} = 100 + 90 / 2 = 95$$

$$Y_{-k_1} = 50 + 10 / 2 = 30$$

cluster	x	y
k1	95	30
k2	35	65

Langkah 6 Jarak dataset ke 1 (65, 40)

$$(95, 30) \leftrightarrow (65, 40) = \sqrt{(65-95)^2 + (40-30)^2}$$

$$= \sqrt{(-30)^2 + (10)^2} = \sqrt{900 + 100} = \sqrt{1000} = 31,62$$

$$(35, 65) \leftrightarrow (65, 40) = \sqrt{(65-35)^2 + (40-65)^2}$$

$$= \sqrt{(30)^2 + (-25)^2} = \sqrt{900 + 625} = \sqrt{1525} = 39,05$$

masuk cluster k1

Langkah 8 memperbarui nilai centroid

$$X_{-k_1} = 95 + 85 / 2 = 90$$

$$Y_{-k_1} = 30 + 40 / 2 = 35$$

cluster	x	y
k1	90	35
k2	55	65

Langkah 9 Jarak dataset 6 (25, 55)

$$(90, 35) \leftrightarrow (25, 55) = \sqrt{(25-90)^2 + (55-35)^2}$$

$$= \sqrt{(-65)^2 + (20)^2} = \sqrt{4225 + 400} = \sqrt{4625} = 68,01$$

$$(55, 65) \leftrightarrow (25, 55) = \sqrt{(25-55)^2 + (55-65)^2}$$

$$= \sqrt{(-30)^2 + (-10)^2} = \sqrt{900 + 100} = \sqrt{1000} = 31,62$$

cluster ke 2

Date.

4

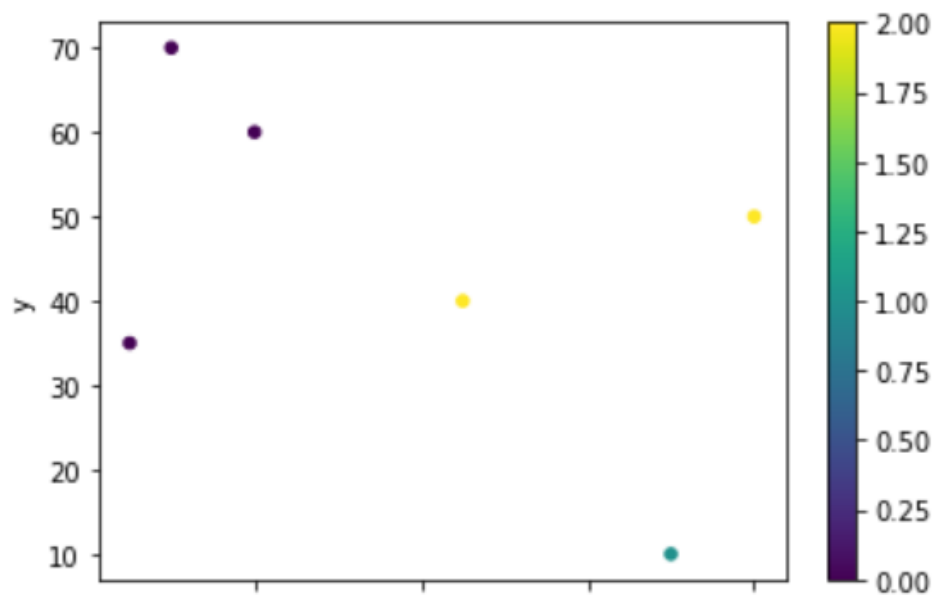
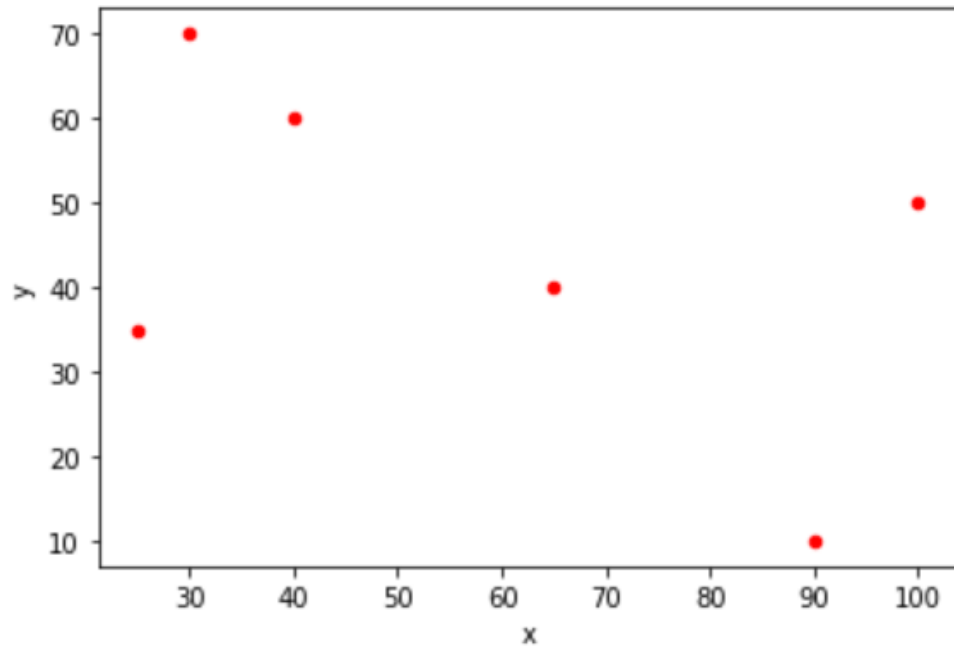
Hasil Perhitungan

Data	x	y	cluster
1	100	50	1
2	40	60	2
3	30	70	2
4	80	10	1
5	65	40	1
6	25	35	2

- Hasil dari data dengan nilai $k = 2$, yaitu akan menghasilkan
1. 2 jenis cluster berdasarkan income dan usia dengan:
 - a. Cluster 1 merupakan kelompok usia 40-70 dengan income antara 2000 s/d 8000
 - b. Cluster 2 merupakan kelompok usia 20-70 dengan income antara 8000 s/d 18000
- Hasil dari data dengan nilai $k = 4$, yaitu akan menghasilkan
1. 4 jenis cluster berdasarkan income dan usia dengan:
 - a. Cluster 1 merupakan kelompok usia 60-70 dengan income antara 2000 s/d 4000
 - b. Cluster 2 merupakan kelompok usia 40-70 dengan income antara 6000 s/d 8000
 - c. Cluster 3 merupakan kelompok usia 30-70 dengan income antara 8000 s/d 12000
 - d. Cluster 4 merupakan kelompok usia 20-70 dengan income antara 14000 s/d 18000
- Hasil dari data dengan nilai $k = 6$, yaitu akan menghasilkan
1. 6 jenis cluster berdasarkan income dan usia dengan:
 - a. Cluster 1 merupakan kelompok usia 60-70 dengan income antara 2000 s/d 4000
 - b. Cluster 2 merupakan kelompok usia 60-70 dengan income antara 6000 s/d 8000
 - c. Cluster 3 merupakan kelompok usia 30-70 dengan income antara 8000 s/d 10000
 - d. Cluster 4 merupakan kelompok usia 30-70 dengan income antara 10000 s/d 12000
 - e. Cluster 5 merupakan kelompok usia 20-70 dengan income antara 12000 s/d 14000
 - f. Cluster 6 merupakan kelompok usia 20-70 dengan income antara 14000 s/d 18000

C. Hasil Visualisasi

Soal A



Soal B

```
Model 1 = [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1  
1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
1 1 1 1 1 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]  
Model 2 = [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0  
0 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2]  
Model 2 = [4 1 4 4 4 4 1 4 1 4 1 4 4 4 4 4 4 4 4 4 0 0 0 0 0 0 5 5 0 0 5 5 0 5 0 0 0 0  
0 0 5 2 2 2 2 2 2 2 2 2 5 2 2 2 1 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
3 3 3 3 3 3 5 2 5 2 5 2 2 1 2 5 2 5 2 2 2 5]
```

