

Table

Sample Data	X	Y	Kelompok / Cluster
1	100	50	
2	40	60	
3	30	70	
4	90	10	
5	65	40	
6	25	35	

Kita akan mengaplikasikan K-Means Clustering untuk data diatas menjadi 2 cluster. Pertama kita akan hitung Cluster Centroid.

Cluster	X	Y
K1	100	50
K2	40	60

Perhitungan menggunakan persamaan Euclidean Distance

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

Perhitungan Pertama ;

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

( 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  $(100, 50) \leftrightarrow (40, 60) =$

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

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

Sehingga ;

Cluster	Centroid		Kelompok Cluster
	X	Y	
K1 (100, 50)	0	60.83	1
K2 (40, 60)	60.83	0	2

Perhitungan kedua.

Langkah selanjutnya kita beralih ke data 3 yaitu (30, 70). Kita mulai menghitung jarak dataset terhadap cluster 1.

$$(100, 50) \leftrightarrow (30, 70) =$$

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

Kemudian kita hitung jarak dataset terhadap cluster 2.

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

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

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(30 , 70)	72.80	14.14	2

Dataset ke-3 masuk dalam kelompok cluster 2 , karena jarak minimum / terdekat dataset adalah terhadap cluster 2 yaitu 14.14.

Kemudian kita update Centroid

Cluster	X	Y
K1	100	50
K2	$= \frac{40 + 30}{2} = 35$	$= \frac{60 + 70}{2} = 65$

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	100	50
K2	35	65

Perhitungan ketiga

Kita lanjutkan kembali menghitung dataset ke-4 yaitu (90, 10). Kita mulai menghitung jarak dataset terhadap cluster 1.

$$(100, 50) \leftrightarrow (90, 10) =$$

$$\sqrt{(90 - 100)^2 + (10 - 50)^2} = \sqrt{(-10)^2 + (-40)^2} = \sqrt{100 + 1600} = \sqrt{1700} = 41.23$$

Kemudian kita hitung jarak dataset terhadap cluster 2. Sangat perlu diingat, centroid 2 yang kita gunakan adalah yang sudah di update di perhitungan kedua diatas yaitu (35, 65).

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

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

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(90 , 10)	41.23	77.78	1

Dataset ke-3 masuk dalam kelompok cluster 1 , karena jarak minimum / terdekat dataset adalah terhadap cluster 1 yaitu 41.23.

Kemudian kita update Centroid

Cluster	X	Y
K1	$= \frac{100 + 90}{2} = 95$	$= \frac{50 + 10}{2} = 30$
K2	35	65

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	95	30
K2	35	65

Perhitungan keempat

Kita lanjutkan kembali menghitung dataset ke-5 yaitu (80, 40). Kita mulai menghitung jarak dataset terhadap cluster 1. Perhatikan kembali, centroid 1 yang digunakan adalah yang telah terupdate yaitu (95, 30).

$$(95, 30) \leftrightarrow (65, 40) =$$

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

Kemudian kita hitung jarak dataset terhadap cluster 2.

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

$$\sqrt{(65 - 35)^2 + (40 - 65)^2} = \sqrt{(30)^2 + (-15)^2} = \sqrt{900 + 225} = \sqrt{1125} = 33.54$$

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(80, 40)	31.62	33.54	1

Dataset ke-4 masuk dalam kelompok cluster 1, karena jarak minimum / terdekat dataset adalah terhadap cluster 1 yaitu 31.62.

Kemudian kita update Centroid

Cluster	X	Y
K1	$= \frac{95 + 65}{2} = 80$	$= \frac{30 + 40}{2} = 35$
K2	35	65

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	80	35
K2	35	65

Perhitungan kelima

Kita lanjutkan kembali menghitung dataset ke-5 yaitu (25, 35). Kita mulai menghitung jarak dataset terhadap cluster 1. Perhatikan kembali, centroid 1 yang digunakan adalah yang telah terupdate yaitu (80, 35).

$$(80, 35) \leftrightarrow (25, 35) =$$

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

Kemudian kita hitung jarak dataset terhadap cluster 2.

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

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

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(80, 40)	55	31.62	<b>2</b>

Semua data telah kita dapatkan, dan masing-masing data sudah terkelompokkan berdasarkan cluster 1 dan cluster 2. Selanjutnya kita input hasil pengelompokkannya pada table pertama. Jadi hasilnya seperti berikut :

Sample 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