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 daa diaas menjadi 2 cluster. Pertama kita akan hitung Centroid.

|  |  |  |
| --- | --- | --- |
| **Cluster** | **X** | **Y** |
| K1 | 100 | 50 |
| K2 | 40 | 60 |

Perhitungan menggunakan persamaan Euclidean Distance

Perhitungan Pertama ;

Cluster 1 (100, 50) =

( jarak cluster 1 ke cluster 1)

Jarak dari Cluster 2 ke cluster 1 (40 , 60) ↔ (100,50) =

Jarak dari cluster 1 ke cluster 2 (100, 50) ↔ (40, 60) =

Jarak cluster 2 ke cluster 2 (40, 60) =

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) ↔ (30, 70) =

Kemudian kita hitung jarak dataset terhadap cluster 2.

(40, 60) ↔ (30, 70) =

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 |  |  |

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) ↔ (90, 10) =

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

(35, 65) ↔ (90, 10) =

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 |  |  |
| 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 yan gtelah terupdate yaitu (95, 30).

(95, 30) ↔ (65, 40) =

Kemudian kita hitung jarak dataset terhadap cluster 2.

(35, 65) ↔ (65, 40) =

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 |  |  |
| 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 yan gtelah terupdate yaitu (80, 35).

(80, 35) ↔ (25, 35) =

Kemudian kita hitung jarak dataset terhadap cluster 2.

(35, 65) ↔ (25, 35) =

Sehingga didapatkan :

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Euclidean Distance** | | **Kelompok Cluster** |
| **Cluster 1** | **Cluster 2** |
| (80, 40) | 55 | 31.62 | **2** |

Semua data telah kita dapatkan, dan masing-masing data sudah terkelompokkan berdasarkan cluster 1 dan cluster 2. Selanjutnya kita input hasil pengelompokkanya 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 |