

ALGORITMA K-MEANS

DATA MINING



Oleh :

Andik Setyawan 11201169

**JURUSAN TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN
KOMPUTER ASIA MALANG**

2015

Pendahuluan

K-means merupakan salah satu algoritma *clustering*. Tujuan algoritma ini yaitu untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas. Hal ini berbeda dengan *supervised learning* yang menerima masukan berupa vektor (x_1, y_1) , (x_2, y_2) , ..., (x_i, y_i) , di mana x_i merupakan data dari suatu data pelatihan dan y_i merupakan label kelas untuk x_i .

Pada algoritma pembelajaran ini, komputer mengelompokkan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pembelajaran ini termasuk dalam *unsupervised learning*. Masukan yang diterima adalah data atau objek dan k buah kelompok (*cluster*) yang diinginkan. Algoritma ini akan mengelompokkan data atau objek ke dalam k buah kelompok tersebut. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut.

K-means ditemukan oleh beberapa orang yaitu Lloyd (1957, 1982), Forgey (1965), Friedman and Rubin (1967), and McQueen (1967). Ide dari *clustering* pertama kali ditemukan oleh Lloyd pada tahun 1957, namun hal tersebut baru dipublikasi pada tahun 1982. Pada tahun 1965, Forgey juga mempublikasi teknik yang sama sehingga terkadang dikenal sebagai Lloyd-Forgey pada beberapa sumber.

Kelebihan dan Kekurangan

Ada beberapa kelebihan pada algoritma k-means, yaitu:

1. Mudah untuk diimplementasikan dan dijalankan.
2. Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat.
3. Mudah untuk diadaptasi.
4. Umum digunakan.

Algoritma k-means memiliki beberapa kelebihan, namun ada kekurangannya juga. Kekurangan dari algoritma tersebut yaitu :

1. Sebelum algoritma dijalankan, k buah titik diinisialisasi secara *random* sehingga pengelompokan data yang dihasilkan dapat berbeda-beda. Jika nilai *random* untuk inisialisasi kurang baik, maka pengelompokan yang dihasilkan pun menjadi kurang optimal.
2. Dapat terjebak dalam masalah yang disebut *curse of dimensionality*. Hal ini dapat terjadi jika data pelatihan memiliki dimensi yang sangat tinggi (Contoh jika data pelatihan terdiri dari 2 atribut maka dimensinya adalah 2 dimensi. Namun jika ada 20 atribut, maka akan ada 20 dimensi). Salah satu cara kerja algoritma ini adalah mencari jarak terdekat antara k buah titik dengan titik lainnya. Jika mencari jarak antar titik pada 2 dimensi, masih mudah dilakukan. Namun bagaimana mencari jarak antar titik jika terdapat 20 dimensi. Hal ini akan menjadi sulit.

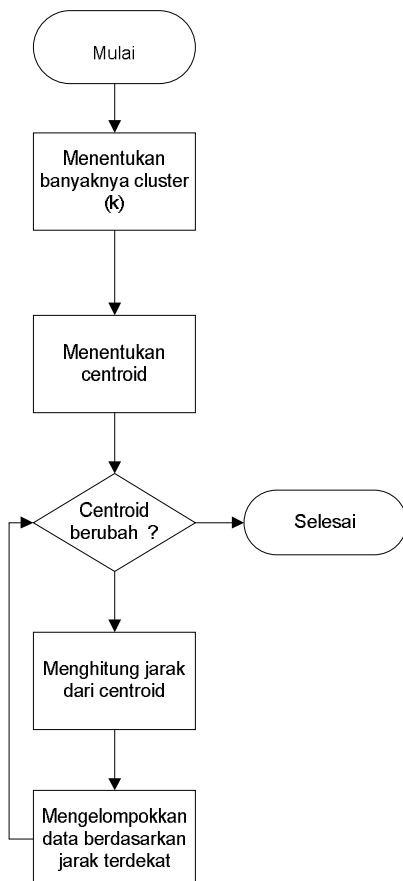
3. Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan k titik yang diinisialisasi secara *random*. Namun jika terdapat banyak sekali titik data (misalnya satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti *kD-Tree* atau *hashing*.

Algoritma k-means merupakan merupakan model *centroid*. Model centroid adalah model yang menggunakan *centroid* untuk membuat *cluster*. *Centroid* adalah "titik tengah" suatu *cluster*. *Centroid* digunakan untuk menghitung jarak suatu objek data terhadap *centroid*. Suatu objek data termasuk dalam suatu *cluster* jika memiliki jarak terpendek terhadap *centroid cluster* tersebut.

Secara umum algoritma k-means adalah:

1. Menentukan banyaknya *cluster* (k).
2. Menentukan *centroid*.
3. Apakah *centroid*-nya berubah ?
 - a. Jika ya, hitung jarak data dari *centroid*.
 - b. Jika tidak, selesai.
4. Mengelompokkan data berdasarkan jarak yang terdekat.

Diagram alir algoritma k-means:



Detail algoritma diatas:

- Tentukan K sebagai jumlah *cluster* yang ingin dibentuk.
- Pilih K *centroid* (titik pusat *cluster*) awal secara random. Dalam menentukan n buah pusat *cluster* awal dilakukan pemilihan bilangan random yang merepresentasikan urutan data input. Pusat awal *cluster* didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan merandom pusat awal dari data.
- Hitung jarak setiap data ke masing-masing *centroids*. Untuk mengukur jarak antara data dengan pusat *cluster* digunakan *Euclidian distance*. Algoritma perhitungan jarak data dengan pusat *cluster*. Langkah-langkahnya, yaitu:
 - Ambil nilai data dan nilai pusat *cluster*
 - Hitung *Euclidian distance* data dengan tiap pusat *cluster*. *Euclidian Distance* merupakan jarak yang didapat dari perhitungan antara semua N data dengan K *centroid* dimana akan memperoleh tingkat kedekatan dengan kelas yang terdekat dengan populasi data tersebut.

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

d = jarak

j = banyaknya data

c = *centroid*

x = data

c = *centroid*

- Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat.
- Tentukan posisi *centroids* yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi sama dengan nilai K atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).
- Jika posisi *centroids* baru dengan *centroids* yang lama tidak sama, maka ulangi kembali dari langkah 3.

Contoh kasus:

Data set

No	NIM	Nama	Nilai Pemrograman		
			VB	PHP	JAVA
1	1120111	Agus Rahman	80	75	75
2	1120112	Ahmad Faril	90	85	80
3	1120113	Ahmad Faizin	77	78	85
4	1120114	Bagus J.	75	80	70
5	1120115	Budi Irawan	89	85	75
6	1120116	Caca S.	95	70	75
7	1120117	Irfan Hakim	75	88	80
8	1120118	Lukman Kurniawan	80	75	85
9	1120119	Sulis Tejo	80	80	85
10	1120120	Tatang	78	80	80

Langkah – langkah:

1. Tentukan banyaknya cluster , misalnya cluster (K) = 2.
2. Tentukan centroid setiap cluster

Untuk menentukan centroid awal (initial centroid) banyak metode yang digunakan. Disini metode yang digunakan adalah mengambil dari data sumber secara acak atau random.

C1 dari data No. 1 sedangkan C2 data No. 7 ,tabel diatas yang bertanda warna kuning.

	VB	PHP	JAVA
c1	80	75	75
c2	75	88	80

Tabel. Centroid pada pengulangan ke-0

3. Hitung jarak data denga centroid. Rumus-rumus untuk menghitung jarak antara lain:
 - a. Euclidean.
 - b. Manhattan / *City Block*.
 - c. Minkowski.

Rumus yang digunakan di sini adalah rumus *Euclidean Distance* :

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

d = jarak

$j = \text{banyaknya data}$

$c = \text{centroid}$

$x = \text{data}$

$c = \text{centroid}$

Supaya lebih mudah kita inisialkan **VB = m** , **PHP = n** , **JAVA = o**.

Jarak data dengan cluster 1 adalah:

$$d(x_1, c_1) = \sqrt{(m_1 - c_{1m})^2 + (n_1 - c_{1n})^2 + (o_1 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (75 - 75)^2 + (75 - 75)^2} = 0$$

$$d(x_2, c_1) = \sqrt{(m_2 - c_{1m})^2 + (n_2 - c_{1n})^2 + (o_2 - c_{1o})^2} = \sqrt{(90 - 80)^2 + (85 - 75)^2 + (80 - 75)^2} = 15$$

$$d(x_3, c_1) = \sqrt{(m_3 - c_{1m})^2 + (n_3 - c_{1n})^2 + (o_3 - c_{1o})^2} = \sqrt{(77 - 80)^2 + (78 - 75)^2 + (85 - 75)^2} = 10,86278$$

$$d(x_4, c_1) = \sqrt{(m_4 - c_{1m})^2 + (n_4 - c_{1n})^2 + (o_4 - c_{1o})^2} = \sqrt{(75 - 80)^2 + (80 - 75)^2 + (70 - 75)^2} = 8,660254$$

$$d(x_5, c_1) = \sqrt{(m_5 - c_{1m})^2 + (n_5 - c_{1n})^2 + (o_5 - c_{1o})^2} = \sqrt{(89 - 80)^2 + (85 - 75)^2 + (75 - 75)^2} = 13,45362$$

$$d(x_6, c_1) = \sqrt{(m_6 - c_{1m})^2 + (n_6 - c_{1n})^2 + (o_6 - c_{1o})^2} = \sqrt{(95 - 80)^2 + (70 - 75)^2 + (75 - 75)^2} = 15,81139$$

$$d(x_7, c_1) = \sqrt{(m_7 - c_{1m})^2 + (n_7 - c_{1n})^2 + (o_7 - c_{1o})^2} = \sqrt{(75 - 80)^2 + (88 - 75)^2 + (80 - 75)^2} = 14,79865$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{1m})^2 + (n_8 - c_{1n})^2 + (o_8 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (75 - 75)^2 + (85 - 75)^2} = 10$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{1m})^2 + (n_9 - c_{1n})^2 + (o_9 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (80 - 75)^2 + (85 - 75)^2} = 11,18034$$

$$d(x_{10}, c_1) = \sqrt{(m_{10} - c_{1m})^2 + (n_{10} - c_{1n})^2 + (o_{10} - c_{1o})^2} = \sqrt{(78 - 80)^2 + (80 - 75)^2 + (80 - 75)^2} = 7,348469$$

Jarak data dengan cluster 2 adalah:

$$d(x_1, c_2) = \sqrt{(m_1 - c_{2m})^2 + (n_1 - c_{2n})^2 + (o_1 - c_{2o})^2} = \sqrt{(80 - 75)^2 + (75 - 88)^2 + (75 - 80)^2} = 14,79865$$

$$d(x_2, c_2) = \sqrt{(m_2 - c_{2m})^2 + (n_2 - c_{2n})^2 + (o_2 - c_{2o})^2} = \sqrt{(90 - 75)^2 + (85 - 88)^2 + (80 - 80)^2} = 15,29706$$

$$d(x_3, c_2) = \sqrt{(m_3 - c_{2m})^2 + (n_3 - c_{2n})^2 + (o_3 - c_{2o})^2} = \sqrt{(77 - 75)^2 + (78 - 88)^2 + (85 - 80)^2} = 11,35782$$

$$d(x_4, c_2) = \sqrt{(m_4 - c_{2m})^2 + (n_4 - c_{2n})^2 + (o_4 - c_{2o})^2} = \sqrt{(75 - 75)^2 + (80 - 88)^2 + (70 - 80)^2} = 12,80625$$

$$d(x_5, c_2) = \sqrt{(m_5 - c_{2m})^2 + (n_5 - c_{2n})^2 + (o_5 - c_{2o})^2} = \sqrt{(89 - 75)^2 + (85 - 88)^2 + (75 - 80)^2} = 15,16575$$

$$d(x_6, c_2) = \sqrt{(m_6 - c_{2m})^2 + (n_6 - c_{2n})^2 + (o_6 - c_{2o})^2} = \sqrt{(95 - 75)^2 + (70 - 88)^2 + (75 - 80)^2} = 27,36786$$

$$d(x_7, c_2) = \sqrt{(m_7 - c_{2m})^2 + (n_7 - c_{2n})^2 + (o_7 - c_{2o})^2} = \sqrt{(75 - 75)^2 + (88 - 88)^2 + (80 - 80)^2} = 0$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{2m})^2 + (n_8 - c_{2n})^2 + (o_8 - c_{2o})^2} = \sqrt{(80 - 75)^2 + (75 - 88)^2 + (85 - 80)^2} = 14,79865$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{2m})^2 + (n_9 - c_{2n})^2 + (o_9 - c_{2o})^2} = \sqrt{(80 - 75)^2 + (80 - 88)^2 + (85 - 80)^2} = 10,67708$$

$$d(x_{10}, c_1) = \sqrt{(m_{10} - c_{2m})^2 + (n_{10} - c_{2n})^2 + (o_{10} - c_{2o})^2} = \sqrt{(78 - 75)^2 + (80 - 88)^2 + (80 - 80)^2} = 8,544004$$

No	Nilai Pemrograman			dc1	dc2
	VB	PHP	JAVA		
1	80	75	75	0	14,79865
2	90	85	80	15	15,29706
3	77	78	85	10,86278	11,35782
4	75	80	70	8,660254	12,80625
5	89	85	75	13,45362	15,16575
6	95	70	75	15,81139	27,36786
7	75	88	80	14,79865	0
8	80	75	85	10	14,79865
9	80	80	85	11,18034	10,67708
10	78	80	80	7,348469	8,544004

4. Kelompokkan data sesuai dengan cluster-nya, yaitu data yang memiliki jarak terpendek.

No	Nilai Pemrograman			dc1	dc2	c1	c2
	VB	PHP	JAVA				
1	80	75	75	0	14,79865	*	
2	90	85	80	15	15,29706	*	
3	77	78	85	10,86278	11,35782	*	
4	75	80	70	8,660254	12,80625	*	
5	89	85	75	13,45362	15,16575	*	
6	95	70	75	15,81139	27,36786	*	
7	75	88	80	14,79865	0		*
8	80	75	85	10	14,79865	*	
9	80	80	85	11,18034	10,67708		*
10	78	80	80	7,348469	8,544004	*	

5. Tentukan centroid yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada *centroid* yang sama. Pusat *cluster* yang baru digunakan untuk melakukan iterasi selanjutnya, jika hasil yang didapatkan belum konvergen. Proses iterasi akan berhenti jika telah memenuhi maksimum iterasi sama dengan nilai K atau hasil yang dicapai sudah konvergen (pusat *cluster* baru sama dengan pusat *cluster* lama).

C1 mempunyai 8 anggota sedangkan c2 mempunyai 2 anggota.

$$C1 = (80+90+77+75+89+95+80+78)/8 = 83$$

$$(75+85+78+80+85+70+75+80)/8 = 78,5$$

$$(75+80+85+70+75+75+85+80)/8 = 78,125$$

$$C2 = (75+80)/2 = 77,5$$

$$(88+80)/2 = 84$$

$$(80+85)/2 = 82,5$$

	VB	PHP	JAVA
c1	83	78,5	78,125
c2	77,5	84	82,5

6. Jika posisi *centroids* baru dengan *centroids* yang lama tidak sama, maka ulangi langkah 3.

Untuk hasil clustering lebih lengkap, berikut tabel-tabel hasil analisis dan perhitungan dari awal sampai selesai:

1. Inialisasi

No	NIM	Nama	Nilai Pemrograman		
			VB	PHP	JAVA
1	1120111		80	75	75
2	1120112		90	85	80
3	1120113		77	78	85
4	1120114		75	80	70
5	1120115		89	85	75
6	1120116		95	70	75
7	1120117		75	88	80
8	1120118		80	75	85
9	1120119		80	80	85
10	1120120		78	80	80

	VB	PHP	JAVA
c1	80	75	75
c2	75	88	80

Jarak data dengan cluster 1 adalah:

$$d(x_1, c_1) = \sqrt{(m_1 - c_{1m})^2 + (n_1 - c_{1n})^2 + (o_1 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (75 - 75)^2 + (75 - 75)^2} = 0$$

$$d(x_2, c_1) = \sqrt{(m_2 - c_{1m})^2 + (n_2 - c_{1n})^2 + (o_2 - c_{1o})^2} = \sqrt{(90 - 80)^2 + (85 - 75)^2 + (80 - 75)^2} = 15$$

$$d(x_3, c_1) = \sqrt{(m_3 - c_{1m})^2 + (n_3 - c_{1n})^2 + (o_3 - c_{1o})^2} = \sqrt{(77 - 80)^2 + (78 - 75)^2 + (85 - 75)^2} = 10,86278$$

$$d(x_4, c_1) = \sqrt{(m_4 - c_{1m})^2 + (n_4 - c_{1n})^2 + (o_4 - c_{1o})^2} = \sqrt{(75 - 80)^2 + (80 - 75)^2 + (70 - 75)^2} = 8,660254$$

$$d(x_5, c_1) = \sqrt{(m_5 - c_{1m})^2 + (n_5 - c_{1n})^2 + (o_5 - c_{1o})^2} = \sqrt{(89 - 80)^2 + (85 - 75)^2 + (75 - 75)^2} = 13,45362$$

$$d(x_6, c_1) = \sqrt{(m_6 - c_{1m})^2 + (n_6 - c_{1n})^2 + (o_6 - c_{1o})^2} = \sqrt{(95 - 80)^2 + (70 - 75)^2 + (75 - 75)^2} = 15,81139$$

$$d(x_7, c_1) = \sqrt{(m_7 - c_{1m})^2 + (n_7 - c_{1n})^2 + (o_7 - c_{1o})^2} = \sqrt{(75 - 80)^2 + (88 - 75)^2 + (80 - 75)^2} = 14,79865$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{1m})^2 + (n_8 - c_{1n})^2 + (o_8 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (75 - 75)^2 + (85 - 75)^2} = 10$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{1m})^2 + (n_9 - c_{1n})^2 + (o_9 - c_{1o})^2} = \sqrt{(80 - 80)^2 + (80 - 75)^2 + (85 - 75)^2} = 11,18034$$

$$d(x_{10},c_1)=\sqrt{(m_{10}-c_{1m})^2+(n_{10}-c_{1n})^2+(o_{10}-c_{1o})^2}=\sqrt{(78-80)^2+(80-75)^2+(80-75)^2}=7,348469$$

Jarak data dengan cluster 2 adalah:

$$d(x_1,c_2)=\sqrt{(m_1-c_{2m})^2+(n_1-c_{2n})^2+(o_1-c_{2o})^2}=\sqrt{(80-75)^2+(75-88)^2+(75-80)^2}=14,79865$$

$$d(x_2,c_2)=\sqrt{(m_2-c_{2m})^2+(n_2-c_{2n})^2+(o_2-c_{2o})^2}=\sqrt{(90-75)^2+(85-88)^2+(80-80)^2}=15,29706$$

$$d(x_3,c_2)=\sqrt{(m_3-c_{2m})^2+(n_3-c_{2n})^2+(o_3-c_{2o})^2}=\sqrt{(77-75)^2+(78-88)^2+(85-80)^2}=11,35782$$

$$d(x_4,c_2)=\sqrt{(m_4-c_{2m})^2+(n_4-c_{2n})^2+(o_4-c_{2o})^2}=\sqrt{(75-75)^2+(80-88)^2+(70-80)^2}=12,80625$$

$$d(x_5,c_2)=\sqrt{(m_5-c_{2m})^2+(n_5-c_{2n})^2+(o_5-c_{2o})^2}=\sqrt{(89-75)^2+(85-88)^2+(75-80)^2}=15,16575$$

$$d(x_6,c_2)=\sqrt{(m_6-c_{2m})^2+(n_6-c_{2n})^2+(o_6-c_{2o})^2}=\sqrt{(95-75)^2+(70-88)^2+(75-80)^2}=27,36786$$

$$d(x_7,c_2)=\sqrt{(m_7-c_{2m})^2+(n_7-c_{2n})^2+(o_7-c_{2o})^2}=\sqrt{(75-75)^2+(88-88)^2+(80-80)^2}=0$$

$$d(x_8,c_1)=\sqrt{(m_8-c_{2m})^2+(n_8-c_{2n})^2+(o_8-c_{2o})^2}=\sqrt{(80-75)^2+(75-88)^2+(85-80)^2}=14,79865$$

$$d(x_9,c_1)=\sqrt{(m_9-c_{2m})^2+(n_9-c_{2n})^2+(o_9-c_{2o})^2}=\sqrt{(80-75)^2+(80-88)^2+(85-80)^2}=10,67708$$

$$d(x_{10},c_1)=\sqrt{(m_{10}-c_{2m})^2+(n_{10}-c_{2n})^2+(o_{10}-c_{2o})^2}=\sqrt{(78-75)^2+(80-88)^2+(80-80)^2}=8,544004$$

No	Nilai Pemrograman			dc1	dc2	c1	c2
	VB	PHP	JAVA				
1	80	75	75	0	14,79865	*	
2	90	85	80	15	15,29706	*	
3	77	78	85	10,86278	11,35782	*	
4	75	80	70	8,660254	12,80625	*	
5	89	85	75	13,45362	15,16575	*	
6	95	70	75	15,81139	27,36786	*	
7	75	88	80	14,79865	0		*
8	80	75	85	10	14,79865	*	
9	80	80	85	11,18034	10,67708		*
10	78	80	80	7,348469	8,544004	*	

$$C1 = (80+90+77+75+89+95+80+78)/8 = 83$$

$$(75+85+78+80+85+70+75+80)/8 = 78,5$$

$$(75+80+85+70+75+75+85+80)/8 = 78,125$$

$$C2 = (75+80)/2 = 77,5$$

$$(88+80)/2 = 84$$

$$(80+85)/2 = 82,5$$

	VB	PHP	JAVA
c1	83	78,5	78,125
c2	77,5	84	82,5

Tabel. Centroid

2. Perulangan ke-1

Jarak data dengan cluster 1 adalah:

$$\begin{aligned}d(x_1, c_1) &= \sqrt{(m_1 - c_{1m})^2 + (n_1 - c_{1n})^2 + (o_1 - c_{1o})^2} = \sqrt{(80 - 83)^2 + (75 - 78,5)^2 + (75 - 78,125)^2} \\ &= 5,569167\end{aligned}$$

$$\begin{aligned}d(x_2, c_1) &= \sqrt{(m_2 - c_{1m})^2 + (n_2 - c_{1n})^2 + (o_2 - c_{1o})^2} = \sqrt{(90 - 83)^2 + (85 - 78,5)^2 + (80 - 78,125)^2} \\ &= 9,734764\end{aligned}$$

$$\begin{aligned}d(x_3, c_1) &= \sqrt{(m_3 - c_{1m})^2 + (n_3 - c_{1n})^2 + (o_3 - c_{1o})^2} = \sqrt{(77 - 83)^2 + (78 - 78,5)^2 + (85 - 78,125)^2} \\ &= 9,138688\end{aligned}$$

$$\begin{aligned}d(x_4, c_1) &= \sqrt{(m_4 - c_{1m})^2 + (n_4 - c_{1n})^2 + (o_4 - c_{1o})^2} = \sqrt{(75 - 83)^2 + (80 - 78,5)^2 + (70 - 78,125)^2} \\ &= 11,50068\end{aligned}$$

$$\begin{aligned}d(x_5, c_1) &= \sqrt{(m_5 - c_{1m})^2 + (n_5 - c_{1n})^2 + (o_5 - c_{1o})^2} = \sqrt{(89 - 83)^2 + (85 - 78,5)^2 + (75 - 78,125)^2} \\ &= 9,381664\end{aligned}$$

$$\begin{aligned}d(x_6, c_1) &= \sqrt{(m_6 - c_{1m})^2 + (n_6 - c_{1n})^2 + (o_6 - c_{1o})^2} = \sqrt{(95 - 83)^2 + (70 - 78,5)^2 + (75 - 78,125)^2} \\ &= 15,03382\end{aligned}$$

$$\begin{aligned}d(x_7, c_1) &= \sqrt{(m_7 - c_{1m})^2 + (n_7 - c_{1n})^2 + (o_7 - c_{1o})^2} = \sqrt{(75 - 83)^2 + (88 - 78,5)^2 + (80 - 78,125)^2} \\ &= 12,56048\end{aligned}$$

$$\begin{aligned}d(x_8, c_1) &= \sqrt{(m_8 - c_{1m})^2 + (n_8 - c_{1n})^2 + (o_8 - c_{1o})^2} = \sqrt{(80 - 83)^2 + (75 - 78,5)^2 + (85 - 78,125)^2} \\ &= 8,277417\end{aligned}$$

$$\begin{aligned}d(x_9, c_1) &= \sqrt{(m_9 - c_{1m})^2 + (n_9 - c_{1n})^2 + (o_9 - c_{1o})^2} = \sqrt{(80 - 83)^2 + (80 - 78,5)^2 + (85 - 78,125)^2} \\ &= 7,649551\end{aligned}$$

$$\begin{aligned}d(x_{10}, c_1) &= \sqrt{(m_{10} - c_{1m})^2 + (n_{10} - c_{1n})^2 + (o_{10} - c_{1o})^2} = \sqrt{(78 - 83)^2 + (80 - 78,5)^2 + (80 - 78,125)^2} \\ &= 5,546677\end{aligned}$$

Jarak data dengan cluster 2 adalah:

$$\begin{aligned}d(x_1, c_2) &= \sqrt{(m_1 - c_{2m})^2 + (n_1 - c_{2n})^2 + (o_1 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (75 - 84)^2 + (75 - 82,5)^2} \\ &= 11,97915\end{aligned}$$

$$\begin{aligned}d(x_2, c_2) &= \sqrt{(m_2 - c_{2m})^2 + (n_2 - c_{2n})^2 + (o_2 - c_{2o})^2} = \sqrt{(90 - 77,5)^2 + (85 - 84)^2 + (80 - 82,5)^2} \\ &= 12,78671\end{aligned}$$

$$\begin{aligned}d(x_3, c_2) &= \sqrt{(m_3 - c_{2m})^2 + (n_3 - c_{2n})^2 + (o_3 - c_{2o})^2} = \sqrt{(77 - 77,5)^2 + (78 - 84)^2 + (85 - 82,5)^2} \\ &= 6,519202\end{aligned}$$

$$\begin{aligned}d(x_4, c_2) &= \sqrt{(m_4 - c_{2m})^2 + (n_4 - c_{2n})^2 + (o_4 - c_{2o})^2} = \sqrt{(75 - 77,5)^2 + (80 - 84)^2 + (70 - 82,5)^2} \\ &= 13,36039\end{aligned}$$

$$d(x_5, c_2) = \sqrt{(m_5 - c_{2m})^2 + (n_5 - c_{2n})^2 + (o_5 - c_{2o})^2} = \sqrt{(89 - 77,5)^2 + (85 - 84)^2 + (75 - 82,5)^2}$$

$$= 13,7659$$

$$d(x_6, c_2) = \sqrt{(m_6 - c_{2m})^2 + (n_6 - c_{2n})^2 + (o_6 - c_{2o})^2} = \sqrt{(95 - 77,5)^2 + (70 - 84)^2 + (75 - 82,5)^2}$$

$$= 23,6326$$

$$d(x_7, c_2) = \sqrt{(m_7 - c_{2m})^2 + (n_7 - c_{2n})^2 + (o_7 - c_{2o})^2} = \sqrt{(75 - 77,5)^2 + (88 - 84)^2 + (80 - 82,5)^2}$$

$$= 5,338539$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{2m})^2 + (n_8 - c_{2n})^2 + (o_8 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (75 - 84)^2 + (85 - 82,5)^2}$$

$$= 9,66954$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{2m})^2 + (n_9 - c_{2n})^2 + (o_9 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (80 - 84)^2 + (85 - 82,5)^2}$$

$$= 5,338539$$

$$d(x_{10}, c_1) = \sqrt{(m_{10} - c_{2m})^2 + (n_{10} - c_{2n})^2 + (o_{10} - c_{2o})^2} = \sqrt{(78 - 77,5)^2 + (80 - 84)^2 + (80 - 82,5)^2}$$

$$= 4,743416$$

No	Nilai Pemrograman			dc1	dc2	c1	c2
	VB	PHP	JAVA				
1	80	75	75	5,569167	11,97915	*	
2	90	85	80	9,734764	12,78671	*	
3	77	78	85	9,138688	6,519202		*
4	75	80	70	11,50068	13,36039	*	
5	89	85	75	9,381664	13,7659	*	
6	95	70	75	15,03382	23,6326	*	
7	75	88	80	12,56048	5,338539		*
8	80	75	85	8,277417	9,66954	*	
9	80	80	85	7,649551	5,338539		*
10	78	80	80	5,546677	4,743416		*

Hitung rata-rata:

$$C1 = (80+90+75+89+95+80)/6 = 84,83333$$

$$(75+85+80+85+70+75)/6 = 78,33333$$

$$(75+80+70+75+75+85)/6 = 76,66667$$

$$C2 = (77+75+80+78)/4 = 77,5$$

$$(78+88+80+80)/4 = 81,5$$

$$(85+80+85+80)/4 = 82,5$$

	VB	PHP	JAVA
c1	84,83333	78,33333	76,66667
c2	77,5	81,5	82,5

Tabel. Centroid

3. Perulangan 2

Jarak data dengan cluster 1 adalah:

$$\begin{aligned}d(x_1, c_1) &= \sqrt{(m_1 - c_{1m})^2 + (n_1 - c_{1n})^2 + (o_1 - c_{1o})^2} \\ &= \sqrt{(80 - 84,83333)^2 + (75 - 78,33333)^2 + (75 - 76,66667)^2} \\ &= 6,103274\end{aligned}$$

$$\begin{aligned}d(x_2, c_1) &= \sqrt{(m_2 - c_{1m})^2 + (n_2 - c_{1n})^2 + (o_2 - c_{1o})^2} \\ &= \sqrt{(90 - 84,83333)^2 + (85 - 78,33333)^2 + (80 - 76,66667)^2} \\ &= 9,069182\end{aligned}$$

$$\begin{aligned}d(x_3, c_1) &= \sqrt{(m_3 - c_{1m})^2 + (n_3 - c_{1n})^2 + (o_3 - c_{1o})^2} \\ &= \sqrt{(77 - 84,83333)^2 + (78 - 78,33333)^2 + (85 - 76,66667)^2} \\ &= 11,44188\end{aligned}$$

$$\begin{aligned}d(x_4, c_1) &= \sqrt{(m_4 - c_{1m})^2 + (n_4 - c_{1n})^2 + (o_4 - c_{1o})^2} \\ &= \sqrt{(75 - 84,83333)^2 + (80 - 78,33333)^2 + (70 - 76,66667)^2} \\ &= 11,99653\end{aligned}$$

$$\begin{aligned}d(x_5, c_1) &= \sqrt{(m_5 - c_{1m})^2 + (n_5 - c_{1n})^2 + (o_5 - c_{1o})^2} \\ &= \sqrt{(89 - 84,83333)^2 + (85 - 78,33333)^2 + (75 - 76,66667)^2} \\ &= 8,036381\end{aligned}$$

$$\begin{aligned}d(x_6, c_1) &= \sqrt{(m_6 - c_{1m})^2 + (n_6 - c_{1n})^2 + (o_6 - c_{1o})^2} \\ &= \sqrt{(95 - 84,83333)^2 + (70 - 78,33333)^2 + (75 - 76,66667)^2} \\ &= 13,25079\end{aligned}$$

$$\begin{aligned}d(x_7, c_1) &= \sqrt{(m_7 - c_{1m})^2 + (n_7 - c_{1n})^2 + (o_7 - c_{1o})^2} \\ &= \sqrt{(75 - 84,83333)^2 + (88 - 78,33333)^2 + (80 - 76,66667)^2} \\ &= 14,18626\end{aligned}$$

$$\begin{aligned}d(x_8, c_1) &= \sqrt{(m_8 - c_{1m})^2 + (n_8 - c_{1n})^2 + (o_8 - c_{1o})^2} \\ &= \sqrt{(80 - 84,83333)^2 + (75 - 78,33333)^2 + (85 - 76,66667)^2} \\ &= 10,19395\end{aligned}$$

$$\begin{aligned}d(x_9, c_1) &= \sqrt{(m_9 - c_{1m})^2 + (n_9 - c_{1n})^2 + (o_9 - c_{1o})^2} \\ &= \sqrt{(80 - 84,83333)^2 + (80 - 78,33333)^2 + (85 - 76,66667)^2} \\ &= 9,776669\end{aligned}$$

$$\begin{aligned}d(x_{10}, c_1) &= \sqrt{(m_{10} - c_{1m})^2 + (n_{10} - c_{1n})^2 + (o_{10} - c_{1o})^2} \\ &= \sqrt{(78 - 84,83333)^2 + (80 - 78,33333)^2 + (80 - 76,66667)^2} \\ &= 7,783526\end{aligned}$$

Jarak data dengan cluster 2 adalah:

$$d(x_1, c_2) = \sqrt{(m_1 - c_{2m})^2 + (n_1 - c_{2n})^2 + (o_1 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (75 - 81,5)^2 + (75 - 82,5)^2}$$

$$= 10,23474$$

$$d(x_2, c_2) = \sqrt{(m_2 - c_{2m})^2 + (n_2 - c_{2n})^2 + (o_2 - c_{2o})^2} = \sqrt{(90 - 77,5)^2 + (85 - 81,5)^2 + (80 - 82,5)^2}$$

$$= 13,2193$$

$$d(x_3, c_2) = \sqrt{(m_3 - c_{2m})^2 + (n_3 - c_{2n})^2 + (o_3 - c_{2o})^2} = \sqrt{(77 - 77,5)^2 + (78 - 81,5)^2 + (85 - 82,5)^2}$$

$$= 4,330127$$

$$d(x_4, c_2) = \sqrt{(m_4 - c_{2m})^2 + (n_4 - c_{2n})^2 + (o_4 - c_{2o})^2} = \sqrt{(75 - 77,5)^2 + (80 - 81,5)^2 + (70 - 82,5)^2}$$

$$= 12,8355$$

$$d(x_5, c_2) = \sqrt{(m_5 - c_{2m})^2 + (n_5 - c_{2n})^2 + (o_5 - c_{2o})^2} = \sqrt{(89 - 77,5)^2 + (85 - 81,5)^2 + (75 - 82,5)^2}$$

$$= 14,16863$$

$$d(x_6, c_2) = \sqrt{(m_6 - c_{2m})^2 + (n_6 - c_{2n})^2 + (o_6 - c_{2o})^2} = \sqrt{(95 - 77,5)^2 + (70 - 81,5)^2 + (75 - 82,5)^2}$$

$$= 22,24298$$

$$d(x_7, c_2) = \sqrt{(m_7 - c_{2m})^2 + (n_7 - c_{2n})^2 + (o_7 - c_{2o})^2} = \sqrt{(75 - 77,5)^2 + (88 - 81,5)^2 + (80 - 82,5)^2}$$

$$= 7,399324$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{2m})^2 + (n_8 - c_{2n})^2 + (o_8 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (75 - 81,5)^2 + (85 - 82,5)^2}$$

$$= 7,399324$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{2m})^2 + (n_9 - c_{2n})^2 + (o_9 - c_{2o})^2} = \sqrt{(80 - 77,5)^2 + (80 - 81,5)^2 + (85 - 82,5)^2}$$

$$= 3,840573$$

$$d(x_{10}, c_1) = \sqrt{(m_{10} - c_{2m})^2 + (n_{10} - c_{2n})^2 + (o_{10} - c_{2o})^2} = \sqrt{(78 - 77,5)^2 + (80 - 81,5)^2 + (80 - 82,5)^2}$$

$$= 2,95804$$

No	Nilai Pemrograman			dc1	dc2	c1	c2
	VB	PHP	JAVA				
1	80	75	75	6,103274	10,23474	*	
2	90	85	80	9,069182	13,2193	*	
3	77	78	85	11,44188	4,330127		*
4	75	80	70	11,99653	12,8355	*	
5	89	85	75	8,036381	14,16863	*	
6	95	70	75	13,25079	22,24298	*	
7	75	88	80	14,18626	7,399324		*
8	80	75	85	10,19395	7,399324		*
9	80	80	85	9,776669	3,840573		*
10	78	80	80	7,783526	2,95804		*

Hitung rata-rata:

$$\begin{aligned}
 C1 &= (80+90+75+89+95)/5 = 85,8 \\
 &(75+85+80+85+70)/5 = 79 \\
 &(75+80+70+75+75)/5 = 75 \\
 C2 &= (77+75+80+80+78)/5 = 78 \\
 &(78+88+75+80+80)/5 = 80,2 \\
 &(85+80+85+85+80)/5 = 83
 \end{aligned}$$

	VB	PHP	JAVA
c1	85	79	75
c2	78	80,2	83

Tabel. Centroid

4. Perulangan ke-3

Jarak data dengan cluster 1 adalah:

$$\begin{aligned}
 d(x_1, c_1) &= \sqrt{(m_1 - c_{1m})^2 + (n_1 - c_{1n})^2 + (o_1 - c_{1o})^2} = \sqrt{(80 - 85)^2 + (75 - 79)^2 + (75 - 75)^2} \\
 &= 7,045566
 \end{aligned}$$

$$\begin{aligned}
 d(x_2, c_1) &= \sqrt{(m_2 - c_{1m})^2 + (n_2 - c_{1n})^2 + (o_2 - c_{1o})^2} = \sqrt{(90 - 85)^2 + (85 - 79)^2 + (80 - 75)^2} \\
 &= 8,86792
 \end{aligned}$$

$$\begin{aligned}
 d(x_3, c_1) &= \sqrt{(m_3 - c_{1m})^2 + (n_3 - c_{1n})^2 + (o_3 - c_{1o})^2} = \sqrt{(77 - 85)^2 + (78 - 79)^2 + (85 - 75)^2} \\
 &= 13,35814
 \end{aligned}$$

$$\begin{aligned}
 d(x_4, c_1) &= \sqrt{(m_4 - c_{1m})^2 + (n_4 - c_{1n})^2 + (o_4 - c_{1o})^2} = \sqrt{(75 - 85)^2 + (80 - 79)^2 + (70 - 75)^2} \\
 &= 11,9432
 \end{aligned}$$

$$\begin{aligned}
 d(x_5, c_1) &= \sqrt{(m_5 - c_{1m})^2 + (n_5 - c_{1n})^2 + (o_5 - c_{1o})^2} = \sqrt{(89 - 85)^2 + (85 - 79)^2 + (75 - 75)^2} \\
 &= 6,8
 \end{aligned}$$

$$\begin{aligned}
 d(x_6, c_1) &= \sqrt{(m_6 - c_{1m})^2 + (n_6 - c_{1n})^2 + (o_6 - c_{1o})^2} = \sqrt{(95 - 85)^2 + (70 - 79)^2 + (75 - 75)^2} \\
 &= 12,87012
 \end{aligned}$$

$$\begin{aligned}
 d(x_7, c_1) &= \sqrt{(m_7 - c_{1m})^2 + (n_7 - c_{1n})^2 + (o_7 - c_{1o})^2} = \sqrt{(75 - 85)^2 + (88 - 79)^2 + (80 - 75)^2} \\
 &= 14,92113
 \end{aligned}$$

$$\begin{aligned}
 d(x_8, c_1) &= \sqrt{(m_8 - c_{1m})^2 + (n_8 - c_{1n})^2 + (o_8 - c_{1o})^2} = \sqrt{(80 - 85)^2 + (75 - 79)^2 + (85 - 75)^2} \\
 &= 12,23274
 \end{aligned}$$

$$\begin{aligned}
 d(x_9, c_1) &= \sqrt{(m_9 - c_{1m})^2 + (n_9 - c_{1n})^2 + (o_9 - c_{1o})^2} = \sqrt{(80 - 85)^2 + (80 - 79)^2 + (85 - 75)^2} \\
 &= 11,60345
 \end{aligned}$$

$$\begin{aligned}
 d(x_{10}, c_1) &= \sqrt{(m_{10} - c_{1m})^2 + (n_{10} - c_{1n})^2 + (o_{10} - c_{1o})^2} = \sqrt{(78 - 85)^2 + (80 - 79)^2 + (80 - 75)^2} \\
 &= 9,318798
 \end{aligned}$$

Jarak data dengan cluster 2 adalah:

$$d(x_1, c_2) = \sqrt{(m_1 - c_{2m})^2 + (n_1 - c_{2n})^2 + (o_1 - c_{2o})^2} = \sqrt{(80 - 78)^2 + (75 - 80,2)^2 + (75 - 83)^2}$$

$$= 9,748846$$

$$d(x_2, c_2) = \sqrt{(m_2 - c_{2m})^2 + (n_2 - c_{2n})^2 + (o_2 - c_{2o})^2} = \sqrt{(90 - 78)^2 + (85 - 80,2)^2 + (80 - 83)^2}$$

$$= 13,26801$$

$$d(x_3, c_2) = \sqrt{(m_3 - c_{2m})^2 + (n_3 - c_{2n})^2 + (o_3 - c_{2o})^2} = \sqrt{(77 - 78)^2 + (78 - 80,2)^2 + (85 - 83)^2}$$

$$= 3,136877$$

$$d(x_4, c_2) = \sqrt{(m_4 - c_{2m})^2 + (n_4 - c_{2n})^2 + (o_4 - c_{2o})^2} = \sqrt{(75 - 78)^2 + (80 - 80,2)^2 + (70 - 83)^2}$$

$$= 13,34316$$

$$d(x_5, c_2) = \sqrt{(m_5 - c_{2m})^2 + (n_5 - c_{2n})^2 + (o_5 - c_{2o})^2} = \sqrt{(89 - 78)^2 + (85 - 80,2)^2 + (75 - 83)^2}$$

$$= 14,42359$$

$$d(x_6, c_2) = \sqrt{(m_6 - c_{2m})^2 + (n_6 - c_{2n})^2 + (o_6 - c_{2o})^2} = \sqrt{(95 - 78)^2 + (70 - 80,2)^2 + (75 - 83)^2}$$

$$= 21,37849$$

$$d(x_7, c_2) = \sqrt{(m_7 - c_{2m})^2 + (n_7 - c_{2n})^2 + (o_7 - c_{2o})^2} = \sqrt{(75 - 78)^2 + (88 - 80,2)^2 + (80 - 83)^2}$$

$$= 8,879189$$

$$d(x_8, c_1) = \sqrt{(m_8 - c_{2m})^2 + (n_8 - c_{2n})^2 + (o_8 - c_{2o})^2} = \sqrt{(80 - 78)^2 + (75 - 80,2)^2 + (85 - 83)^2}$$

$$= 5,919459$$

$$d(x_9, c_1) = \sqrt{(m_9 - c_{2m})^2 + (n_9 - c_{2n})^2 + (o_9 - c_{2o})^2} = \sqrt{(80 - 78)^2 + (80 - 80,2)^2 + (85 - 83)^2}$$

$$= 2,835489$$

$$d(x_{10}, c_1) = \sqrt{(m_{10} - c_{2m})^2 + (n_{10} - c_{2n})^2 + (o_{10} - c_{2o})^2} = \sqrt{(78 - 78)^2 + (80 - 80,2)^2 + (80 - 83)^2}$$

$$= 3,006659$$

No	Nilai Pemrograman			dc1	dc2	c1	c2
	VB	PHP	JAVA				
1	80	75	75	7,045566	9,748846	*	
2	90	85	80	8,86792	13,26801	*	
3	77	78	85	13,35814	3,136877		*
4	75	80	70	11,9432	13,34316	*	
5	89	85	75	6,8	14,42359	*	
6	95	70	75	12,87012	21,37849	*	
7	75	88	80	14,92113	8,879189		*
8	80	75	85	12,23274	5,919459		*
9	80	80	85	11,60345	2,835489		*
10	78	80	80	9,318798	3,006659		*

Hitung rata-rata:

$$C1 = (80+90+75+89+95)/5 = 85,8$$

$$(75+85+80+85+70)/5 = 79$$

$$(75+80+70+75+75)/5 = 75$$

$$C2 = (77+75+80+80+78)/5 = 78$$

$$(78+88+75+80+80)/5 = 80,2$$

$$(85+80+85+85+80)/5 = 83$$

	VB	PHP	JAVA
c1	85	79	75
c2	78	80,2	83

Tabel. Centroid

5. Perulangan ke-4

Karena hasil centroid sama dengan centroid sebelumnya ,Iterasi berhenti.