

NAMA: RADEN ISNAWAN ARGY ARYASATYA  
NIM: 195410257  
KELAS: IF-5  
MATKUL: DATA MINING

---

#### UJIAN AKHIR SEMESTER

#### 1. Terdapat data sebagai berikut :

Pelanggan	Kartu	Panggilan	Blok	Bonus
1	Prabayar	Sedikit	Sedang	Tidak
2	Pascabayar	Banyak	Sedang	Ya
3	Prabayar	Banyak	Sedang	Ya
4	Prabayar	Banyak	Rendah	Tidak
5	Pascabayar	Cukup	Tinggi	Ya
6	Pascabayar	Cukup	Sedang	Ya
7	Prabayar	Cukup	Sedang	Ya
8	Prabayar	Cukup	Rendah	Tidak
9	Pascabayar	Sedikit	Tinggi	Ya
10	Pascabayar	Banyak	Tinggi	Ya
11	Pascabayar	Sedikit	Rendah	Ya

Hitung probabilitas bersyarat untuk setiap kelas  $P(X|C_i), i=1,2$  (menggunakan algoritma Naïve Bayes).

- Jika terdapat seorang pelanggan yang memiliki nilai atribut Kartu='Pascabayar', Panggilan='Banyak', dan Blok='Rendah', layak mendapatkan bonus atau tidak (Bonus=Ya)?
- Jika terdapat seorang pelanggan yang memiliki nilai atribut Kartu='Prabayar', Panggilan='Cukup', dan Blok='Tinggi', layak mendapatkan bonus atau tidak (Bonus=Ya)?

JAWAB:

- Jika terdapat seorang pelanggan yang memiliki nilai atribut Kartu='Pascabayar', Panggilan='Banyak', dan Blok='Rendah', layak mendapatkan bonus atau tidak (Bonus=Ya)?

*Tahap 1: menghitung jumlah class/label*

- $P(Y=YA) = 8/11$  (Jumlah data "YA" pada kolom "BONUS" dibagi jumlah data). Probabilitas prior untuk kelas pertama  $C_1$ , yaitu Bonus=Ya, adalah  $P(C_1) = 8/11 = 0,7273$ .
- $P(Y=TIDAK) = 3/11$  (Jumlah data "TIDAK" pada kolom "BONUS" dibagi dengan semua jumlah data). Probabilitas prior untuk kelas pertama  $C_2$ , yaitu Bonus=Tidak, adalah  $P(C_2) = 3/11 = 0,2727$ .

*Tahap 2: menghitung jumlah class/label*

- $P(KARTU = PASACABAYAR | Y= YA) = 6/8$   
(jumlah data kartu "PASCABAYAR" dengan bonus "YA" dibagi jumlah data YA)
- $P(KARTU = PASACABAYAR | Y= TIDAK) = 0/3$   
(jumlah data kartu "PASCABAYAR" dengan bonus "TIDAK" dibagi jumlah data TIDAK)

- $P(\text{PANGGILAN} = \text{BANYAK} \mid Y = \text{YA}) = 3/8$   
(jumlah data panggilan "BANYAK" dengan bonus "YA" dibagi jumlah data YA)
- $P(\text{PANGGILAN} = \text{BANYAK} \mid Y = \text{TIDAK}) = 1/3$   
(jumlah data panggilan "BANYAK" dengan bonus "TIDAK" dibagi jumlah data TIDAK)
- $P(\text{BLOK} = \text{RENDAH} \mid Y = \text{YA}) = 1/8$   
(jumlah data blok "RENDAH" dengan bonus "YA" dibagi jumlah data YA)
- $P(\text{BLOK} = \text{RENDAH} \mid Y = \text{TIDAK}) = 2/3$   
(jumlah data blok "RENDAH" dengan bonus "TIDAK" dibagi jumlah data TIDAK)

*Tahap 3: kalikan semua hasil variabel YA dan variabel TIDAK*

- YA:  

$$(P(\text{KARTU} = \text{PASCABAYAR} \mid Y = \text{YA}) * P(\text{PANGGILAN} = \text{BANYAK} \mid Y = \text{YA}) * P(\text{BLOK} = \text{RENDAH} \mid Y = \text{YA}))$$

$$= 6/8 * 3/8 * 1/8$$

$$= 0,03515625$$
- TIDAK:  

$$(P(\text{KARTU} = \text{PASCABAYAR} \mid Y = \text{TIDAK}) * P(\text{PANGGILAN} = \text{BANYAK} \mid Y = \text{TIDAK}) * P(\text{BLOK} = \text{RENDAH} \mid Y = \text{TIDAK}))$$

$$= 0/3 * 1/3 * 2/3$$

$$= 0$$

*Tahap 4: Probabilitas posterior bahwa tuple X berada di kelas Ci didapat dengan memaksimalkan  $P(X|Ci)P(Ci)$ , untuk  $i=1,2$ . Dengan kata lain, kalikan Bonus YA dan TIDAK dengan probabilitas prior YA dan TIDAK*

- YA  

$$(P(\text{BONUS} = \text{"YA"}) * (\text{JUMLAH YA})) = 0,03515625 * 0,7273 = \mathbf{0,02556914062}$$
- TIDAK  

$$(P(\text{BONUS} = \text{"TIDAK"}) * (\text{JUMLAH TIDAK})) = 0 * 0,2727 = \mathbf{0}$$

*Tahap 5: Kesimpulan*

$P(X \mid \text{Bonus} = \text{Ya}) \times P(\text{Bonus} = \text{Ya}) = 0,03515625 \times 0,7273 = \mathbf{0,02556914062}$  lebih besar dibanding  $P(X \mid \text{Bonus} = \text{Tidak}) \times P(\text{Bonus} = \text{Tidak}) = 0 \times 0,2727 = \mathbf{0}$ . Maka, Naïve Bayesian Classifier mengklasifikasikan pelanggan dengan Kartu='Pascabayar', Panggilan='Banyak', Blok='Rendah' ke dalam kelas **BONUS=YA**

- b. Jika terdapat seorang pelanggan yang memiliki nilai atribut Kartu='Prabayar', Panggilan='Cukup', dan Blok='Tinggi', layak mendapatkan bonus atau tidak (Bonus=Ya)?**

*Tahap 1: menghitung jumlah class/label*

- $P(Y = \text{YA}) = 8/11$  (Jumlah data "YA" pada kolom "BONUS" dibagi jumlah data). Probabilitas prior untuk kelas pertama C1, yaitu Bonus=Ya, adalah  $P(Ci) = 8/11 = 0,7273$ .
- $P(Y = \text{TIDAK}) = 3/11$  (Jumlah data "TIDAK" pada kolom "BONUS" dibagi dengan semua jumlah data). Probabilitas prior untuk kelas pertama C2, yaitu Bonus=Tidak, adalah  $P(Ci) = 3/11 = 0,2727$ .

Tahap 2: menghitung jumlah class/label

- $P(\text{KARTU} = \text{PRABAYAR} \mid Y = \text{YA}) = 2/8$   
(jumlah data kartu "PRABAYAR" dengan bonus "YA" dibagi jumlah data YA)
- $P(\text{KARTU} = \text{PRABAYAR} \mid Y = \text{TIDAK}) = 3/3$   
(jumlah data kartu "PRABAYAR" dengan bonus "TIDAK" dibagi jumlah data TIDAK)
- $P(\text{PANGGILAN} = \text{CUKUP} \mid Y = \text{YA}) = 3/8$   
(jumlah data panggilan "CUKUP" dengan bonus "YA" dibagi jumlah data YA)
- $P(\text{PANGGILAN} = \text{CUKUP} \mid Y = \text{TIDAK}) = 1/3$   
(jumlah data panggilan "CUKUP" dengan bonus "TIDAK" dibagi jumlah data TIDAK)
- $P(\text{BLOK} = \text{TINGGI} \mid Y = \text{YA}) = 3/8$   
(jumlah data blok "TINGGI" dengan bonus "YA" dibagi jumlah data YA)
- $P(\text{BLOK} = \text{TINGGI} \mid Y = \text{TIDAK}) = 0/3$   
(jumlah data blok "TINGGI" dengan bonus "TIDAK" dibagi jumlah data TIDAK)

Tahap 3: kalikan semua hasil variabel YA dan variabel TIDAK

- YA:  

$$(P(\text{KARTU} = \text{PRABAYAR} \mid Y = \text{YA}) * P(\text{PANGGILAN} = \text{CUKUP} \mid Y = \text{YA}) * P(\text{BLOK} = \text{TINGGI} \mid Y = \text{YA}))$$

$$= 2/8 * 3/8 * 3/8$$

$$= \mathbf{0.03515625}$$
- TIDAK:  

$$(P(\text{KARTU} = \text{PRABAYAR} \mid Y = \text{TIDAK}) * P(\text{PANGGILAN} = \text{CUKUP} \mid Y = \text{TIDAK}) * P(\text{BLOK} = \text{TINGGI} \mid Y = \text{TIDAK}))$$

$$= 3/3 * 1/3 * 0/3$$

$$= \mathbf{0}$$

Tahap 4: Probabilitas posterior bahwa tuple  $X$  berada di kelas  $C_i$  didapat dengan memaksimalkan  $P(X|C_i)P(C_i)$ , untuk  $i=1,2$ . Dengan kata lain, kalikan Bonus YA dan TIDAK dengan probabilitas prior YA dan TIDAK

- YA  

$$(P(\text{BONUS} = \text{"YA"}) * (\text{JUMLAH YA})) = 0,03515625 * 0,7273 = \mathbf{0,02556914062}$$
- TIDAK  

$$(P(\text{BONUS} = \text{"TIDAK"}) * (\text{JUMLAH TIDAK})) = 0 * 0,2727 = \mathbf{0}$$

Tahap 5: Kesimpulan

$P(X \mid \text{Bonus} = \text{Ya}) \times P(\text{Bonus} = \text{Ya}) = 0,03515625 \times 0,7273 = \mathbf{0,02556914062}$  lebih besar dibanding  $P(X \mid \text{Bonus} = \text{Tidak}) \times P(\text{Bonus} = \text{Tidak}) = 0 \times 0,2727 = \mathbf{0}$ . Maka, Naïve Bayesian Classifier mengklasifikasikan pelanggan dengan Kartu='Prabayar', Panggilan='Cukup', Blok='Tinggi' ke dalam kelas **BONUS=YA**

## 2. Perhatikan data berikut.

Objek	Nilai	Kehadiran
-------	-------	-----------

Mahasiswa 1	90	24
Mahasiswa 2	75	21
Mahasiswa 3	50	17
Mahasiswa 4	80	16
Mahasiswa 5	90	24
Mahasiswa 6	75	23
Mahasiswa 7	85	22
Mahasiswa 8	70	20
Mahasiswa 9	55	19
Mahasiswa 10	55	18

Buatlah pengelompokan menggunakan :

- a. Algoritma FCM, dengan ketentuan :

Jumlah cluster = 2

Pangkat = 2

Maksimum iterasi = 5

Error terkecil yang diharapkan = 0.1

Fungsi obyektif awal =  $P_0 = 0$

Iterasi awal =  $t = 0$

- b. Algoritma K-means dengan jumlah cluster = 2.

JAWAB :

**a. Algoritma FCM**

Langkah-langkah:

- 1) Jumlah data sebanyak 10 mahasiswa dengan data ke- $i$  ( $i = 1, 2, \dots, 10$ ). Dari kriteria yang ditentukan dibuat suatu tingkat kepentingan kriteria berdasarkan nilai bobot yang telah ditentukan dalam bilangan fuzzy.

Rating kecocokan setiap alternatif pada setiap kriteria sebagai berikut:

Rendah (R) = 0

Cukup (C) = 0.5

Tinggi (T) = 1

Nilai bobot dari setiap variabel ditentukan dalam bilangan fuzzy. Hasil perhitungan dilakukan berdasarkan langkah-langkah pada algoritma FCM. Langkah yang pertama yaitu memasukkan data yang akan di-cluster ke dalam matriks  $X$  sebagai berikut:

1	0,5
1	0,5
1	0,5
1	1
1	0
1	0,5
1	0
1	0
1	1
0	0,5

- 2) Bangkitkan bilangan random untuk matrik  $U$  dengan ukuran  $10 \times 2$ . Matriks ini menunjukkan derajat keanggotaan data ke- $i = 1,2,3,4,5,6,7,8,9,10$  pada cluster ke  $k = 1,2$ .

Pertama, bangkitkan matriks  $U$  dengan komponen  $\mu_{ik}$ ,  $i = 14$ ;  $k = 2$ , nilai  $\mu_{ik}$  ditentukan secara acak dengan syarat jumlah nilai elemen matriks dari kolom dalam setiap baris harus 1. Matriks  $U$  yang terbentuk (secara random) adalah sebagai berikut:

$U =$   
 $\begin{matrix} 0,3 & 0,7 \\ 0,6 & 0,4 \\ 0,7 & 0,3 \\ 0,4 & 0,6 \\ 0,8 & 0,2 \\ 0,3 & 0,7 \\ 0,6 & 0,4 \\ 0,7 & 0,3 \\ 0,4 & 0,6 \\ 0,8 & 0,2 \end{matrix}$

Menghitung pusat cluster:

Data (x)	$u_{ij}$		$x_{ij}$		$(\mu_{i1})^2$	$(\mu_{i2})^2$	$(\mu_{i1})^2 * x_{i1}$	$(\mu_{i1})^2 * x_{i2}$	$(\mu_{i2})^2 * x_{i1}$	$(\mu_{i2})^2 * x_{i2}$
	1	2	1	2						
1	0,3	0,7	90	24	0,09	0,49	8,1	2,16	44,1	11,76
2	0,6	0,4	75	21	0,36	0,16	27	7,56	12	3,36
3	0,7	0,3	50	17	0,49	0,09	24,5	8,33	4,5	1,53
4	0,4	0,6	80	16	0,16	0,36	12,6	2,56	28,8	5,76
5	0,8	0,2	90	24	0,64	0,04	57,6	15,36	3,6	0,96
6	0,3	0,7	75	23	0,09	0,49	6,75	2,07	36,75	11,27
7	0,6	0,4	85	22	0,36	0,16	30,6	7,92	13,6	3,52
8	0,7	0,3	70	20	0,49	0,09	34,3	9,8	6,3	1,8
9	0,4	0,6	55	19	0,16	0,36	8,8	3,04	19,8	6,84
10	0,8	0,2	55	18	0,64	0,04	35,2	11,52	2,2	0,72
$\Sigma$					3,48	2,28	245,45	70,32	171,65	47,16

Penjelasan:

Pada kolom  $(\mu_{i1})^2$  dan  $(\mu_{i2})^2$ , nilai hasil didapatkan dari hasil pangkat dari matriks  $U$  pada setiap kolomnya dan akan dipakai pada perkalian dengan matriks  $X$ . Nilai dari  $\Sigma$  adalah hasil penjumlahan dari setiap kolomnya dan akan dipakai untuk menghitung hasil pusat cluster 1 dan 2.

Kolom  $(\mu_{i1})^2 * x_{i1}$  dan  $(\mu_{i2})^2 * x_{i2}$  diperoleh dari hasil perkalian antara kolom  $(\mu_{i1})^2$  dan  $(\mu_{i2})^2$  dengan setiap kolom pada matriks  $X$  dan nilai dari  $\Sigma$  adalah hasil penjumlahan untuk setiap kolomnya dan nilai-nilai dari  $\Sigma$  tersebut akan dipakai dalam menghitung hasil pusat cluster 1

- 3) Hasil pusat cluster iterasi 1:

$V_{kj}$	1	2
1	70,532	20,207
2	75,286	20,685

Penjelasan:

Hasil untuk cluster 1 pada baris pertama diperoleh dari baris pada hasil penjumlahan setiap  $\Sigma$  pada kolom  $(\mu_1)^2 * x_{i1}$  dibagi dengan nilai  $\Sigma$  pada kolom  $\mu_1$  ( $245,5/3,48 = 70,532$ ) dan ( $70,32/3,48 = 20,207$ )

Hasil untuk cluster 2 pada baris kedua diperoleh dari baris pada hasil penjumlahan setiap  $\Sigma$  pada kolom  $(\mu_i)^2 * x_{i2}$  dibagi dengan nilai  $\Sigma$  pada kolom  $\mu_i$  ( $171,65/2,28 = 75,286$ ) dan ( $47,16/2,28 = 20,685$ )

1 0,5  
 1 0,5  
 1 0,5  
 1 1  
 1 0  
 1 0,5  
 1 0  
 1 0  
 1 1  
 0 0,5

#### 4) Hasil Perhitungan C1

Data (x)	C1		$\Sigma C1$	$\Sigma C1 * (\mu_1)^2$
	$(X_{i1} - V_{i1})^2$	$(X_{i1} - V_{i1})^2$		
1	4809,69	388,365	5198,055	
2	4809,69	388,365	5198,055	
3	4809,69	388,365	5198,055	
4	4809,69	368,908		
5	4809,69	408,323		
6	4809,69	388,365	5198,055	
7	4809,69	408,323		
8	4809,69	408,323		
9	4809,69	368,908		
10	4949,40	388,365		

Mohon maaf ya Bu.. saya Algoritma FCM tidak selesai karena kehabisan waktu, saya telah mengerjakan semampu saya sesuai dengan kemampuan saya saat ini. Mohon tetap dinilai ya bu.. Saya berjanji akan belajar di bidang ini dengan lebih giat karena saya berminat untuk berkarir di bidang data science. Terima kasih Bu..

#### b. Algoritma K-means dengan jumlah cluster = 2.

Diketahui:

Objek	Nilai	Kehadiran
Mahasiswa 1	90	24
Mahasiswa 2	75	21
Mahasiswa 3	50	17
Mahasiswa 4	80	16
Mahasiswa 5	90	24
Mahasiswa 6	75	23
Mahasiswa 7	85	22
Mahasiswa 8	70	20
Mahasiswa 9	55	19
Mahasiswa 10	55	18

Langkah-langkah:

1) Tahap pertama:

ambil data ke-1 dan ke-2 sebagai perhitungan pertama. Kita menggunakan rumus Euclidean Distance untuk mendapatkan jarak minimum data terhadap centroid.

Cluster	Nilai	Kehadiran
K1	90	24
K2	75	21

Kemudian hitung centroid pertama. Tentukan jarak dari data 1 ke data 1, data 1 ke data 2, data 2 ke data 1 dan data 2 ke data 2.

- Cluster 1 ke cluster 1 (90, 24)  

$$\sqrt{(90 - 90)^2 + (24 - 24)^2} = 0$$
- Jarak dari cluster 2 ke cluster 1 (75, 21) <-> (90, 24) =  

$$\sqrt{(75 - 90)^2 + (21 - 24)^2} = \sqrt{(-15)^2 + (-3)^2} = \sqrt{225 + 9} = \sqrt{234} = 15.3$$
- Jarak dari cluster 1 ke cluster 2 (90, 24) <-> (75, 21) =  

$$\sqrt{(90 - 75)^2 + (24 - 21)^2} = \sqrt{(15)^2 + (3)^2} = \sqrt{225 + 9} = \sqrt{234} = 15.3$$
- Cluster 2 ke cluster 2 (75, 21)  

$$\sqrt{(75 - 75)^2 + (21 - 21)^2} = 0$$

Dari perhitungan di atas, didapatkan hasil seperti berikut:

Cluster	Centroid		Kelompok Cluster
	Nilai	Kehadiran	
K1 (90, 24)	0	15.3	1
K2 (75, 21)	15.3	0	2

Penjelasan:

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. Selanjutnya kita beralih ke perhitungan kedua untuk data ke-3.

2) Tahap kedua:

Setelah mendapatkan centroid, kita beralih ke data ke-3 yaitu (50, 17). Pertama kita hitung jarak data ke-3 terhadap centroid 1 dan centroid 2 supaya hasilnya nanti adalah data 3 akan masuk ke cluster K1 atau K2.

- Jarak dari data 3 ke cluster 1 (50, 17) <-> (90, 24) =  

$$\sqrt{(50 - 90)^2 + (17 - 24)^2} = \sqrt{(-40)^2 + (-7)^2} = \sqrt{1600 + 49} = \sqrt{1649} = 40.6$$
- Jarak dari data 3 ke cluster 2 (50, 17) <-> (75, 21) =  

$$\sqrt{(50 - 75)^2 + (17 - 21)^2} = \sqrt{(-25)^2 + (-4)^2} = \sqrt{625 + 16} = \sqrt{641} = 25.3$$

Hasil perhitungan Euclidean Distance kita masukkan dalam tabel berikut.

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(50, 17)	40.6	25.3	2

Penjelasan:

Kenapa data ke-3 masuk kelompok cluster 2? Karena jika dilihat dari jarak data terhadap centroid, K2 lebih dekat dengan data ke-3 dengan nilai 25.3 (jaraknya lebih pendek dibanding K1 yaitu 40.6)

- UPDATE nilai Centroid.  
 $X\_centroid\_baru = (x\_K2 + x\_data3)/2$   
 $Y\_centroid\_baru = (y\_K2 + y\_data3)/2$

Perhitungan centroid baru:

Cluster	Nilai	Kehadiran
K1	90	24
K2	$(75 + 50)/2$	$(21 + 17)/2$

Hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	90	24
K2	62.5	19

### 3) Tahap ketiga:

Kita lanjutkan perhitungan ke data ke-4 yaitu (80, 16). Kita hitung jarak antar dataset dan Centroid K1, sama seperti langkah di tahap kedua.

- Jarak dari data 4 ke cluster 1 (80, 16) <-> (90, 24) =  
 $\sqrt{(80 - 90)^2 + (16 - 24)^2} = \sqrt{(-10)^2 + (-8)^2} = \sqrt{100 + 64} = \sqrt{164}$   
 $= 12.8$
- Jarak dari data 4 ke cluster 2 (80, 16) <-> (62.5, 19) =  
 (karena K2 telah diupdate, jadi kita harus menggunakan nilai centroid K2 yang baru yaitu : (62.5, 19).  
 $\sqrt{(80 - 62.5)^2 + (16 - 19)^2} = \sqrt{(17.5)^2 + (-3)^2} = \sqrt{306.25 + 9} =$   
 $\sqrt{315.25}$   
 $= 17.76$

Dari kedua perhitungan jarak dataset ke K1 dan K2, didapatkan hasil berikut :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(80, 16)	12.8	17.78	1

Penjelasan

data ke-4 masuk kelompok cluster 1 karena jika dilihat dari jarak data terhadap centroid, data ke-4 lebih dekat dengan K1 dengan nilai 12.8 (jaraknya lebih pendek dibanding K2 yaitu 17.78)

perhitungan Culster Centroid baru:



Cluster	Nilai	Kehadiran
K1	$(90 + 80) / 2$	$(24 + 16) / 2$
K2	62.5	19

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	85	20
K2	62.5	19

4) Tahap keempat:

Kita lanjutkan perhitungan ke data ke-5 yaitu (90, 24).

- Jarak dari data 5 ke cluster 1 (90, 24)  $\leftrightarrow$  (85, 20) =  

$$\sqrt{(90 - 85)^2 + (24 - 20)^2} = \sqrt{(5)^2 + (4)^2} = \sqrt{25 + 16} = \sqrt{41}$$

$$= 6.4$$
- Jarak dari data 5 ke cluster 2 (90, 24)  $\leftrightarrow$  (62.5, 19) =  

$$\sqrt{(90 - 62.5)^2 + (24 - 19)^2} = \sqrt{(27.5)^2 + (5)^2} = \sqrt{756.25 + 25} =$$

$$\sqrt{781.25}$$

$$= 27.95$$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(90, 24)	6.4	27.95	1

data ke-5 masuk kelompok cluster 1 karena jika dilihat dari jarak data terhadap centroid, data ke-5 lebih dekat dengan K1 dengan nilai 6.4

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	$(85 + 90) / 2$	$(20 + 24) / 2$
K2	62.5	19

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	87.5	22
K2	62.5	19

5) Tahap kelima:

Lanjutkan perhitungan ke data ke-6 yaitu (75, 23).

- Jarak dari data 6 ke cluster 1 (75, 23)  $\leftrightarrow$  (87.5, 22) =  

$$\sqrt{(75 - 87.5)^2 + (23 - 22)^2} = \sqrt{(-12.5)^2 + (1)^2} = \sqrt{156.25 + 1} =$$

$$\sqrt{157.25}$$

$$= 12.53$$
- Jarak dari data 6 ke cluster 2 (75, 23)  $\leftrightarrow$  (62.5, 19) =

$$\sqrt{(75 - 62.5)^2 + (23 - 19)^2} = \sqrt{(12.5)^2 + (4)^2} = \sqrt{156.25 + 16} = \sqrt{172.25} = 13.12$$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(75, 23)	12.53	13.12	1

data ke-6 masuk kelompok cluster 1 karena jika dilihat dari jarak data terhadap centroid, data ke-6 lebih dekat dengan K1 dengan nilai 12.53

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	$(87.5 + 75) / 2$	$(22 + 23) / 2$
K2	62.5	19

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	81.25	22.5
K2	62.5	19

6) Tahap keenam:

Lanjutkan perhitungan ke data ke-7 yaitu (85, 22).

- Jarak dari data 7 ke cluster 1  $(85, 22) \leftrightarrow (81.25, 22.5) =$   
 $\sqrt{(85 - 81.25)^2 + (22 - 22.5)^2} = \sqrt{(3.75)^2 + (-0.5)^2} = \sqrt{14.06 + 0.25}$   
 $= \sqrt{14.31}$   
 $= 3.78$
- Jarak dari data 7 ke cluster 2  $(85, 22) \leftrightarrow (62.5, 19) =$   
 $\sqrt{(85 - 62.5)^2 + (22 - 19)^2} = \sqrt{(22.5)^2 + (3)^2} = \sqrt{506.25 + 9} =$   
 $\sqrt{515.25}$   
 $= 22.7$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(85, 22)	3.78	22.7	1

data ke-7 masuk kelompok cluster 1 karena jika dilihat dari jarak data terhadap centroid, data ke-7 lebih dekat dengan K1 dengan nilai 3.78

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	$(81.25 + 85) / 2$	$(22.25 + 22) / 2$
K2	62.5	19

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
---------	-------	-----------

K1	83.125	22.25
K2	62.5	19

7) Tahap ketujuh:

Lanjutkan perhitungan ke data ke-8 yaitu (70, 20).

- Jarak dari data 8 ke cluster 1 (70, 20)  $\leftrightarrow$  (83.125, 22.25) =  

$$\sqrt{(70 - 83.125)^2 + (20 - 22.25)^2} = \sqrt{(-13.125)^2 + (-2.25)^2} =$$

$$\sqrt{172.27 + 5.06} = \sqrt{177.33}$$

$$= 13.32$$
- Jarak dari data 8 ke cluster 2 (70, 20)  $\leftrightarrow$  (62.5, 19) =  

$$\sqrt{(70 - 62.5)^2 + (20 - 19)^2} = \sqrt{(7.5)^2 + (1)^2} = \sqrt{56.25 + 1} = \sqrt{57.25}$$

$$= 7.57$$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(70, 20)	13.32	7.57	2

data ke-8 masuk kelompok cluster 2 karena jika dilihat dari jarak data terhadap centroid, data ke-8 lebih dekat dengan K2 dengan nilai 7.57

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	$(62.5 + 70)/2$	$(19 + 20)/2$

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	66.25	19.5

8) Tahap kedelapan:

Lanjutkan perhitungan ke data ke-9 yaitu (55, 19).

- Jarak dari data 9 ke cluster 1 (55, 19)  $\leftrightarrow$  (83.125, 22.25) =  

$$\sqrt{(55 - 83.125)^2 + (19 - 22.25)^2} = \sqrt{(-28.125)^2 + (-3.25)^2} =$$

$$\sqrt{791.01 + 10.56} = \sqrt{801.57}$$

$$= 28.31$$
- Jarak dari data 9 ke cluster 2 (55, 19)  $\leftrightarrow$  (66.25, 19.5) =  

$$\sqrt{(55 - 66.25)^2 + (19 - 19.5)^2} = \sqrt{(-11.25)^2 + (-0.5)^2} =$$

$$\sqrt{126.5 + 0.25} = \sqrt{126.75}$$

$$= 11.25$$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	

(55, 19)	28.31	11.25	2
----------	-------	-------	---

data ke-9 masuk kelompok cluster 2 karena jika dilihat dari jarak data terhadap centroid, data ke-9 lebih dekat dengan K2 dengan nilai 11.25

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	$(66.25 + 55)/2$	$(19.5 + 19)/2$

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	60.625	19.25

9) Tahap kesembilan:

Lanjutkan perhitungan ke data ke-10 yaitu (55, 18).

- Jarak dari data 9 ke cluster 1 (55, 18)  $\leftrightarrow$  (83.125, 22.25) =  
 $\sqrt{(55 - 83.125)^2 + (18 - 22.25)^2} = \sqrt{(-28.125)^2 + (-4.25)^2} =$   
 $\sqrt{791.01 + 18.06} = \sqrt{809,07}$   
 $= 28.44$
- Jarak dari data 9 ke cluster 2 (55, 18)  $\leftrightarrow$  (60.625, 19.25) =  
 $\sqrt{(55 - 60.625)^2 + (18 - 19.25)^2} = \sqrt{(-5.625)^2 + (-1.25)^2} =$   
 $\sqrt{31.64 + 1.56} = \sqrt{33,2}$   
 $= 5.76$

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(55, 18)	28.44	5.76	2

data ke-10 masuk kelompok cluster 2 karena jika dilihat dari jarak data terhadap centroid, data ke-10 lebih dekat dengan K2 dengan nilai 5.76

perhitungan Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	$(60.625 + 55)/2$	$(19.25 + 18)/2$

hasil Culster Centroid baru:

Cluster	Nilai	Kehadiran
K1	83.125	22.25
K2	57.81	18.625

10) Tahap kesepuluh:

Buat tabel untuk menentukan kluster setiap objek/data dari perhitungan-perhitungan yang sudah dikerjakan tadi.

<b>Objek</b>	<b>Nilai</b>	<b>Kehadiran</b>	<b>Kluster</b>
Mahasiswa 1	90	24	1
Mahasiswa 2	75	21	2
Mahasiswa 3	50	17	2
Mahasiswa 4	80	16	1
Mahasiswa 5	90	24	1
Mahasiswa 6	75	23	1
Mahasiswa 7	85	22	1
Mahasiswa 8	70	20	2
Mahasiswa 9	55	19	2
Mahasiswa 10	55	18	2

**Terima Kasih**