

Clustering di Data Mining

Contoh Perhitungan K-Means

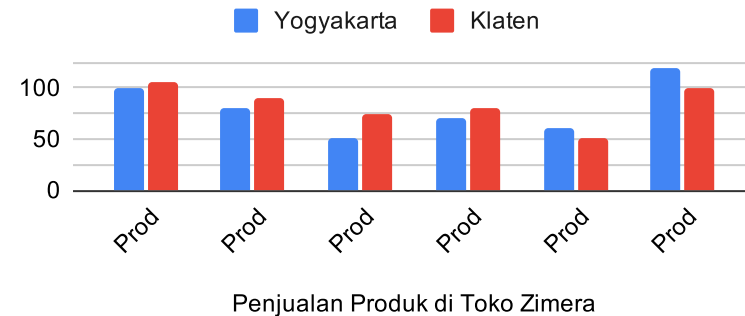
Dr. Bambang Purnomosidi D. P.



Penjualan Produk di Toko Zimera

	Yogyakarta	Klaten
Produk A	100	105
Produk B	80	90
Produk C	50	75
Produk D	70	80
Produk E	60	50
Produk F	120	100

Yogyakarta dan Klaten



Langkah Pertama

Tetapkan K	2	
K1	100	105 => Produk A
K2	70	80 => Produk D

Jumlah (K) cluster ditetapkan dengan cara yang sembarangan tapi make sense. Misal, data penjualan bisa dibuat 2 cluster: laku dan cukup laku
 Penentuan titik awal untuk centroid (K1 dan K2) juga sembarangan tetapi masuk akal.

Rumus Euclidian untuk menghitung jarak:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}$$

Langkah Kedua

Menghitung jarak antar cluster

Jarak K1 ke K1	K1 (100, 105)	akar dari (100-100) kuadrat + (105-105) kuadrat	0
Jarak K2 ke K1	K2 (70, 80)	akar dari (70-100) kuadrat + (80-105) kuadrat	39,05124838

	Jarak K2 ke K2	K2 (70, 80)	akar dari (70-70) kuadrat + (80-80) kuadrat		0
	Jarak K1 ke K2	K1 (100, 105)	akar dari (100-70) kuadrat + (105-80) kuadrat		39,05124838
Update Tabel					
	Produk	Centroid		Kelompok Cluster	
		K1	K2		
	Produk A	0	39,05	1	
	Produk D	39,05	0	2	
Langkah Ketiga:					
Iterasi Setiap Dataset	Menghitung jarak ke masing-masing centroid (K1 dan K2)				
Data Produk B	(80, 90)				
	Jarak ke K1	B (80, 90)	akar dari (80-100) kuadrat + (90-105) kuadrat		25
	Jarak ke K2	B (80, 90)	akar dari (80-70) kuadrat + (90-80) kuadrat		14,14213562
Update Tabel					
	Produk	Centroid		Kelompok Cluster	
		K1	K2		
	Produk A	0	39,05	1	
	Produk D	39,05	0	2	
	Produk B	25	14,14	2	=> diambil jarak terdekat
	Update Centroid K2				
			K2 (70, 80)	(70+80)/2	(80+90)/2
			Produk B (80, 90)	75	85
	Jadi, cluster centroid baru =				
		K1	100	105	
		K2	75	85	=> akan digunakan untuk
					perhitungan berikutnya
Data Produk C	(50, 75)				
	Jarak ke K1	C (50, 75)	akar dari (50-100) kuadrat + (75-105) kuadrat		58,30951895

	Jarak ke K2	C (50, 75)	akar dari (50-75) kuadrat + (75-85) kuadrat		26,92582404
Update Tabel					
		Centroid			
	Produk	K1	K2	Kelompok Cluster	
	Produk A	0	39,05	1	
	Produk D	39,05	0	2	
	Produk B	25	22,36	2	
	Produk C	58,3	26,92	2	=> diambil jarak terdekat
		Update Centroid K2			
			K2 (75, 85)	(75+50)/2	(85+75)/2
			Produk B (50,75)	62,5	80
	Jadi, cluster centroid baru =				
		K1	100	105	
		K2	62,5	80	=> akan digunakan untuk
					perhitungan berikutnya
Data Produk E	(60, 50)				
	Jarak ke K1	E (60, 50)	akar dari (60-100) kuadrat + (50-105) kuadrat		68,00735254
	Jarak ke K2	E (60, 50)	akar dari (50-62,5) kuadrat + (50-80) kuadrat		32,5
Update Tabel					
		Centroid			
	Produk	K1	K2	Kelompok Cluster	
	Produk A	0	39,05	1	
	Produk D	39,05	0	2	
	Produk B	25	22,36	2	
	Produk C	58,3	26,92	2	
	Produk E	68	32,5	2	=> diambil jarak terdekat
		Update Centroid K2			
			K2 (62,5, 80)	(62,5+60)/2	(80+50)/2
			Produk E (60, 50)	61,25	65

	Jadi, cluster centroid baru =						
		K1	100	105			
		K2	61,25	65	=> akan digunakan untuk		
					perhitungan berikutnya		
Data Produk F	(120, 100)						
	Jarak ke K1	F (120, 100)	akar dari (120-100) kuadrat + (100-105) kuadrat			20,61552813	
	Jarak ke K2	F (120, 100)	akar dari (120-61,25) kuadrat + (100, 65) kuadrat			68,38539683	
Update Tabel							
	Produk	Centroid		Kelompok Cluster			
		K1	K2				
		Produk A	0		39,05	1	
		Produk D	39,05		0	2	
		Produk B	25		22,36	2	
		Produk C	58,3		26,92	2	
		Produk E	68		32,5	2	
		Produk F	20,62		68,39	1	=> diambil jarak terdekat
			Update Centroid K1				
			K1 (100, 105)	(100+120)/2	(105+100)/2		
			Produk F (120, 100)	110	102,5		
	Jadi, cluster centroid baru =						
		K1	110	102,5	=> akan digunakan untuk		
		K2	61,25	65	perhitungan berikutnya		