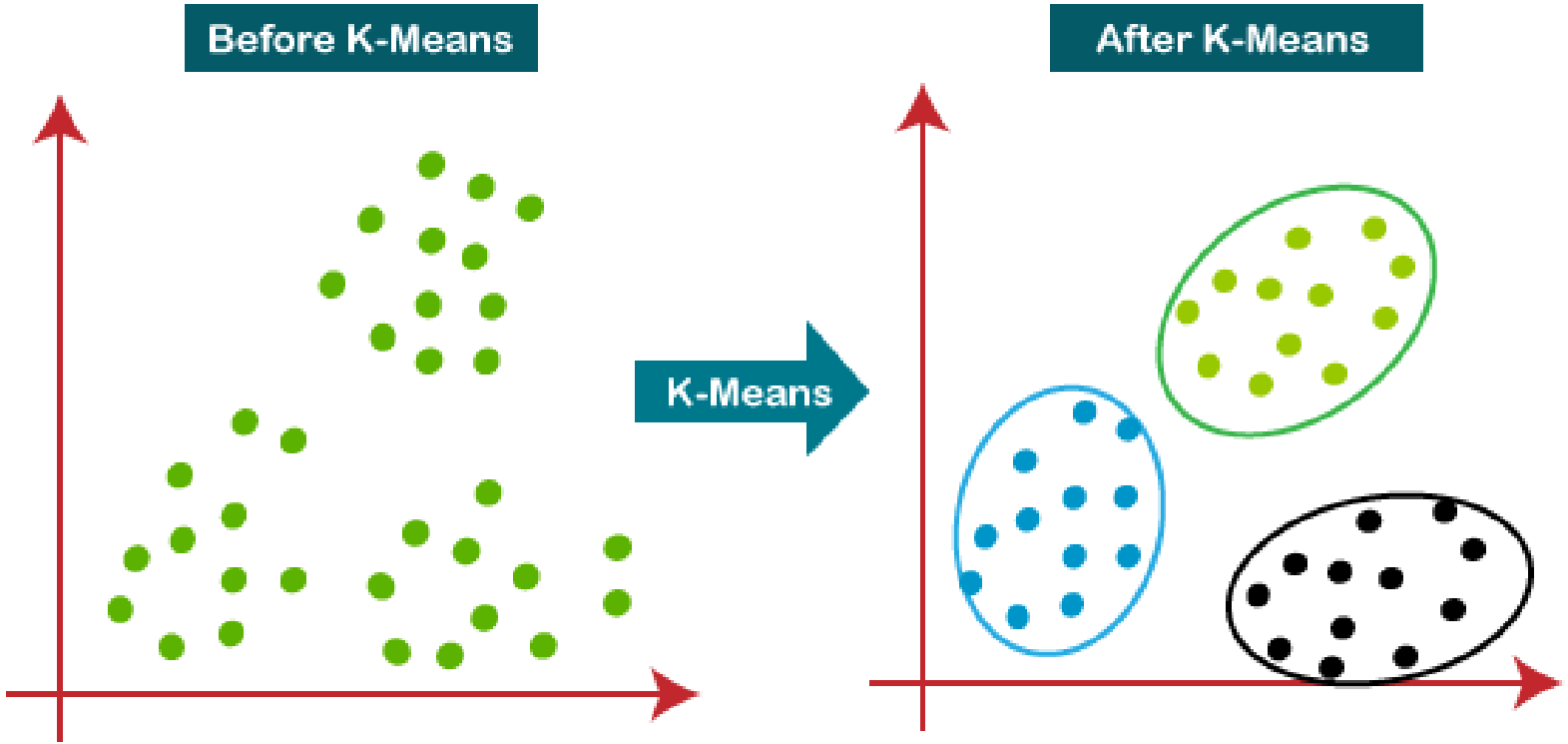
Course Data Sains untuk Bisnis dan Perkantoran

Analisis Cluster K-Means

Bobby Poerwanto, S.Pd., M.Si
Dosen Prodi Statistika FMIPA UNM

Kredensial Mikro Mahasiswa Indonesia
Universitas Negeri Makassar

Ilustrasi



Contoh Permasalahan

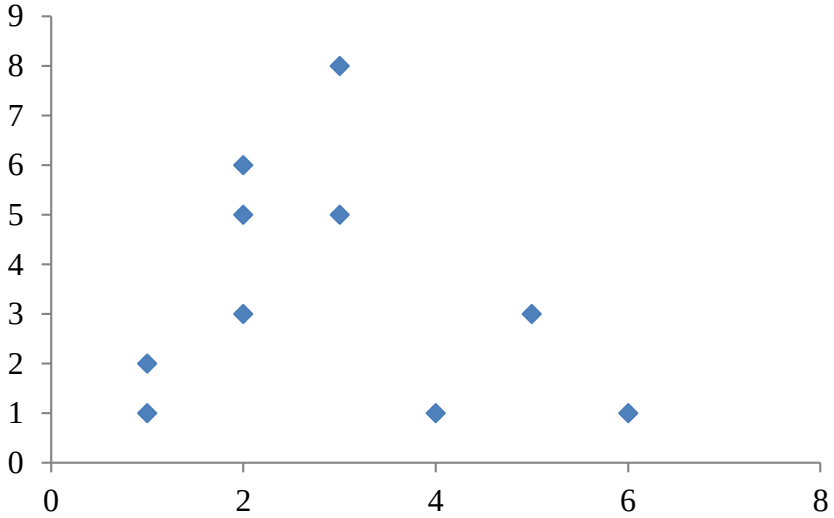
Seorang peneliti ingin mengelompokkan wilayah menjadi 3 kategori yaitu wilayah yang produktif, sedang, dan kurang produktif berdasarkan luas lahan (Ha) dan produksi padi (ton) sehingga dapat membantu pemerintah dalam menentukan secara proporsional bantuan dan tindakan yang akan diberikan pada daerah tersebut

Seorang guru ingin mengelompokkan siswa di kelas XI IPA 2 berdasarkan hasil belajar matematika, biologi, fisika, dan kimia. Pengelompokan ini bertujuan agar memudahkan guru dalam memberikan perlakuan ke siswa-siswa yang cepat menyerap materi dan yang lambat.

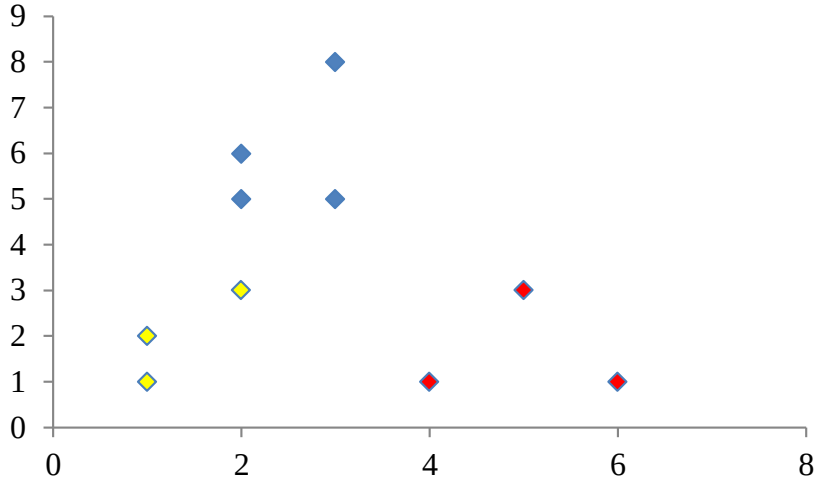
Analisis Cluster K-Means

Analisis Cluster K-Means adalah teknik statistika yang berguna untuk mengelompokkan objek ke dalam K cluster yang telah ditentukan di awal dimana setiap objek:

1. Mempunyai tingkat homogenitas yang tinggi **dalam** satu kelompok
2. Mempunyai tingkat heterogenitas yang tinggi **antar** kelompok



Gambar 1. Plot Awal



Gambar 2. Plot Setelah Clustering

Algoritma

1. Tentukan nilai K sebanyak jumlah cluster atau kelompok yang diinginkan
2. Pilih sebanyak K data dari set data sebagai pusat cluster (*centroid*) secara random
3. Menghitung jarak antara objek dengan masing-masing *centroid*. Bisa menggunakan rumus *Euclidean Distance*

$$d\left(x_i, x_j\right) = \sqrt{\left(\left|x_{i1} - x_{j1}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 + \cdots + \left|x_{ip} - x_{jp}\right|^2\right)}$$

4. Mengelompokkan objek berdasarkan jarak terdekat dengan *centroid*
5. Menentukan *centroid* baru dengan menggunakan rumus

$$C_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_{i(q)}$$

6. Ulangi langkah 3 dan 4 hingga tidak ada lagi objek yang berpindah cluster

Contoh Kasus

Tabel di bawah berisi tentang indeks tingkat kemiskinan 10 wilayah. Seorang peneliti ingin mengelompokkan wilayah tersebut menjadi 3 kelompok berdasarkan indeks kedalaman kemiskinan (X) dan indeks keparahan (Y). Sebagai seorang data scientist, kalian diminta untuk membantu analisis dengan menggunakan analisis cluster K-Means.

Wilayah	X	Y
A	1	1
B	4	1
C	6	1
D	1	2
E	2	3

Wilayah	X	Y
F	5	3
G	2	5
H	3	5
I	2	6
J	3	8