**Muh. Sulthan Nazhim**

**H071191047**

1. **Clustering (Pengelompokan)**

Pengelompokan data variable dimana data ini menjadi beberapa kelompok sedemikian rupa sehingga titik data dalam kelompok sama lebih mirip dengan titik data lain dalam kelompok yang sama dan tidak mirip dengan titik pada kelompok lain tersebut .

Metode metode Clustering :

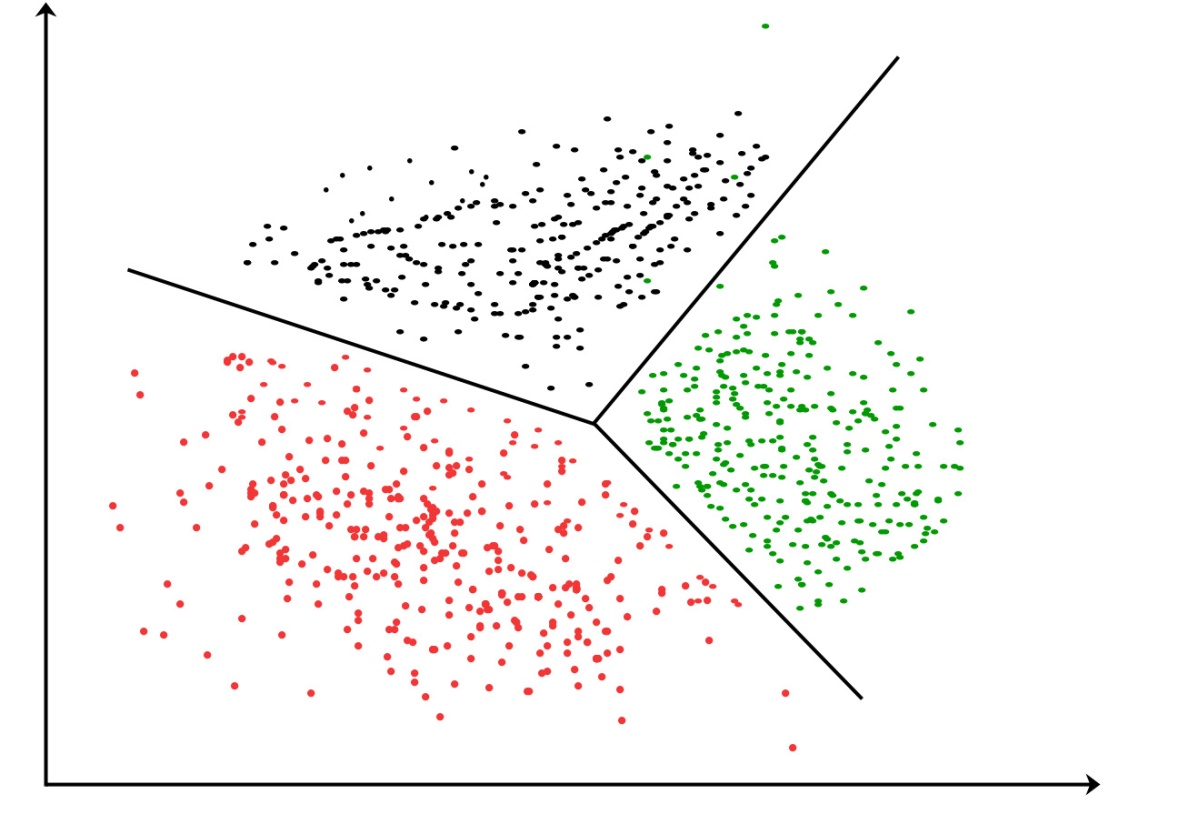
* + Metode Berbasis Kepadatan (Density-Based Method) : Metode ini anggap sebagai wilayah padat yang memiliki kesamaan dan berbeda dengan wilayah rapat dalam bawah ruangan.
  + Hirerarchical Based-Method(Metode Berbasis Hiralki) : Cluster yang terbentuk dalam sebuah metode yang membentuk struktur tipe pohon berdasarkan hiralki. Cluster baru dibentuk berdasarkan kelompok menjadi 2 Sebelumnya terbentuk :
    - Agglomerative (Pendekatan dari bawah ke atas)
    - Divisive(Pendekatan dari atas ke bawah)

Contoh CURE, BIRCH

* + Partitioning Methods (Metode Partisi) : Metode Partisi ini mengobjek menjadi *k* cluster dan setiap partisi dalam membentuk satu cluster. Fungsi dari metode partisi ini digunakan untuk mengoptimalkan pada fungsi kesamaan kriteria bersifat objek tif seperti ketika kita jarak adalah parameter utama itu contohnya *k*-means, CLARANS (Clustering Large Applications based on Randomized Search) dll.

* + Grid Based Methods (Metode Berbasis Grid) : Metode Grid ini memiliki ruang data yang dapat diformalisasikan menjadi sejumlah sel yang membentuk struktur seperti grid. Semua operasi pengelompokan yang dilakukan dengan pada kisi-kisi ini cepat dan tidak bergantung pada jumlah objek data-data . contoh STING.

Algoritma Clustering *k* means ini Pembelajaran tanpa pengawasan paling sederhana untuk memecahkan permasalahan data Pengelompokan. Algoritmat *k-means* mempartisi *n* pengamatan menjadi *k*-*cluster* dimana dalam setiap pengamatan milik cluster dengan means terdekat sehingga berfungsi dengan sebagai prototipe Cluster.



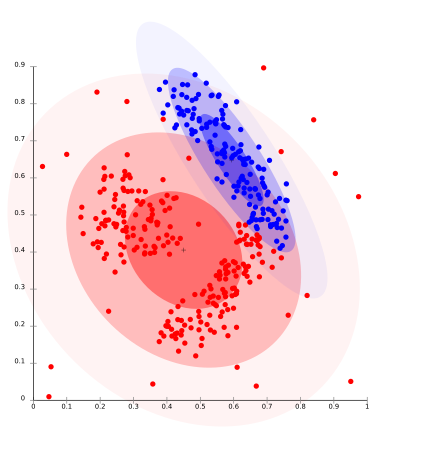
Aplikasi Aplikasi Pada Algoritma Clustering di bidang misalnya :

* Pemasaran : Digunakan Untuk mengkarakteristik dan menemukan segmen pada pelanggan untuk tujuan pemasaran.
* Biologi (Biology): Untuk mengklasifikasi berbagai spesies tumbuhan dan hewan dan berbagai aspek aspek dalam biologi .
* Perpustakaan (Library) : Untuk mengelompokan buku-buku yang ada kategori dalam series beserta berbeda dalam setiap buku berdasarkan topik dan informasi.
* Asuransi : Untuk Mengakui pelanggan, Kebajikan mereka dan Mengidentifikasi Penipuan.
* Perencanaan Kota (City Planning) : Untuk membuat kelompok rumah dan mempelajari nilainya berdasarkan lokasi geografis dan faktor dari lain.
* Studi Gempa (Studies Earthquake) : Untuk mempelajari daerah yang terkena gempa kita dapat menentukan zona berbahaya .

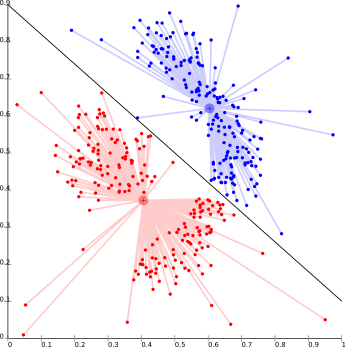
1. **Jenis Jenis Algoritma Clustering**

* Metode Berbasis Distribusi (Distribution based Methods): model pengelompokan dimana menyesuaikan data dengan probabilitas bahwa data tersebut akan termasuk dalam distribusi yang sama. Pengelompokan bisa dilakukan dengan *normal* atau *gaussian*. Distribusi Gaussian Lebih menonjong pada jumlah distribusi data akan tetap sama.

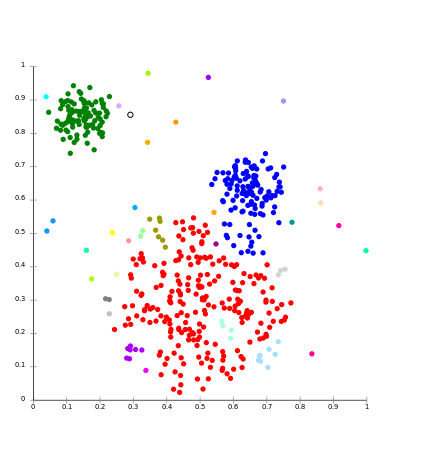
Contoh Gambar Metode Berbasis Distribusi :



* Metode Berbasis Centroid : jarak titik minimum dengan pusat. Algoritma ini digunakan Dengan Clustering interaktif dimana cluster ini dibentuk dengan kedekatan titik data ke *centroid* cluster. Masalah Terbesar dalam algoritma adalah menentukan *k* terlebih dahulu. Ini harus juga mengelompokan dalam distribusi berbasis kepadatan.



* Connectivity Based Methods (Metode Berbasis Koknetivitas) :  
  Intinya dari metode ini mirip dengan Metode Berbasis Centroid pada dasarnya mendefinisikan cluster berdasarkan kedekatan titik data. Fungsi jarak ini bersifat subjektif.



* Model Density (Density Models) : Model ini dilakukan dengan pencarian ruang data daerah-daerah dengan kepadatan yang bervariasi dari titik-titik data dalam ruang data tersebut. Khusus Ini EX-DBSCAN dan OPTIK

