

Nama : Akmal Zuhdy Prasetya  
NIM : H071191035  
Kelas : Machine Learning

### **Resume Pertemuan 3**

#### **A. Apa itu Clustering?**

Pada dasarnya clustering adalah jenis metode pembelajaran tanpa pengawasan (unsupervised), yaitu metode di mana kita menarik referensi dari kumpulan data yang terdiri dari data input tanpa tanggapan berlabel. Sedangkan clustering adalah tugas membagi populasi atau titik data menjadi beberapa kelompok sedemikian rupa sehingga titik data dalam kelompok yang sama lebih mirip dengan titik data lain dalam kelompok yang sama dan tidak mirip dengan titik data pada kelompok lain. Pada dasarnya clustering adalah proses pengumpulan objek berdasarkan kesamaan dan ketidaksamaan di antara mereka.

#### **B. Tujuan Clustering**

Clustering sangat penting karena proses ini bertujuan untuk menentukan pengelompokan intrinsik di antara data yang tidak berlabel yang ada. Tidak ada kriteria untuk clustering yang baik. Itu tergantung pada penggunaannya masing-masing, kriteria apa yang dapat mereka gunakan yang memenuhi kebutuhan mereka. Misalnya, kita mungkin tertarik untuk menemukan perwakilan untuk kelompok homogen (reduksi data), dalam menemukan "klaster alami" dan menggambarkan sifat mereka yang tidak diketahui (tipe data "alami"), dalam menemukan pengelompokan yang berguna dan cocok (kelas data "berguna") atau dalam menemukan objek data yang tidak biasa (outlier detection). Algoritma ini harus membuat beberapa asumsi yang merupakan kesamaan poin dan setiap asumsi membuat cluster yang berbeda dan sama-sama valid.

#### **C. Metode Clustering**

Berikut adalah beberapa metode yang digunakan dalam Clustering.

##### **1. Partitioning**

Metode ini mempartisi objek menjadi sebanyak  $k$  cluster dan setiap partisi membentuk satu cluster. Metode ini digunakan untuk mengoptimalkan fungsi kesamaan kriteria objektif seperti ketika jarak adalah parameter utama, contohnya

adalah K-means serta CLARANS (Clustering Large Applications based on Randomized Search).

## 2. Grid-Based

Dalam metode ini ruang data dirumuskan menjadi sejumlah sel terbatas yang membentuk struktur seperti grid. Semua operasi pengelompokan yang dilakukan pada grid-grid ini cepat dan tidak bergantung pada jumlah objek data, contohnya adalah STING (Statistical Information Grid), wave cluster, dan CLIQUE (Clustering In Quest).

## 3. Density-Based

Metode ini menganggap cluster sebagai wilayah padat yang memiliki beberapa kesamaan dan berbeda dari wilayah padat bawah ruang. Metode ini memiliki akurasi dan kemampuan yang baik untuk menggabungkan dua cluster. Contohnya adalah DBSCAN (Density-Based Spatial Clustering of Applications with Noise) dan OPTICS (Ordering Points to Mengidentifikasi Clustering Structure) dll.

## 4. Hierarchical Based

Cluster yang terbentuk dalam metode ini membentuk struktur tipe pohon yang terbagi berdasarkan hierarki. Cluster baru dibentuk menggunakan cluster yang sebelumnya terbentuk. Cluster ini dibagi menjadi dua kategori yaitu Agglomerative (pendekatan dari bawah ke atas) dan Divisive (pendekatan dari atas ke bawah). Contohnya adalah CURE (Clustering Using Representatives) dan BIRCH (Balanced Iterative Reducing Clustering and using Hierarchies).

## D. Algoritma Clustering

Salah satu algoritma yang cukup sederhana adalah algoritma clustering K-means. K-means adalah algoritma unsupervised yang mempartisi  $n$  pengamatan ke dalam  $k$  cluster di mana setiap pengamatan milik cluster dengan mean terdekat berfungsi sebagai prototipe cluster.

## E. Aplikasi Clustering

Adapun beberapa pengaplikasian proses clustering dalam beberapa bidang adalah sebagai berikut.

1. Pemasaran, dapat digunakan untuk mengkarakterisasi & menemukan segmen pelanggan untuk tujuan pemasaran.

2. Biologi, dapat digunakan untuk klasifikasi di antara berbagai spesies tumbuhan dan hewan.
3. Perpustakaan, digunakan untuk mengelompokkan buku yang berbeda berdasarkan topik dan informasi.
4. Perencanaan Kota, digunakan untuk membuat kelompok rumah dan mempelajari nilainya berdasarkan lokasi geografis dan faktor lain yang ada.
5. Studi gempa, dengan mempelajari daerah yang terkena gempa kita dapat menentukan zona berbahaya.