

1. Jelaskan kelebihan dan kekurangan *k-Means Clustering* dengan contoh kasus dan ilustrasi yang representatif. Gunakan bahasa Anda sendiri. (CLO1: 5%).

Ada beberapa kelebihan k-means, yaitu

- Mudah untuk di implementasikan
  - o Dibandingkan dengan algoritma clustering lainnya, algoritma k-means clustering adalah yang paling mudah
- Waktu *learning* relatif cepat
  - o Karena algoritma k-means clustering adalah *unsupervised learning*, maka proses learning dilakukan lebih cepat karena tidak ada data yang dibandingkan
- Sangat umum digunakan
  - o Umum digunakan dan banyak tutorial tentang algoritma kmeans di internet

Sedangkan kekurangannya adalah sebagai berikut :

- Hasil bisa jadi tidak optimal
  - o Hasil tidak optimal karena nilai *K* diacak dalam setiap *running* algoritma
- Tidak optimal digunakan dengan data yang besar
  - o Algoritma k-means akan optimal untuk jumlah data yang lebih kecil dibanding data yang besar ukurannya

Sumber: <https://garudacyber.co.id/artikel/1514-kelebihan-dan-kekurangan-k-means-clustering>  
<https://brotodata.com/2018/06/28/penjelasan-k-means-clustering/>

2. Jelaskan konsep dasar *Agglomerative Hierarchical Clustering* dengan contoh kasus dan ilustrasi yang representatif. Gunakan bahasa Anda sendiri. (CLO1: 5%).

Algoritma Hierarchical Clustering memiliki 2 macam pengelompokan, yaitu Agglomerative dan Devisive

Pada masing masing algoritma juga terdapat 3 macam perbandingan nilai untuk melakukan pengelompokan

- Single linkage (perbandingan nilai minimum)
- Average linkage (perbandingan nilai rata rata)
- Complete linkage (perbandingan nilai maksimum)

Misalnya menggunakan complete linkage, maka berikut adalah algoritma yang perlu dilakukan:

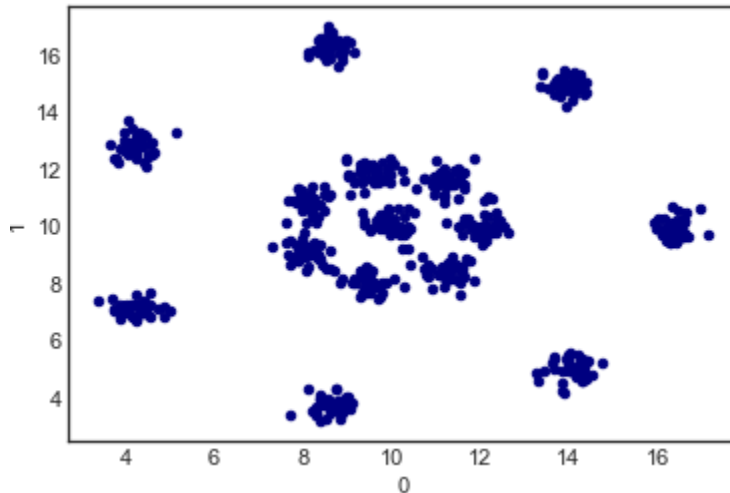
1. Melakukan pemilihan jarak paling minimum (untuk pertama kali lakukan pencarian minimum)
2. Gabung kedua kelompok dan hitung jarak antara kelompok gabungan dengan kelompok lain
  - a. Misal jarak minimum pertama adalah kelompok 1-3 maka gabung menjadi kelompok 13 dan hitung dengan kelompok lain misalnya kelompok 13 dengan kelompok 2, kelompok 13 dengan 4 dan lainnya.
3. Melakukan pemilihan jarak maksimum dari hasil perhitungan jarak untuk menjadi jarak baru pada tabel yang baru
4. Membuat tabel baru berdasarkan jarak yang baru
5. Melakukan langkah 1-4 hingga tidak ada lagi kelompok yang bisa digabungkan

Source : <https://informatikalogi.com/algoritma-hierarchical-clustering/>

## Laporan Self Organizing Map

### 1. Analisis Masalah

Diberikan 600 data berisi titik koordinat, berikut adalah *scatter plot* nya



Dari 600 data di atas, kita bisa lihat bahwa ada beberapa kelompok data (cluster) namun data tersebut belum dikelompokkan, oleh karena itu perlu dilakukan pengelompokkan dari setiap data agar data dapat dikelompokkan dengan benar.

Algoritma SOM di implementasikan dengan Bahasa python dengan beberapa library seperti numpy, pandas, matplotlib serta seaborn

### 2. Sistem yang dibangun

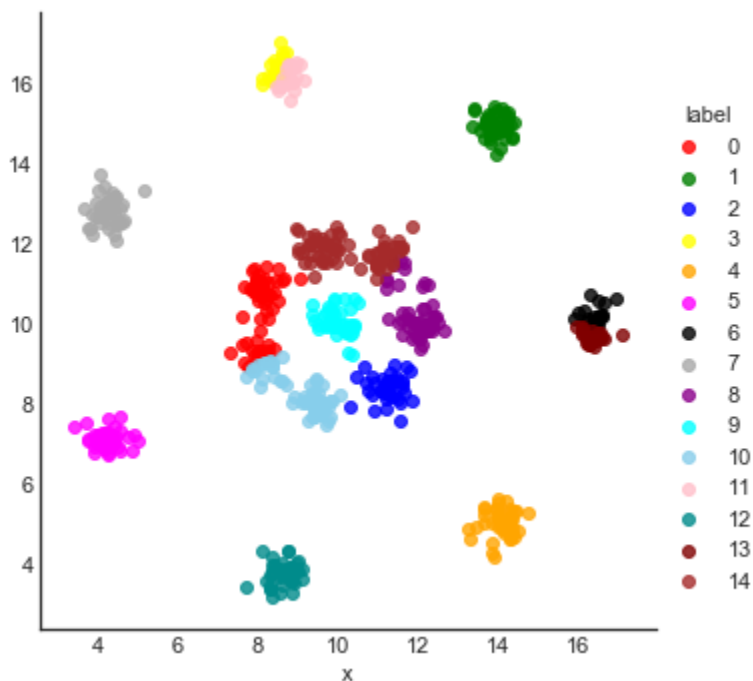
Algoritma SOM merupakan algoritma *unsupervised learning* dimana tidak dibutuhkan data train untuk menjalankannya. Berikut adalah susunan algoritmanya secara umum :

- Membuat titik weight (centroid) (x,y) sejumlah cluster yang kita inginkan
- Set learning rate
- Set epoch
- Set tau
- Melakukan perulangan sebanyak epoch
  - o Melakukan perulangan sebanyak jumlah data
    - Melakukan perulangan sebanyak jumlah cluster

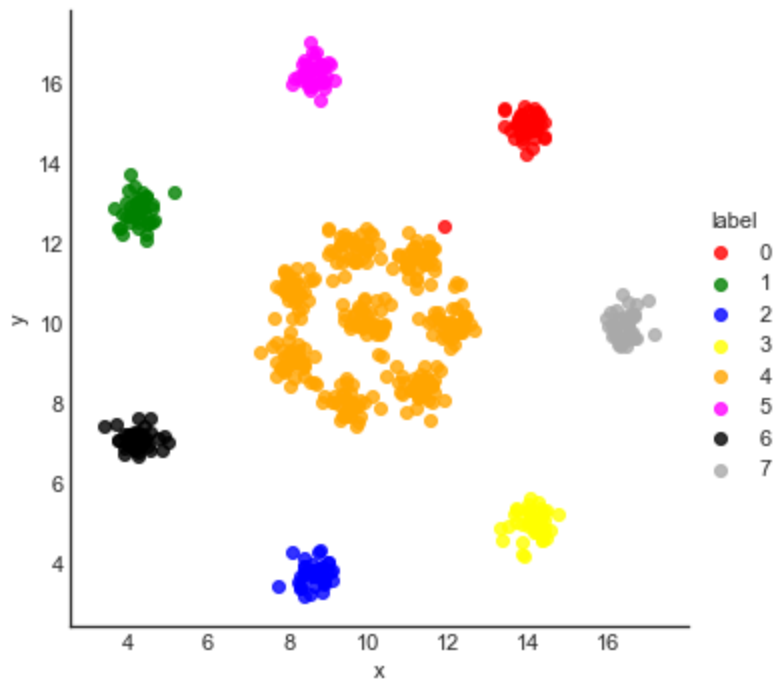
- Melakukan perhitungan jarak data ke masing masing weight
- Membandingkan semua jarak dan mencari jarak minimum
  - Meng-*update* weight yang menang (jarak minimum)
  - Mengelompokkan data kepada cluster yang menang
- Melakukan update learning rate

### 3. Hasil Pengujian,

- Terdapat 2 macam hasil yang bisa kita lihat optimal, yaitu dengan set 8 titik weight ataupun 15 titik weight



- 
- Di atas adalah scatter plot dari 15 cluster, mengapa tidak terlihat rapi ? karena setiap weight adalah acak, jadi nilai weight bisa berada dimana saja dan tidak beraturan, bila dilakukan *running* kembali, bisa saja ditemukan hasil yang optimum



- Berikut adalah hasil running dari 8 cluster, sama seperti 15 cluster, hasil random weight adalah acak dan tidak bisa diprediksi.