**TUGAS 02**

**MACHINE LEARNING**

**Sebuah gambar berisi teks

Deskripsi dibuat secara otomatis**

**SITI NISRINA NABILAH PUTERI SULFI YUDO**

**H071191001**

**Program Studi Sistem Informasi**

**Departemen Matematika**

**Fakultas Matematika dan IlmuPengetahuan Alam**

**Universitas Hasanuddin**

**2021**

**Jenis-jenis Metode Machine Learning**

* Supervised learning (ada kelas pada data)
  + Classification (Klasifikasi)
  + Regression (Regresi)
* Unsupervised learning (tidak ada kelas pada data)
  + Clustering (Klasterisasi)

**Clustering**

Clustering termasuk dalam unsupervised learning dimana datanya tidak memiliki label. Algoritma clustering ini bekerja dengan mengelompokkan objek-objek berdasarkan polanya, entitasnya, kejadiannya, hasil observasinya ke dalam sebuah kelompok atau dengan kata lain melakukan pemecahan atau pemisahan data ke dalam sejumlah kelompok menurut karakter tertentu. Clustering bisa diterapkan di berbagai bidang contohnya pada marketing untuk membantu pihak pemasaran menentukan grup dan membuat program khusus grup tersebut, pada bidang asuransi untuk mengidentifikasi grup yang memiliki tingkat claim tinggi, pada bidang biologi untuk taksonomi makhluk hidup, pada bidang ekonomi untuk riset pasar, dan lain sebagainya.

**Algoritma Clustering**

Partisi : *K-means, K-medoids*

* + - * *K-means*: setiap cluster dipresentasikan oleh pusat cluster (centroid)
      * *K-medoids* atau PAM (*Partition Around Medoids*): setiap cluster dipresentasikan oleh suatu objek dalam cluster (medoid)
      * Partisi himpunan ***D*** sebanyak ***n*** ke ***k*** clusters, yang memiliki jarak terdekat (dimana adalah centroid atau medoid dari cluster ). Partisi kedua algoritma menggunakan metode jarak terdekat dengan menggunakan Euclidean Distance:

Algoritma *K-Means*

1. Pilih jumlah cluster *k*. Inisiasi pusat cluster (centroid). Biasanya dilakukan random
2. Tempatkan setiap data ke cluster terdekat. Penentuan cluster berdasarkan jarak data paling dekat ke cluster mana
3. Hitung kembali pusat cluster. Bisa menggunakan mean dari semua anggota dalam setiap cluster
4. Ulangi langkah 2 dan 3 memakai pusat cluster yang baru hingga tidak ada lagi perubahan cluster

Algoritma *K-Medoids*

1. K-means sensitive dengan data pencilan atau outliers. Solusinya menggunakan *k-metoids*
2. Langkah menentukan anggota cluster sama. Letak perbedaanya pada cluster. Pusat cluster k-medoid merupakan salah satu objek data dalam cluster yang berada di tengah.

Hirarki : *Agnes, Diana*

1. Menggunakan jarak matriks sebagai kriteria melakukan clustering. Tidak membutuhkan input *k* sebagai banyaknya cluster.

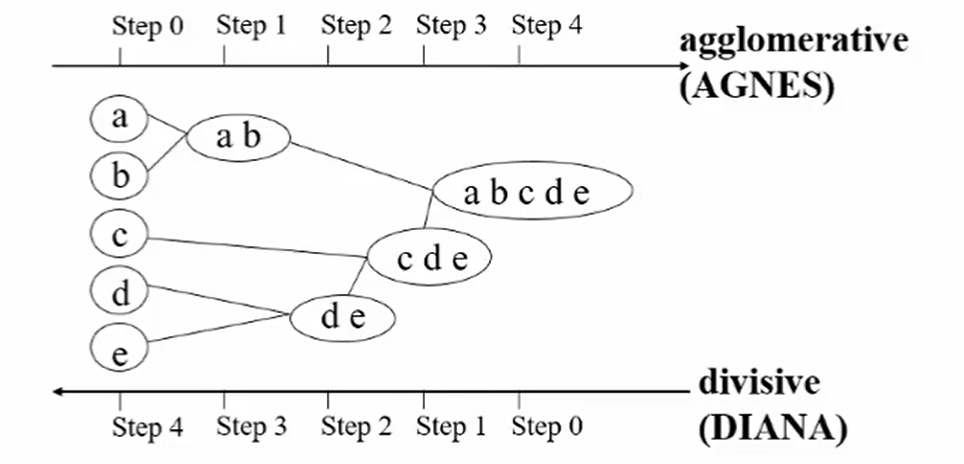
****

Figure 1

1. Ada dua cara untuk membentuk hierarchical clustering
   1. *Agglomerative Nesting* (*Agnes*). Dimulai dari beberapa *flat clusters;* pada setiap Langkah iterasi, kita menggabungkan dua *clusters* termirip dengan kata lain mendefinisikan arti “kedekatan” dua *clusters*.
   2. *Divisive Analysis* (*Diana*). Dimulai dari satu *cluster* (seluruh data), kemudian memecah belah *cluster* dengan kata lain mendefinisikan *cluster* mana yang harus dipecah dan bagaimana cara memecahnya.

Density-based : *DBSCAN*

1. 2 parameter penting :
   1. *Eps* : radius maksimal untuk node tetangga
   2. *MinPts* : jumlah minimal tetangga yang mencapai *Eps*
2. Langkah-langkah *DBSCAN*:
   1. Pilih sembarang titik p
   2. Ambil semua titik yang dapat dijangkau berdasarkan *Eps* dan *MinPts*
   3. Jika p adalah titik inti (memenuhi *MinPts*) maka terbentuk cluster
   4. Jika p adalah titik batas (tidak memenuhi *MinPts*). *DBSCAN* pindah ke titik p lain yang belum dikunjungi
   5. Ulangi hingga semua titik telah diproses

Model-based : *SOM*

1. Pada clustering network, terdapat banyak input sebagai komponen
2. Setiap output mempresentasikan suatu cluster dan banyaknya output berdasarkan banyaknya cluster
3. Selama training, network mencari output terbaik dari input
4. Bobot vektor yang selalu diupdate selama proses training berlangsung