

## UNSUPERVISED LEARNING

a. Apa itu unsupervised learning?

Unsupervised Learning merupakan metode pelatihan machine learning dimana dataset yang diberikan kepada komputer tidak diberikan labelnya, sehingga mesin atau komputer harus bisa mengelompokkan data yang ada secara mandiri (clustering)

b. Jelaskan bagaimana cara kerja dari algoritma yang anda implementasikan!

- K-Means Clustering

1. Pilih secara acak k buah datapoint yang akan menjadi centroid awal setiap cluster
2. Hitung jarak datapoint lainnya terhadap seluruh centroid, beri label sesuai dengan centroid yang terdekat sehingga sehingga telah terbentuk k cluster
3. Setelah semua datapoint diberi label, hitung rata-rata setiap fitur untuk setiap cluster, jadikan data rata-rata tersebut sebagai centroid yang baru dari cluster yang bersangkutan
4. Ulangi langkah 2 dan 3 sampai menuju konvergensi, dimana setelah perhitungan, tidak ada datapoint yang label atau clusternya berubah

- K-Medoids Clustering

1. Pilih secara acak k buah datapoint yang akan menjadi medoid awal setiap cluster
2. Hitung jarak datapoint lainnya terhadap seluruh medoid, beri label sesuai dengan medoid yang terdekat sehingga telah terbentuk sebanyak k cluster
3. Setelah semua datapoint diberi label atau sudah memiliki clusternya masing-masing, untuk setiap cluster, iterasi setiap datapoint dan hitung cost dari datapoint tersebut. Datapoint dengan cost terendah akan menjadi medoid yang terbaru  
\* cost merupakan jumlah jarak datapoint lain ke datapoint tersebut
4. Ulangi langkah 2 dan 3 sampai menuju konvergensi, dimana setelah perhitungan, tidak ada datapoint yang label atau clusternya berubah\

- DBSCAN

1. Pilih datapoint awal yang belum dikunjungi (selanjutnya akan disebut datapoint  $i$ )
  2. Periksa datapoint yang ada di radius sebesar epsilon di sekitar datapoint  $i$
  3. Jika jumlah datapoint yang ada pada radius datapoint  $i$  lebih besar sama dengan minimum point yang telah ditentukan (datapoint yang memiliki label sama dengan datapoint  $i$  atau yang belum diberi label), maka suatu cluster berhasil dibuat (beri label datapoint yang ada di radius tersebut secara seragam)
  4. Ulangi langkah 2 dan 3 sampai berhenti di kasus dasar sebagai berikut:
    - Datapoint  $i$  telah memiliki cluster dari perhitungan datapoint sebelumnya, namun pada perhitungan datapoint  $i$ , tidak terpenuhi sejumlah minimum point pada radius datapoint  $i$  (ini dinamakan border point)
    - Datapoint  $i$  merupakan datapoint inisial untuk sebuah cluster yang ingin dibuat, namun tidak terpenuhi syarat pembuatan cluster yaitu sebanyak minimum point yang telah ditentukan di radius datapoint  $i$  (ini dinamakan noise point)
    - (Kasus yang memenuhi syarat pembuatan cluster dinamakan core point)
  5. Setelah seluruh datapoint dikunjungi, beri label noise point sebagai cluster sendiri yang disebut noise/outliers
- c. Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kelemahannya!
- K-Means Clustering
- Kelebihan:
- Komputasi cepat dan terhitung sederhana
  - Prinsipnya sederhana
- Kekurangan:

- Sensitif pada centroid awal (terutama jika dimensi dari fiturnya sudah banyak)
- Sensitif terhadap outlier
- Outlier atau noise tidak ditangani dengan baik

- K-Medoid Clustering

Kelebihan:

- Tidak sesensitif k-means jika berhadapan dengan outlier
- Cluster direpresentasikan dengan datapoint yang ada dan nyata

Kekurangan:

- Komputasi cenderung lebih mahal
- Tidak sensitif terhadap outlier namun tidak ada pengelompokan tersendiri terhadap outlier

- DBSCAN

Kelebihan:

- Menangani outlier dengan baik
- Tidak perlu menentukan jumlah cluster yang ingin dibentuk

Kekurangan:

- Sulitnya menentukan nilai epsilon dan minimum point yang tepat

d. Jelaskan penerapan dari algoritma unsupervised! (misal di bidang kesehatan atau industri)

Beberapa contoh aplikasi unsupervised learning:

- Melakukan klasterisasi data kesehatan anak-anak untuk meneliti stunting pada anak
- Melakukan klasterisasi segmen pasar suatu bidang bisnis tertentu
- Melakukan klasterisasi data penduduk oleh pemerintah daerah untuk mengetahui kebutuhan masyarakat