

Jawaban Implementasi Algoritma *Unsupervised*

Soal

- a. Apa itu *unsupervised learning*?
- b. Jelaskan bagaimana cara kerja dari algoritma yang anda implementasikan!
- c. Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kelemahannya!
- d. Jelaskan penerapan dari algoritma *unsupervised*! (misal di bidang kesehatan atau industri)

Jawaban

Jawaban a

Unsupervised learning adalah salah satu cara mesin belajar dengan data yang tidak dilabeli. Dengan kata lain, hanya input saja yang diberikan sebagai acuan. Umumnya untuk mencari pola atau *cluster*.

Jawaban b

DBSCAN (*Density-Based Spatial Clustering of Application with Noise*)

DBSCAN adalah metode pencarian *cluster* dengan memerhatikan kepadatan (*density*) di sekitar setiap data (data dianggap titik). Kepadatannya berupa banyak titik yang berada pada *radius* tertentu dengan nilai yang disebut epsilon. Suatu *data* membentuk *cluster* jika banyak titik di sekitarnya mencapai banyak titik minimum. Jika tidak, itu dapat dikatakan sebagai data yang *outlier*.

Cara melakukannya adalah pilih salah satu data dan jadikan itu sebagai *cluster* baru untuk sementara, kemudian cari data yang berada dalam *radius* epsilon. Jika mencapai banyak titik minimum, proses titik-titik tersebut dalam cakupan *cluster* yang sama. Lakukan hingga tidak ada lagi titik yang dapat diproses. Lalu, ambil lagi titik lain sebagai *cluster* baru dan lakukan lagi hal yang sama. Jika *cluster* hanya berisi satu titik, jadikan itu sebagai *outlier*.

K-Means

K-means adalah metode pencarian *cluster* dengan memerhatikan rata-rata jarak setiap titik terhadap *centroid* untuk masing-masing *cluster*. Dalam metode ini, *centroid* didefinisikan sebagai posisi (nilai) rata-rata titik-titik dalam suatu *cluster*. Metode ini menghasilkan K *cluster*.

Cara melakukannya adalah masukkan titik ke salah satu dari K *cluster* secara acak. (Dalam algoritme yang saya gunakan, titik-titik dibagi ke dalam cluster sama rata sehingga banyaknya maksimal satu.) Kemudian hitung *centroid* untuk masing-masing *cluster*, lalu pindahkan semua titik ke *cluster* dengan *centroid* terdekat. (Dalam algoritme yang saya gunakan, jika *cluster* hanya tersisa satu titik, titik tersebut tidak akan dipindahkan untuk menghindari pencarian rata-rata tanpa data.) Lakukan hingga tidak ada perubahan.

K-Moloids

K-moloids adalah metode pencarian *cluster* dengan memerhatikan rata-rata jarak setiap titik terhadap *centroid* untuk masing-masing *cluster*. Dalam metode ini, *centroid* didefinisikan sebagai titik tengah dalam suatu *cluster* sehingga jumlah jarak setiap titik pada *centroid* memiliki nilai yang

minimum. Metode ini menghasilkan K *cluster*. Algoritmenya sama dengan *k-mean*, yang membedakan hanyalah definisi *centroid*-nya.

Jawaban c

- DBSCAN dapat mengidentifikasi *cluster* dengan bentuk yang bebas. Bentuknya tidak harus radial seperti yang dilakukan pada *k-means* dan *k-medoids*. Namun, konsekuensinya adalah *centroid* dari DBSCAN tidak ditemukan. Selain itu, banyak *cluster* yang dihasilkan tidak perlu dispesifikan; biarkan metode ini yang melakukan. *Outlier* juga dapat dideteksi. Namun, konsekuensinya adalah perlu banyak titik minimum dan epsilon yang tepat untuk *cluster* yang baik.
- *K-means* dan *k-medoids* dapat mengidentifikasi *centroid* dari setiap *cluster*. Namun, konsekuensinya adalah bentuk *cluster*-nya tidak bebas; dia tidak dapat mendeteksi *cluster* dengan bentuk donat, misalnya. Bentuknya harus radial. Selain itu, yang hanya dimasukkan hanyalah data dan nilai K ; tidak sebanyak DBSCAN dan hasilnya lebih *straightforward*. Namun, *outlier* tidak dapat dideteksi dan hasilnya dapat acak setiap proses.
- *K-means* lebih berfokus pada *centroid* dengan tengah dari hasil rata-rata, tidak peduli apakah titik ada di sana atau tidak. Sedangkan, *centroid* dari *k-medoids* adalah salah satu titik yang cocok.

Jawaban d

Unsupervised learning, sesuai yang disebutkan di awal, dapat mengidentifikasi pola dan *cluster*. Pola dan *cluster* ini membantu untuk mengambil keputusan, misalkan mendeteksi pusat penyebaran *covid-19* atau data yang tidak normal.