

Apa itu unsupervised learning?

Unsupervised learning adalah salah satu jenis pembelajaran mesin yang membiarkan mesin tersebut mengklasifikasikan secara mandiri data yang diberikan. Berbeda dengan supervised yang memiliki label di setiap entri datanya, unsupervised tidak memiliki label, yang akhirnya membuat pengklasifikasian didasarkan oleh ukuran lain (sebagai contoh, jarak).

Cara kerja algoritma yang diimplementasikan

DBScan

DBScan dilakukan dengan melihat jumlah tetangga-tetangga terdekat yang masih berada dalam suatu nilai jarak dari sebuah titik data yang terpilih. DBScan kemudian secara berkelanjutan menjadikan tetangga-tetangga titik tersebut sebuah titik inti tersendiri yang juga memiliki titik-titik tetangga mereka sendiri. Proses pencarian titik inti dan tetangga akan dilanjutkan sampai tidak ada lagi titik yang bisa melanjutkan pencarian. Pada akhir pencarian, seluruh titik yang bertetangga akan dijadikan sebuah kategori tersendiri. Algoritma ini dilakukan selama masih ada sekumpulan titik yang belum dikategorikan.

K Means

Algoritma K Means menerima input n kelas yang akan diprediksi. Algoritma K Means akan memilih n titik random yang dijadikan patokan seluruh titik data, kemudian akan ditentukan untuk seluruh titik data, salah satu dari n titik patokan yang paling dekat, membuatnya menjadi satu kategori. Algoritma K Means kemudian akan secara iteratif melakukan penghitungan ulang titik tengah yang merupakan rata-rata dari seluruh titik data dari sebuah kategori tersebut. Titik tengah yang baru didapatkan (tidak harus sebuah anggota dari dataset) dan pengkategorian titik berdasarkan titik inti diulang. Proses ini berulang sampai tidak ada perubahan pada waktu pembaruan nilai titik tengah.

K Medoids

Hampir mirip dengan algoritma K Means, K Medoids menerima input n jumlah kelas yang akan diprediksi, K Medoids juga mengambil n titik random untuk dijadikan patokan. Perbedaan K Means dan K Medoids adalah K Means bisa memiliki titik patokan yang tidak merupakan anggota dari dataset, sedangkan K Medoids selalu memiliki titik patokan yang merupakan anggota dari dataset. Penghitungan ulang titik tengah K Medoids adalah yaitu memilih sebuah titik yang paling dekat dengan semua titik dalam satu kategori, yang artinya seluruh jarak antar titik harus dihitung untuk mendapatkannya. Proses akan lanjut sampai pembaruan titik tengah menghasilkan titik yang sama dari penghitungan sebelumnya.

Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kekurangannya!

Algoritma	Kelebihan	Kekurangan
DBScan	<ul style="list-style-type: none">- Bisa handle cluster yang memiliki bentuk-bentuk yang beragam- Mengatasi noise dan outlier dengan baik	<ul style="list-style-type: none">- Kurang baik dalam data berdimensi tinggi- Kurang baik dalam data yang memiliki densitas beragam
K Means	<ul style="list-style-type: none">- Bisa untuk dataset berukuran besar- Pasti konvergen- Lebih cepat dari K Medoids	<ul style="list-style-type: none">- Harus memilih jumlah kelas secara manual- Tergantung nilai titik patokan awal- Tidak handle outlier
K Medoids	<ul style="list-style-type: none">- Pasti konvergen- Beradaptasi ke contoh-contoh baru	<ul style="list-style-type: none">- Harus memilih jumlah kelas secara manual- Tergantung nilai titik patokan awal- Tidak handle outlier- Komputasi mahal

Jelaskan penerapan dari algoritma unsupervised!

Contoh penerapan algoritma unsupervised adalah pembuatan program rekomendasi yang user-specific untuk e-commerce