

Jawaban Tugas Seleksi Asisten Lab AI '23

Dipersiapkan Oleh Brian A. Hadian (13523048)

Versi: 1.0 07/08/2025

Jawaban

DBSCAN

Jawablah pertanyaan-pertanyaan berikut:

1. Jelaskan cara kerja dari algoritma tersebut! (boleh dalam bentuk *pseudocode* ataupun narasi)
cara kerja algoritma DBSCAN adalah pertama dengan mengambil suatu titik (sampel) acak dari dataset dan kemudian melakukan kalkulasi jarak terhadap sampel lainnya menggunakan suatu fungsi jarak. Apabila jumlah sampel lain yang memiliki jarak dibawah batas threshold tersebut bernilai $> e$ (batas threshold tertentu), maka sampel lain tersebut berperan menjadi core point. Hal ini diteruskan untuk setiap core point sedemikian sehingga dapat diperoleh banyak core point. Setiap core point memiliki label yang sama dengan core point pertama yang dipilih.

Setelah ditemukan core point terakhir, dilakukan pengujian kembali untuk titik (sampel) yang tidak termasuk core point, tetapi masih berada dalam jangkauan threshold tertentu, maka titik tersebut dilabeli dengan label yang dimiliki oleh core point terdekat tersebut. Hal ini dilakukan terus menerus untuk setiap non-core point hingga tidak tersisa potential core point lagi (titik yang memiliki jarak $<$ batas tertentu).

Kemudian, diambil kembali suatu titik acak yang tidak memiliki label dan titik ini dijadikan sebagai core point yang baru. Proses evaluasi titik titik lain yang belum memiliki label dan memiliki jarak terdekat dengan core point baru tersebut dilakukan kembali hingga seluruh titik memiliki labelnya sendiri. Apabila terjadi kondisi bahwa terlalu banyak titik yang berjauhan (terlalu banyak outlier), maka akan diambil suatu label acak yang belum dimiliki oleh titik manapun sebelumnya dan label tersebut akan dipasangkan untuk titik titik outlier tersebut.

2. Bandingkan hasil evaluasi model pada nomor 2 dan 3, bagaimana hasil perbandingannya? Jika ada perbedaan, jelaskan alasannya!

Terdapat perbedaan terhadap hasil prediksi bahwa nilai prediksi yang dibuat menggunakan library khusus cenderung memiliki nilai yang lebih sesuai dengan model dibandingkan fungsi yang telah dibuat. Hal ini dapat diakibatkan oleh tidak dilakukannya normalisasi pada implementasi manual sehingga data cenderung memiliki noise. Selain itu, library khusus juga biasanya memiliki fungsi untuk mengganti nilai null / tidak ada secara lebih baik dibanding menggunakan rata-rata seperti pada implementasi manual. Terakhir, implementasi manual juga tidak menerapkan metode tambahan untuk meningkatkan akurasi dari weight dan bias diluar training, sementara library khusus memiliki langkah lebih lanjut untuk meningkatkan akurasi tersebut melalui metode lain. Tak lupa, library khusus juga bisa menerapkan metode validasi secara tertutup sehingga dapat meningkatkan akurasi.

3. Jelaskan *improvement* apa saja yang bisa Anda lakukan untuk mencapai hasil yang lebih baik dibandingkan dengan hasil yang Anda punya saat ini! *Improvement* yang dimaksud tidak terbatas pada bagaimana algoritma diimplementasikan, namun juga mencakup tahap sebelum *modeling and validation*.
 - a. melakukan normalisasi terhadap dataset (memindahkan pusat ke 0)
 - b. menerapkan metode validasi yang bersifat cross (melakukan acak data dan melatih ulang data dengan variansi dan range yang cukup besar)
 - c. menggabungkan beberapa metode learning menjadi satu untuk meningkatkan akurasi