

DBSCAN

1. Cara kerja algoritma DBSCAN

DBSCAN adalah algoritma clustering berbasis densitas yang dapat menemukan cluster dengan bentuk arbitrer dan mengidentifikasi data noise (outliers). Algoritma ini bekerja berdasarkan konsep kepadatan, di mana cluster didefinisikan sebagai area dengan titik-titik data yang memiliki kepadatan tinggi, dipisahkan oleh area dengan kepadatan rendah. Ide utamanya adalah bahwa untuk setiap titik klaster, lingkungan dengan radius tertentu harus berisi setidaknya sejumlah titik minimum. Langkah-langkah utama dalam algoritma DBSCAN adalah sebagai berikut.

Inisialisasi Variabel:

- Mulai dengan membuat daftar untuk menyimpan cluster dan menandai semua titik sebagai belum dikunjungi.

Iterasi pada Semua Titik Data:

- Untuk setiap titik data, jika titik tersebut belum dikunjungi, tandai titik sebagai sudah dikunjungi.
- Ambil semua tetangga titik dalam radius ϵ .

Periksa Titik sebagai Core Point:

- Jika jumlah tetangga lebih kecil dari MinPts, tandai titik sebagai noise.
- Jika tidak, buat cluster baru dan tambahkan titik ini serta tetangga-tetangganya ke dalam cluster.

Perluas Cluster:

- Untuk setiap tetangga dari core point, jika tetangga belum dikunjungi, tandai sebagai sudah dikunjungi dan cari tetangga-tetangganya dalam radius ϵ .
- Jika tetangga-tetangga ini juga core points, tambahkan mereka ke dalam cluster.
- Jika tetangga tidak ada dalam cluster manapun, masukkan ke dalam cluster.

Lanjutkan Sampai Semua Titik Diproses:

- Proses berlanjut dengan memperluas cluster dengan cara di atas sampai semua tetangga yang memenuhi syarat telah ditambahkan.
- Ulangi proses untuk titik-titik yang belum dimasukkan ke dalam cluster manapun sampai semua titik telah diproses.

4. Hasil Evaluasi Model

From Scratch	From Scikit-Learn
Silhouette Score (DBSCAN from scratch): -0.41258868246287606 Davies-Bouldin Index (DBSCAN from scratch): 1.484042732733766	Silhouette Score (DBSCAN from scratch): -0.41258868246287606 Davies-Bouldin Index (DBSCAN from scratch): 1.484042732733766

Hasil Silhouette Score dan Davies-Bouldin Index dari implementasi DBSCAN from scratch dan scikit-learn adalah sama. Hal ini menunjukkan bahwa kedua implementasi memberikan hasil yang konsisten dalam hal penilaian clustering, yang biasanya merupakan indikasi bahwa fungsi clustering bekerja dengan benar dalam hal menghasilkan output yang dapat dievaluasi dengan metrik yang sama. Implementasi from scratch mengikuti prinsip dasar DBSCAN dengan benar, termasuk perhitungan jarak dan pengelompokan, yang konsisten dengan bagaimana scikit-learn mengimplementasikan algoritma tersebut.

Silhouette Score yang rendah menunjukkan bahwa jarak antara data dalam cluster relatif dekat, tetapi jarak antara data di cluster yang berbeda cukup jauh namun masih rendah, yang menunjukkan bahwa hasil clustering tidak terlalu baik. Davies-Bouldin Index yang lebih tinggi mengindikasikan bahwa rata-rata rasio jarak intra-cluster terhadap inter-cluster adalah lebih tinggi, yang berarti ada ruang untuk perbaikan dalam segregasi cluster.

5. Improvement

Meskipun hasil konsisten, melakukan pengujian tambahan dengan berbagai parameter dan dataset lain akan lebih memperkuat validitas implementasi dari awal. Hal ini akan membantu memastikan bahwa implementasi dapat menangani berbagai kasus dan kondisi. Penanganan noise juga dapat dilakukan dengan memeriksa bagaimana setiap implementasi menangani titik data yang dianggap noise (label -1) dan memastikan tidak ada perbedaan dalam cara titik-titik tersebut diperlakukan.