

Jawaban Tugas Seleksi Asisten Lab AI '23

Dipersiapkan Oleh Brian A. Hadian (13523048)

Versi: 1.0 07/08/2025

Jawaban

K-Nearest Neighbor

Jawablah pertanyaan-pertanyaan berikut:

1. Jelaskan cara kerja dari algoritma tersebut! (boleh dalam bentuk *pseudocode* ataupun narasi)
function kNearestNeighbor(dataset : row[row[features]], x : features, neighborCount : int):
 { tentukan jarak dari setiap row pada dataset ke x menggunakan rumus jarak seperti Euclidean, minkowski, etc }
 distance = distanceFunction(dataset, x)
 {diambil 'tetangga' dengan jarak terpendek sebanyak neighborCount}
 nearestNeighbor = distance.sorted().head(neighborCount)
 {distance merupakan map[row] -> distance, kemudian dibuat suatu map baru yang mengkategorikan label dari setiap nearestNeighbor dengan jumlah nearestNeighbor yang termasuk dalam label tersebut }
 category = map(label, len(nearestNeighbor[label])
 {x kemudian dikategorikan memiliki label yang sama dengan label yang memiliki jumlah neighbor terbanyak}
 finalCat = category[label.maxNearestNeighbor()]
 -> finalCat
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