

Tugas 3 Pengantar Kecerdasan buatan

1. Masalah yang harus diselesaikan

Merancang program dengan K-Nearest Neighbor untuk mencari nilai k terbaik dan rata-rata akurasi menggunakan 5-fold cross-validation dengan diberikan himpunan data Pima India Diabetes Dataset (PIDD) pada file “Diabetes.csv”

2. Spesifikasi implementasi

Dalam membangun program menggunakan bahasa python dengan menggunakan jupyter notebook

3. Strategi Penyelesaian Masalah

3.1 Perhitungan jarak

Untuk menghitung jarak menggunakan rumus Euclidean, berikut rumus:

$$\text{L2 (Euclidean) distance}$$
$$d_1(x_1, x_2) = \sqrt{\sum_p (x_{1p} - x_{2p})^2}$$

3.2 Prapemrosesan data

Teknik yang digunakan dalam melakukan Prapemrosesan data adalah normalisasi min-max dengan rumus sebagai berikut :

$$x_{new} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}}$$

3.3 5-fold cross-validation

Dengan metode ini membagi dataset yang diberikan menjadi 5 dataset dengan komposisi objek-objek training set dan testing set , setelah dibuat 5 dataset di program ini dilakukan pengubahan dari dataframe ke sebuah list.

4. Klasifikasi Knn

Dengan melakukan perulangan sebanyak data training kemudian memanggil fungsi Euclidean untuk melakukan perhitungan jarak setelah itu di masukan ke list dan terakhir hasil dari list tersebut di sorting.

5. Pemilihan k terbaik

Dalam proses pemilihan k terbaik, mencari nilai terbesar dari rata –rata akurasi 5 dataset yang telah dibagi dengan nilai akurasi masing masing. kemudian jika sudah didapatkan nilai terbsesar maka selanjutnya mencari index dari nilai terbesar tersebut , dari index nilai terbesar itu didapatkan nilai k terbaik.

6. Output program

Dengan 20 tetangga yang diuji, maka didapatkan output sebagai berikut

Nilai k terbaik : 16

Rata-rata akurasi : 76.53176382967655