**LAPORAN TUGAS PROGRAM 2**

**KECERDASAN BUATAN**

**K-Nearest Neighbor**

* Analisis

Terdapat file DataTrain\_Tugas3\_AI.csv yang berisi himpunan 800 data yang memiliki 5 atribut input (X1, X2, X3, X4, dan X5) dan 1 output yang memiliki 4 kelas/label (0, 1, 2, dan 3). Terdapat file DataTest\_Tugas3\_AI.csv yang berisi himpunan 200 data yang memiliki 5 atribut input (X1, X2, X3, X4, dan X5) dan 1 output yang belum memiliki label. Untuk itu diminta untuk membangun sebuah sistem klasifikasi menggunakan metode k-Nearest Neighbors untuk menentukan kelas/label data testing dalam file DataTest\_Tugas3\_AI.csv. Sistem membaca masukan file DataTrain\_Tugas3\_AI.csv dan DataTest\_Tugas3\_AI.csv dan mengeluarkan *output* berupa file **TebakanTugas3.csv** berupa satu kolom berisi **200 baris angka bernilai *integer*/bulat** (0, 1, 2, atau 3) yang menyatakan kelas/label baris atau *record* yang bersesuaian pada file DataTest\_Tugas3\_AI.csv.

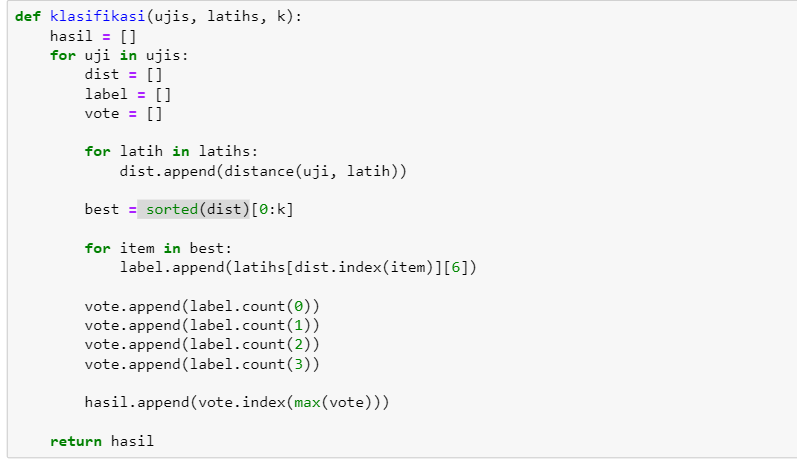
* Strategi Penyelesaian masalah yang digunakan menggunakan metode KNN

1. Menentukan jumlah tetangga yang akan kita perhitungkan (k) dengan bilangan bulat positif.
2. Hitung jarak setiap tetangga menggunakan rumus ***Euclidean Distance***

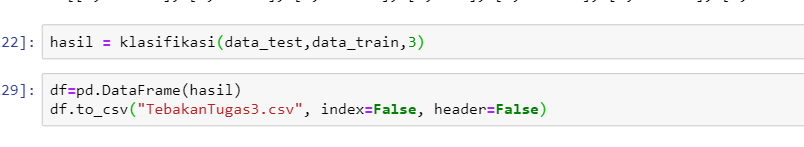




1. Pilih tetangga terdekat dari data baru sebanyak k atau dengan mensorting secara ascending untuk mencari jarak yang terdekat kemudian mengambil data sebanyak k
2. Tentukan klasifikasi paling umum pada hasil dari langkah 3, dengan menggunakan frekuensi terbanyak.



1. Outputkan klasifikasi data dari sampel data test berupa .csv

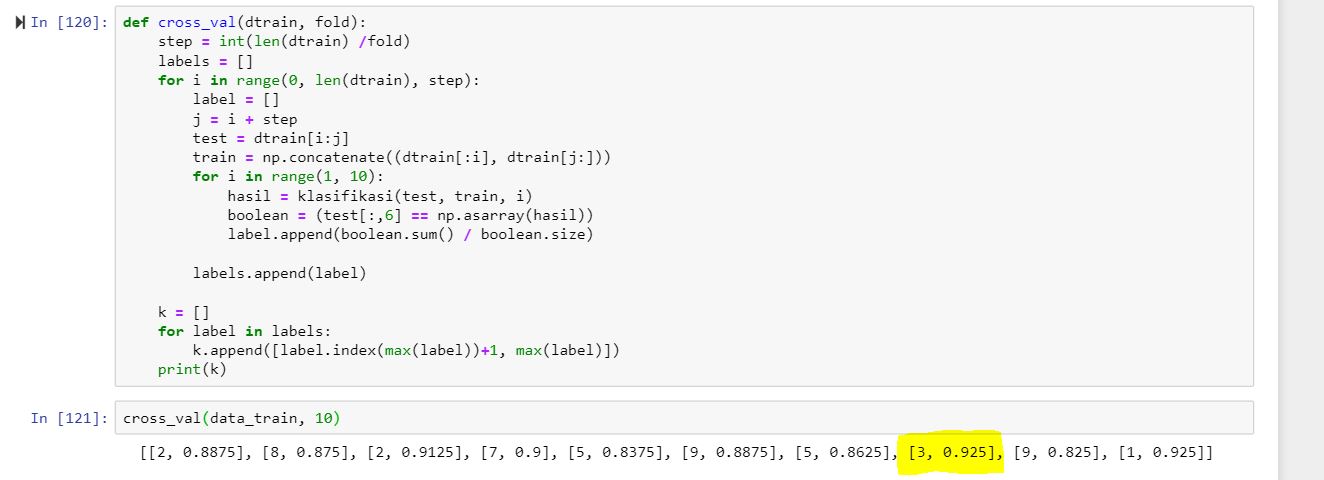


* Strategi menentukan K optimum dengan K-fold cross validation

Untuk menentukan nilai parameter k yang optimum, menggunakan metode K-fold cross validation yaitu Strategi menentukan k optimum berdasarkan akurasi yang paling optimum dengan cara mempartisi data atau melipat data

Konsep :

1. Mempartisi atau melipat data training(sebanyak 10 bagian)
2. Menentukan K (menguji 10 parameter yaitu dari 1-10 untuk mencari k yang optimum)
3. Menentukan rumus akuransi
4. Mengklasifikaskan data yang sudah di lipat dengan seluruh k yang akan di uji
5. Menentukan k optimum dengan akuransi optimum



K yang optimum adalah 3 dengan akuransi 0.925