Nama : Gagah Ghalistan (1301164203) – IF 40 04
Tugas Pemrograman 3 AI – K Nearest Neighbors

Diberikan file DataTrain_Tugas3_AI.csv berupa himpunan data berisi 800 data yang memiliki 5 atribut input (X1, X2, X3, X4, X5) dan 1 output yang memiliki 4 kelas / label (0, 1, 2, dan 3). Bangunlah 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.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.

Analisa :

- DataTrain_Tugas3_AI.csv berisikan 800 data dengan tiap datanya berisikan 7 kolom (Index, X1, X2, X3, X4, X5, Y(Kelas /Label))
- DataTest_Tugas3_AI.csv berisikan 800 data dengan tiap datanya berisikan 7 kolom (Index, X1, X2, X3, X4, X5, Y(Kelas / Label))
- Menebak Kelas / Label dari file DataTest_Tugas3_AI.csv dengan algoritma k-Nearest Neighbors
- Setelah melakukan validasi data (dengan hasil random index dari DataTrain_Tugas3_AI.csv) menggunakan **k fold cross validation dengan 8 fold,** akurasi dari suatu k mulai menurun / stagnan / mengeluarkan pattern perulangan setelah k > 20, dengan diketahuinya hal tersebut pengetesan sebaiknya dilakukan hanya hingga k=20 untuk meminimalisir running time validasi data.
- Terdapat **206 data dengan kelas 0**, **194 data dengan kelas 1**, **199 data dengan kelas 2**, dan **201 data dengan kelas 3** pada DataTrain_Tugas3_AI.csv
- **Rumus Manhattan dan Euclidean memberikan hasil yang berbeda.**

Strategi Penyelesaian:
Untuk menyelesaikan masalah yang diberikan dengan k-Nearest Neighbor, pertama adalah dengan menentukan nilai k yang dianggap terbaik untuk kasus yang diberikan, salah satu caranya dengan proses validation **K Fold Cross Validation,** Proses yang dilakukan adalah:

1. Ambil data dari DataTrain_Tugas3_AI.csv dan lakukan pengacakan urutan data
2. Bagi data yang teracak tersebut kedalam beberapa fold, dalam kasus ini fold yang digunakan adalah 8, setiap fold memiliki 100 data.
3. Lakukan proses k-nearest neighbors dengan 1 fold berperan sebagai dataTest dan fold lainnya sebagai dataTrain, Ulangi langkah ini hingga semua fold mendapatkan giliran menjadi dataTest.
4. Pilih nilai k yang akan digunakan dalam algoritma, dalam kasus ini algortima distance yang digunakan dalam proses k-Nearest Neighbors adalah Manhattan.
5. Hitung rata rata akurasi dari suatu k dan lakukan proses ke-3 dengan nilai k lainnya.
6. Dari hasil test validasi 8 fold dengan Manhattan, **nilai k terbaik yang didapatkan adalah 9 dengan akurasi 87.875%.**
7. Dengan telah didapatkannya nilai k yang dianggap paling optimum, implementasikan k-Nearest Neighbor kepada data test yang akan digunakan, menggunakan k = 9 dan rumus Manhattan.

```
fold=8, K=1, Accuracy=83.25%
fold=8, K=2, Accuracy=84.125%
fold=8, K=3, Accuracy=84.75%
fold=8, K=4, Accuracy=85.125%
fold=8, K=5, Accuracy=86.25%
fold=8, K=6, Accuracy=86.125%
fold=8, K=7, Accuracy=87.125%
fold=8, K=8, Accuracy=87.75%
fold=8, K=9, Accuracy=87.875%
fold=8, K=10, Accuracy=87.625%
fold=8, K=11, Accuracy=87.75%
fold=8, K=12, Accuracy=86.875%
fold=8, K=13, Accuracy=86.625%
fold=8, K=14, Accuracy=86.5%
fold=8, K=15, Accuracy=86.75%
fold=8, K=16, Accuracy=86.375%
fold=8, K=17, Accuracy=86.25%
fold=8, K=18, Accuracy=86.5%
fold=8, K=19, Accuracy=86.25%
fold=8, K=20, Accuracy=86.5%
```

Figure 1 Hasil Validasi Data

Nama : Gagah Ghalistan (1301164203) – IF 40 04
Tugas Pemrograman 3 AI – K Nearest Neighbors

Program Mencari K (program validasi data) dan program asli (Program penghitungan datatest) adalah 2 program yang berbeda. Program yang saya kumpulkan hanyalah program asli yang sudah menggunakan nilai K dari hasil pencarian K di file program berbeda.

Dengan telah ditemukannya k yang optimum, proses k-Nearest Neighbor bisa dilakukan ke data test asli, hasil dari proses k-Nearest Neighbor kepada data asli mengeluarkan index kelas yang akan dimasukkan kedalam file TebakanTugas3.csv, Berikut beberapa hasil tebakan index dari k=9 dengan rumus Manhattan.

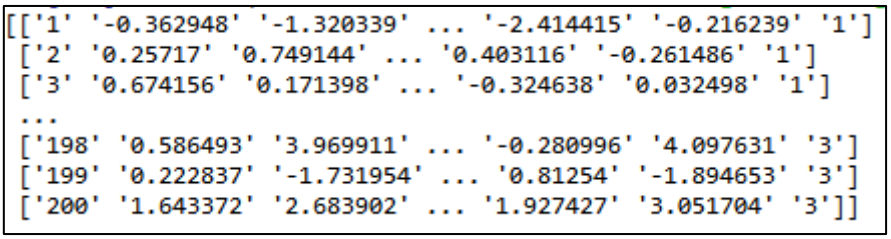

Figure 2 Hasil Tebakan DataTest