

**Laporan Pengerjaan Tugas Besar 3**  
**K-Nearest neighbor**



**Mata Kuliah :**  
**Artificial Intellegence**

**Nama :**  
**Irvan Naufali Rahmanto**

**1301164062**

**IF – 40 – 08**

- **Deskripsi Masalah :**

*K-Nearest neighbor* atau KNN adalah sebuah metode algoritma untuk melakukan klasifikasi data terhadap objek berdasarkan data pembelajaran jaraknya paling dekat dengan objek tersebut. Pada kali ini terdapat sebuah kasus diberikan sebuah data train berupa **DataTrain\_Tugas3\_AI.csv** yang berisikan 800 himpunan data yang memiliki 5 atribut input (X1,X2,X3,X4 dan X5) dan 1 output an yang memiliki kelas/label (0,1,2,3). Kita diminta untuk membangun sebuah sistem klasifikasi menggunakan metode *k-nearest neighbor* untuk menentukan kelas/label data testing berupa file **DataTest\_Tugas3\_AI.csv**. Sistem akan membaca outputan file dari **DataTrain\_Tugas3\_AI.csv** untuk mengeluarkan outputan berupa **TebakanTugas3.csv** berupa satu kolom berisi 200 baris angka bernilai *integer*/bulat yang menyatakan label/kelas *record* yang bersesuaian pada file **DataTest\_Tugas3\_AI.csv**.

- **Strategi Penyelesaian :**

- A. Perhitungan rumus :**

Pada step pertama strategi penyelesaian kita dapat kan dari data train yang berada pada file **DataTrain\_Tugas3\_AI.csv** adalah 800 data yang akan dicoba kan pada program yang dijalankan, sebelum itu kita perlu *import* terlebih dahulu data yang akan dicoba ke pada algoritma nya, kemudian setelah dari sana setelah data terbaca kita bisa lakukan perhitungan rumus. Rumus yang digunakan pada kasus kali ini ialah *manhattan distance*, berikut rumus nya 
$$d_1(x_1, y_2) = \sum_p |x_1^p - y_2^p|$$

- B. Pengurutan (sorting) :**

Proses selanjut nya setelah pembacaan file atau *import* untuk kemudian dimasukkan ke dalam perhitungan rumus menggunakan *manhattan distance* adalah melakukan pengurutan data nya tersebut atau *sorting* yang bertujuan untuk mengetahui dan mendapatkan jarak antara data train dengan data test. Dari sana untuk mendapatkan K nya harus menggunakan uji coba test K nya.

- C. Pencarian dari sorting :**

Kemudian dari pengurutan data berdasarkan data train dan data test yang didapatkan setelah digunakan rumus akan menghasilkan data yang terurut, untuk itu dari sana kita perlu melakukan data berdasarkan nilai terbanyak dari nilai distance yang ditemukan atau nilai K yang didapatkan. Kita lakukan metode uji coba untuk mendapatkan K nya , maka dari hasil sorting yang didapatkan dari sana kita lakukan pencarian data distance terbanyak untuk mendapatkan parameter dari ouputan yang diminta atau hasil dari parameter Y nya. Cara nya dengan menghitung secara intens terhadap jumlah dari setiap kateori nya untuk membandingkan K nilai pada baris di list tersebut agar mendapatkan nilai Y yang diminta berdasarkan data nya.

- D. Pengumpulan data hasil tes :**

Strategi penyelesaian terakhir yang digunakan pada metode k-nearest neighbor adalah melakukan pengumpulan data hasil berdasarkan step dan metode yang dilakukan sebelum nya untuk dimasukkan kepada data tebakan. Setelah melakukan perhitungan rumus dari *import* data yang dilakukan kemudian dilanjutkan dengan pengurutan data nya (*import*) untuk nanti nya dilakukan pencarian data terbanyak berdasarkan distance yang disesuaikan maka dari data train yang diujikan kepada data tes nya untuk menghasilkan data yang berisikan 200 baris/angka bertipe *integer*/bulat yang mencocokkan data tebakan yang dihasilkan.

➤ Screenshoot output :

```
793
-0.234877
-1.197815
3.152064
1.837798
-1.075383
3
794
1.027813
1.395114
-0.125166
2.644815
2.458163
3
795
1.397522
2.020321
-1.314494
2.501385
2.402569
3
796
1.346162
0.996949
0.842938
1.56318
3.051993
3
797
1.080028
6.290745
-3.27249
0.70706
3.433035
3
798
1.792138
-1.532045
-0.212688
1.346853
3.14493
2
799
0.63085
5.350815
-3.716385
1.064686
2.790904
3
800
1.320487
0.129181
3.289529
0.63392
-0.57673
3
```