**TUPRO 3 Artificial Intelligence**

**[k-Nearest Neighbour]**

1. **Analisa dan Pembahasan**
2. **k-Nearest Neighbour**

k-Nearest Neighbour adalah salah satu metode klasfikasi pada machine learning dan juga kNN yang paling mudah dimengerti (straightforward) karena proses klasfikasi pada metode ini dengan mengidentifikasi sebuah data testing dengan tetangga terdekat dari data training dan menggunakan tetangga tersebut untuk memberikan data testing tersebut sebuah class berdasarkan class tetangga. Untuk k lebih sering di pakai tidak hanya satu dan berguna untuk k lebih dari satu sehingga metode ini sering disebut dengan k-Nearest Neighbour(k-NN) Classification dimana k nearest neighbours dipakai untuk menetukan label class. k-NN juga memiliki istilah lain yang disebut dengan Memory-Based Classification karena data training ada di memory saat program berjalan kemudian disebut juga Lazy Learning teknik karena tertunda akibat run time dan disebut juga sebagai Example-Based Classification/ Case-Based Classification karena secara langsung mengklasifikasikan dengan data training.

Secara pseudocode k-NN dapat dijelaskan sebagai berikut :

*Untuk setiap point dataTraining:*

*Hitung jarak antara dataTesting dan point sekarang.*

*Urutkan jarak secara ascending*

*Ambil k buah dengan jarak terendah dengan dataTesting*

*Temukan class yang terbanyak di k buah tetangga tersebut*

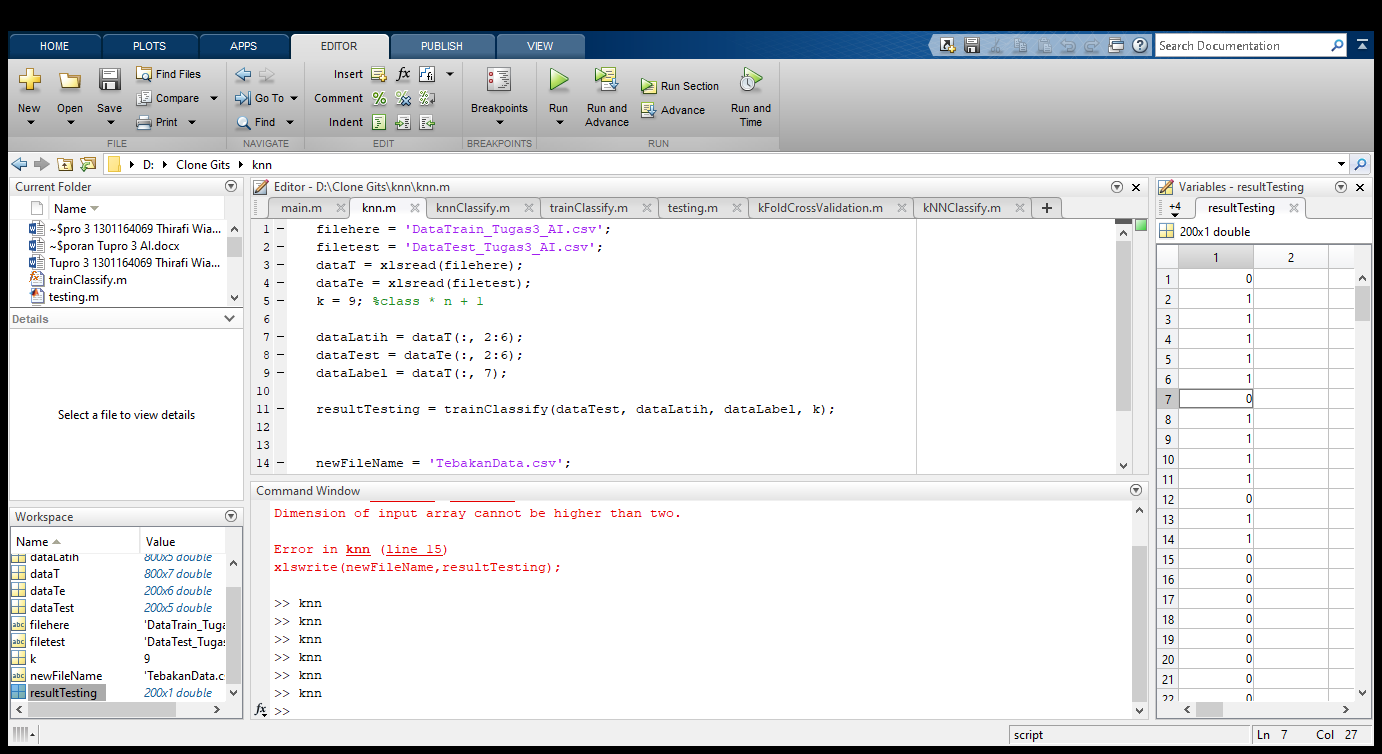
*Kembalikan class yang terbanyak tersebut sebagai class untuk dataTesting.*

1. **Hasil**

Implementasi dari kedua algoritma diatas adalah pada sebuah data set yang terdiri dari 800 baris data training, dan 200 baris data testing. Pada setiap data terdapat 6 variable kolom yaitu X1, X2, X3, X4 dan X5. Untuk variable yang menunjukan ID dari setiap baris adalah index. Penulis menggunakan Matlab untuk melakukan kedua algoritma tersebut.

Adapun langkah-langkah yang dilakukan adalah sebagai berikut :

1. Mengimport data set pada sheet 1 ke variable dataTraining
2. Melakukan k-Fold Cross Validation pada dataTraining dengan k adalah 9 dan k maksimal untuk kNN adalah 200 dengan pertambahan k adalah bilangan ganjil, dan hasil dari step ini adalah sebuah k = [1,3,5..199] dan akurasi masing-masing k.
3. Mengimport data set pada sheet 2 ke variable dataTesting
4. Menggunakan k terbaik pada step 2 untuk dilakukan kNN klasifikasi untuk mengklasifikasikan data Testing dan menghasilkan output klasifikasi.
5. Membuat file excel baru dengan hasil dari step 5.



Gambar 1Output Program