|  |
| --- |
| Tugas Program K-Nearest Neighbor |
|  |
| Dosen Pengampu: Bernadus Anggo Seno Aji, S.Kom, M.Kom  Nama: Aziza Hayupratiwi  NIM: 1301150440  Kelas : IF 39-06 |

10/28/2017

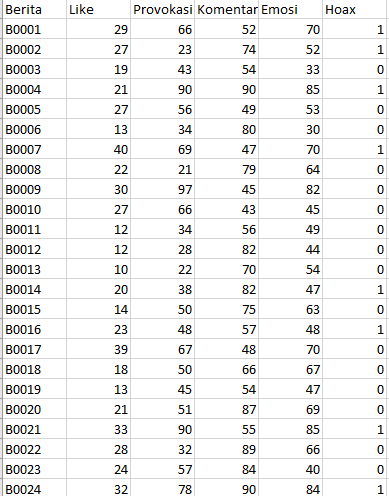
**2017**

# Deskripsi Kasus

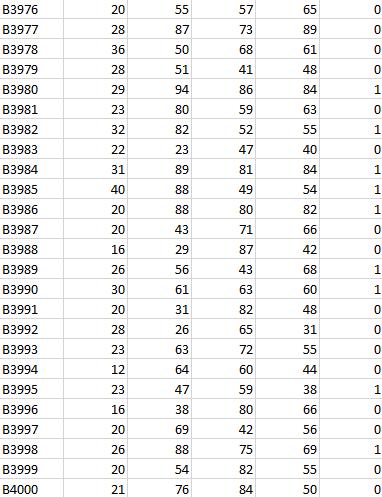
K-Nearest Neighbor adalah metode untuk klasifikasi data berdasarkan data dengan jarak terdekat disekitarnya. Dalam tugas program 3 ini dibuat solusi berupa model klasifikasi untuk permasalahan yang ada. Klasifikasi yang akan dibuat akan menentukan suatu data termasuk kelas 1 atau 0 yang berarti 1 data berita termasuk hoax dan 0 berarti bukan hoax.

# Data Pre-Processing

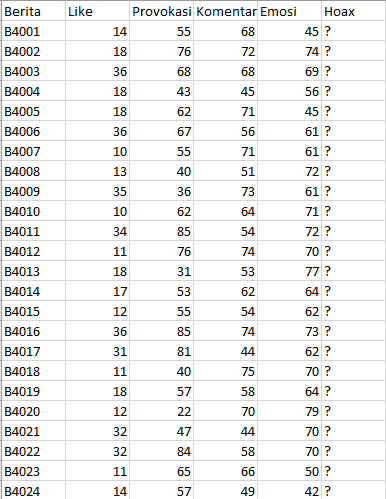
Data yang digunakan sebagai sebanyak 5000 data dengan 4000 data training dan 1000 data testing yang mana akan ditentukan berdasarkan hasil dari KNN. 1000 data tersebut akan diklasifikasi berdasarkan hasil Euclidean Distance dan dengan nilai K yang sesuai. Untuk hal ini nilai K yang sesuai yaitu 75.



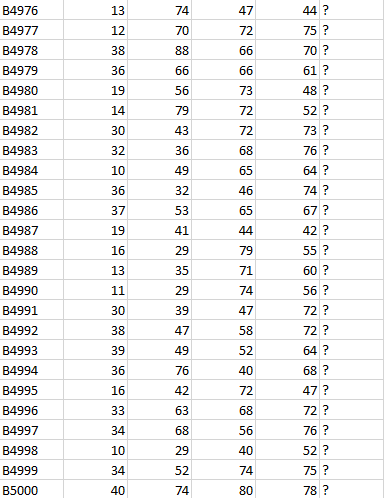
Gambar 1 Data Training



Gambar 2 Data Training



Gambar 3 Data Testing



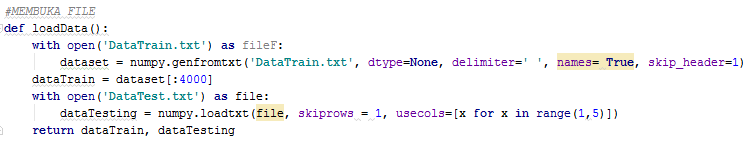
Gambar 4 Data Testing

# Rancangan Metode

Untuk menyelesaikan masalah ini digunakan metode K-Fold Validation terhadap data training untuk dijadikan sebagai data testing. Pada tugas program ini digunakan 400 data terakhir sebagai data testing. Nilai fold yang digunakan yaitu 100. Setelah dihitung menggunakan fold yang ditetapkan, hitung nilai akurasi setiap fold lalu hitung akurasi rata-rata dari fold tersebut. Ambil nilai akurasi tertinggi dan dengan nilai disetiap data yang stabil agar jika ada data dengan nilai akurasi rendah tidak terlalu jauh bedanya dengan nilai akurasi yang tinggi. Selanjutnya dapat ditetapkan nilai K mana yang akan digunakan untuk diterapkan pada 1000 data testing.

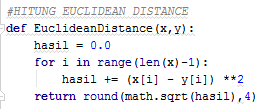
# Simulasi Metode

Tahap pertama yaitu membuka file data training dan data testing dalam bentuk txt.



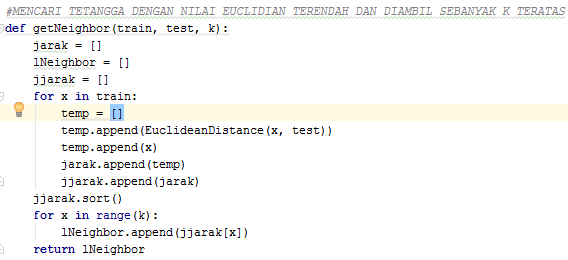
Gambar 5 Membuka dan Membaca File

Pada tahap ini digunakan 400 data training yang akan digunakan sebagai data testing dengan menghitung nilai Euclidean Distance agar didapat nilai akurasi untuk setial fold.



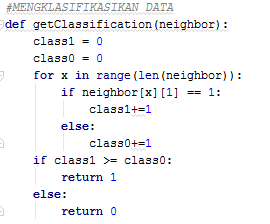
Gambar 6 Euclidean Distance

Setelah dilakukan penghitungan jarak dengan Euclidean Distance, maka dilakukan pencarian tetangga denga nilai eulidean terendah lalu diambil sebanyak K teratas.

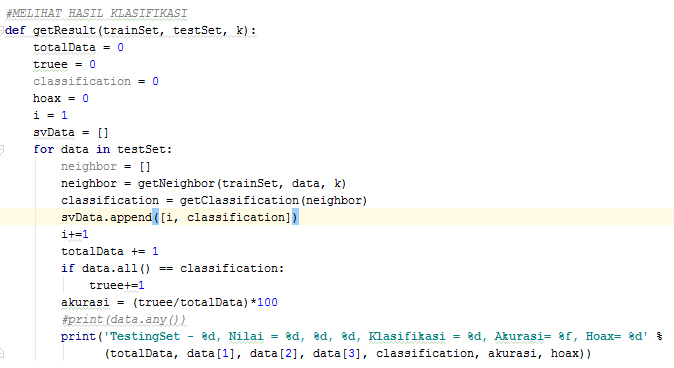


Gambar 7 Mencari nilai K

Pada tahap ini data berita akan diklasifikasi apakah termasuk kelas 0 atau 1.



Gambar 8 Klasifikasi Data

Tahap terakhir yaitu melihat hasil klasifikasi dan nilai hoax dari data testing.

Gambar 9 Hasil Klasifikasi