|  |
| --- |
| *Menentukan kategori berita dengan metode klasifikasi k-Nearest Neighbor (kNN)* |
| Laporan Tugas Program 3 Kecerdasan Buatan |
| Dosen Pengampu:  Untari Novia Wisesty  Disusun oleh :  Febry Ghaisani (1301154576) |

12/2/2017

**2017**

# Spesifikasi Program

1. Studi Kasus

Diberikan suatu himpunan data berisi 5000 berita dengan empat atribut: Jumlah Like, Emosi, Komentar, dan Provokasi yang bernilai 0 sampai 100, serta atribut kelas *Hoax* yang bernilai 1 yang berarti “Hoax” dan 0 yang berarti “Bukan Hoax”, seperti terdapat dalam file “Dataset Tugas 3.xlsx”. 4000 data dalam sheet “DataTrain” digunakan sebagai data latih untuk mendeteksi 1000 berita dalam sheet “DataTest” yang belum diketahui kelasnya. Metode klasifikasi yang digunakan dalam Tugas Program 3 ini adalah **k-Nearest Neighbor (kNN)**. Data latih tersebut dapat dibagi menjadi data latih dan validasi (dapat menggunakan metode cross validation) untuk mengukur seberapa akurat system klasifikasi yang dibangun. Dataset yang digunakan dalam Tugas Program 3 terlampir dalam folder.

1. Batasan Masalah

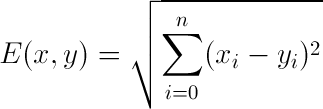
* Program dibangun menggunakan Bahasa pemrograman Python.
* Program menggunakan metode klasifikasi k-Nearest Neighbor (kNN).
* Tidak ada inputan user, untuk nilai k di-inisialisasi di awal.

1. Spesifikasi Kebutuhan Program

* PC/Laptop
* OS Minimum Windows 7
* Python 2.7
* JetBrains PyCharm
* Anaconda 2

# Rancangan Metode

Metode Klasifikasi k-Nearest Neighbor memanfaatkan rumus distance Euclidean, yaitu sebagai berikut :



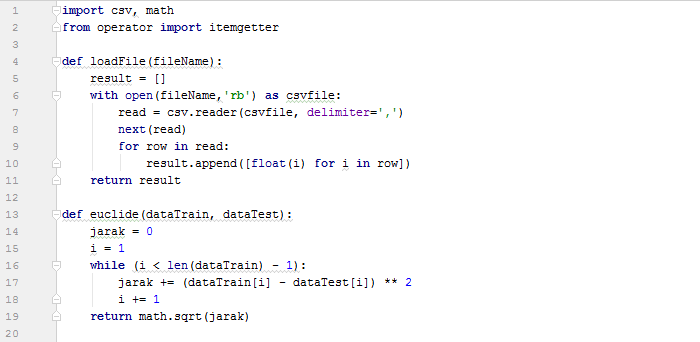
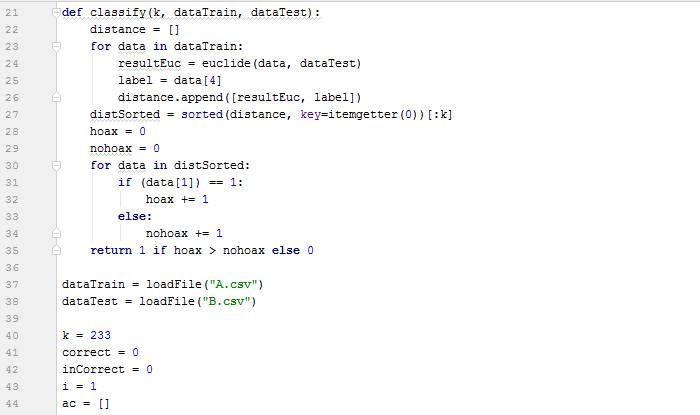
Algoritma k-NN :

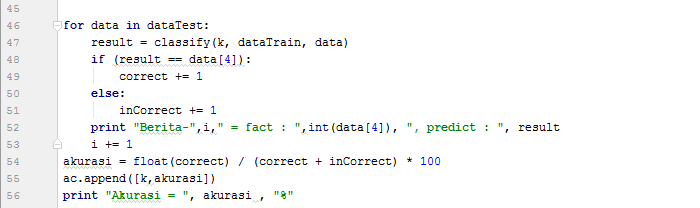
1. Hitung jarak antara data yang akan diuji dengan masing-masing data latih dengan rumus Euclidean distance lalu simpan hasil perhitungan beserta label dari data latih ke dalam array.
2. Lakukan sorting array secara ascending berdasarkan jarak.
3. Ambil data dari array sebanyak “k” dimulai dari data terkecil atau data pertama.
4. Bandingkan jumlah data atau berita yang hoax dan yang tidak hoax (voting).
5. Label baru didapat dari hasil voting. Kemudian simpan label baru tersebut.

# Pengujian Program

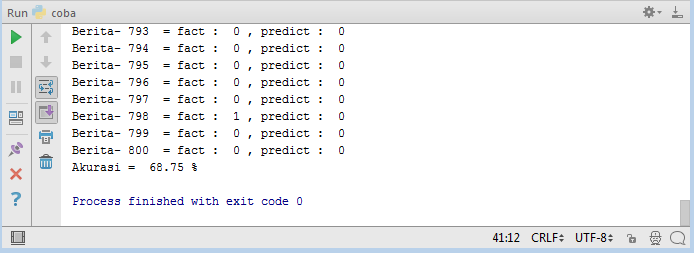
Pengujian data validasi yang berjumlah 800 berita terhadap data latih dengan jumlah 3200 berita.

Code coba.py :

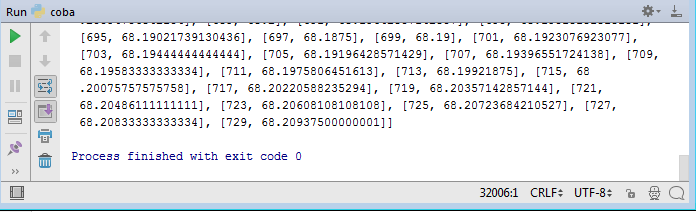
 



Hasil running :



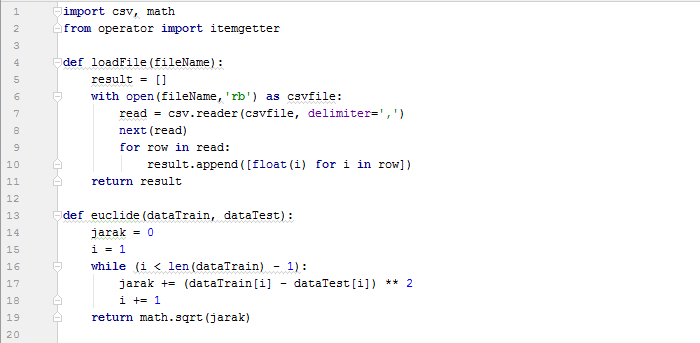
Nilai k dicari dengan menggunakan skema perulangan dengan menguji semua nilai k ganjil dari 1 sampai 729. Hasil pencarian nilai k beserta nilai akurasi (%) :

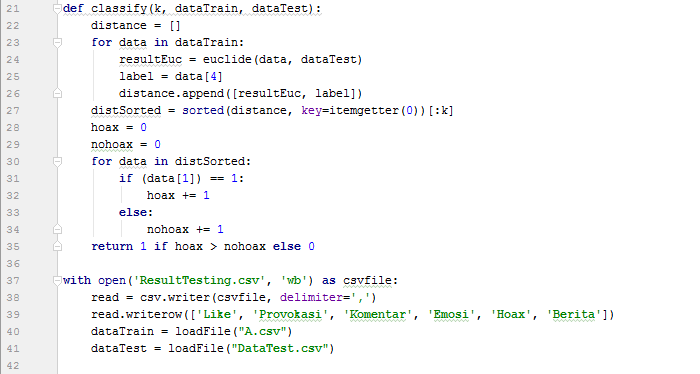
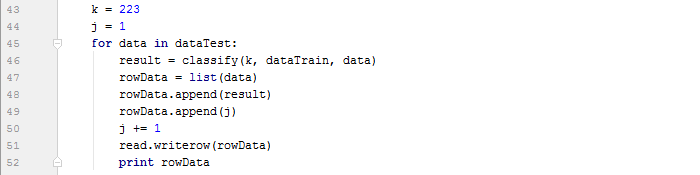


Grafik nilai k terhadap akurasi :

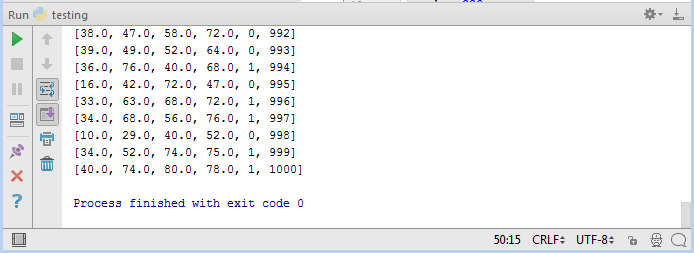
Nilai k yang digunakan adalah 233. Kemudian nilai k diterapkan ke data testing dan disimpan hasilnya ke file dengan ekstensi .csv

Code testing.py :



Hasil running :



Hasil klasifikasi disimpan di resultTesting.csv (file terlampir).

# Referensi

Slide mata kuliah Artificial Intelligence 2017.

https://en.wikipedia.org/wiki/K-nearest\_neighbors\_algorithm