**Laporan Tugas Program 2  
Kecerdasan Buatan  
  
*KNN***

******

**Novelty Octaviani Faomasi Daeli**

**1301154522**

**IF 39-04**

**Telkom University**

**S1 Teknik Informatika**

# Deskripsi Masalah

Pada file tugas program ke-3 diberikan data train dan data test dengan format xls. Pada data train terdapat 4000 data dengan 5 atribut yang terdiri dari jumlah like, emosi, komentar, provokasi dengan range tiap atribut sebesar 0 hingga 100 dan atribut kelas dengan nilai 1 untuk ‘Hoax’ atau 0 untuk ‘Bukan Hoax’. Pada data test terdapat 1000 data dengan 4 atribut yang terdiri dari jumlah like, emosi, komentar, dan provokasi dengan range nilai tiap atribut yaitu 0 hingga 100 dan nilai atribut kelas yang belum diketahui nilainya. Melalui data test akan dilakukan learning dengan metode KNN (k-Nearest Neighbor) pada data train, hasil learning akan digunakan untuk mencari nilai atribut kelas yang belum diketahui pada 1000 data test. Sebelum hasil learning pada data train diterapkan, akan dilakukan validasi hasil learning terlebih dahulu dengan cara mengukur tingkat akurasi learning dengan teknik penggunaan data train sebagai data validasi.

1. **Rancangan KNN**
2. Rancangan Input

Rancangan awal input pada tugas besar KNN ini yaitu dengan cara memanggil file yang telah disimpan sebelumnya pada folder yang sama dimana file aplikasi matlab disimpan. Terdapat 2 file yang akan diolah pada tugas besar kali ini, nama file yang akan diolah yaitu Data Train dan Data Test. Kedua file tersebut berformat xlsx. Sehingga, rancangan input yang digunakan yaitu dengan cara memanggil file dan menampungnya dalam satu variable.

Contoh penulisan penggunaan pemanggilan file dan menampungnya pada satu variable yaitu:

A = ‘Data Train.xlsx’

Dari contoh di atas A merupakan variable penampung, ‘Data Train’ adalah nama file yang akan diolah, dan ‘.xlsx’ adalah jenis format file.

Setelah dilakukan pemanggilan file, hal yang dilakukan selanjutnya yaitu membaca file yang telah diinput. Cara pembacaan filenya yaitu dengan memanggil nama variable yang telah menampung file sebelumnya. Berikut contoh pembacaan file yaitu :

DataTrain = xlsread(a)

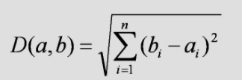
Dari contoh di atas, ‘DataTrain’ merupakan variable penampung, ‘xlsread’ adalah syntax untuk membaca file excel, dan ‘a’ adalah variable sebelumnya yang telah menampung isi data file ‘data train’.

1. Proses Klasifikasi

Proses klasifikasi yang digunakan pada tugas besar kali ini yaitu dengan cara mencari kfold dan K terbaik. Data yang digunakan untuk mencari kfold dan K terbaik yaitu data yang terdapat pada data train. Pencarian kfold dan K terbaik ditentukan melalui besar tingkat akurasi yang dicapai. Range pencarian kfold yang digunakan yaitu dari rentang 4 hingga 500. Range pencarian K yang digunakan yaitu dari rentang 1 hingga 500. Setelah di dapat nilai K dan kfold terbaik, maka nilai K akan digunakan untuk melakukan klasifikasi pada data test.

Berikut langkah-langkah klasifikasi yang digunakan :

1. Akan dicari jarak terdekat antara data train dengan data test sejumlah K. Rumus yang digunakan yaitu :



Keterangan :

a = data train

b = data test

ai = nilai data train yang ke-i

bi = nilai data test yang ke-i

n = banyak atribut

1. Akan dilakukan voting label pada data yang terdapat pada K terdekat.
2. Voting label terbanyak akan menentukan jenis kalisifikasi pada data test.
3. Proses Mencari Nilai K yang tepat

Proses mencari nilai K yang tepat yaitu dengan cara validasi. Berikut validasi yang akan digunakan untuk mencari nilai K yang tepat :

1. Akan dicari kfold terbaik terlebih dahulu dari range 1 hingga 500. Data akan dibagi sebanyak kfold dengan sama rata tiap bagiannya.
2. Tiap bagian data yang telah dibagi dengan beberapa kfold akan dicari tingkat akurasinya dengan cara menggunakan K dari range 1 hingga 500.
3. Tiap satu bagian data pada data train akan menjadi data validasi pada percobaan validasi ini. Lalu hasil data validasi yang telah didapat akan dicocokan dengan data train yang telah ada sebelumnya.
4. Hasi akurasi tiap bagian data sebelumya akan dijumlah dan dibagi sebesar kfold sehingga didapat besar rata-rata akurasi dengan kfold dan K tertentu.
5. Hasil rata-rata akurasi yang telah didapat sebelumnya akan disimpan, lalu akan diulang pencarian kfold dan K terbaik dengan cara melakukan perulangan dari langkah ke-1 hingga langkah ke-4
6. Setelah perulangan selesai, nilai akurasi terbesar akan diambil. Nilai kfold dan K yang digunakan pada rata-rata akurasi terbesar akan digunakan untuk melakukan validasi pada data test.
7. Output yang dihasilkan

Output yang dihasilkan yaitu hasil klasifikasi tiap data test yang telah tersimpan pada data excel dengan menggunakan kfold dan K yang telah divalidasi sebelumnya.

1. **Screenshoot**

