**LAPORAN TUGAS PROGRAM 3 KECERDASAN BUATAN MENGGUNAKAN METODE KNN**

****

**DISUSUN OLEH:**

**ARIEF BUDHIMAN**

**(1301154360)**

**IF 39-10**

**S1 TEKNIK INFORMATIKA – FAKULTAS INFORMATIKA**

**TELKOM UNIVERSITY**

1. **DESKRIPSI MASALAH (STUDI KASUS)**

KNN atau K-Nearest Neighbor adalah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Pada kasus ini, diberikan 4000 data training dan 1000 data testing. Untuk membuat program, digunakanlah data training untuk melakukan validasi. Dengan menggunakan konsep cross validation K-Fold, data training dibagi menjadi 100 untuk data validasi dan 3900 untuk training. Dengan menggunakan cross validation, setiap data yang masuk, berfungsi ganda, suatu saat menjadi data validasi, dan di saat yang lain menjadi data training.

Setelah dibuat program untuk validasi, dicarilah nilai K yang paling optimal. Setalah nilai K ditentukan, maka program ditambahkan untuk mengisi data testing berdasarkan data training.

1. **RANCANGAN METODE YANG DIGUNAKAN**

Sesuai dengan judul, metode untuk menyelesaikan tugas program 3 adalah menggunakan KNN atau K-Nearest Neighbor. Diberikan 4000 data training dan 1000 data testing. Data training sudah disertai dengan nilai HOAX. Sedangkan untuk data testing tidak disediakan nilai HOAX.

Tujuan dari program dibuat adalah untuk mengisi kolom HOAX pada data testing. Bahasa pemrograman yang digunakan adalah python 3.6.3. Aplikasi bantuan yang digunakan adalah pycharm, dan dibutuhkan library XLRD agar dapat memasukkan data excel ke dalam python.

**RANCANGAN INPUT**

Rancangan input pada tugas program 3 adalah dari file excel yang diberikan. Untuk membuat file excel dapat dibaca di python, diperlukan library tambahan yang telah dijelaskan sebelumnya, yaitu XLRD. Setelah dilakukan import menggunakan XLRD, data pada pada kedua sheet dimasukkan ke dalam list yang berbeda. Tujuan membedakan list ini adalah pada saat melakukan test, data tidak tercampur/terjadi ambigu.

**PROSES KLASIFIKASI**

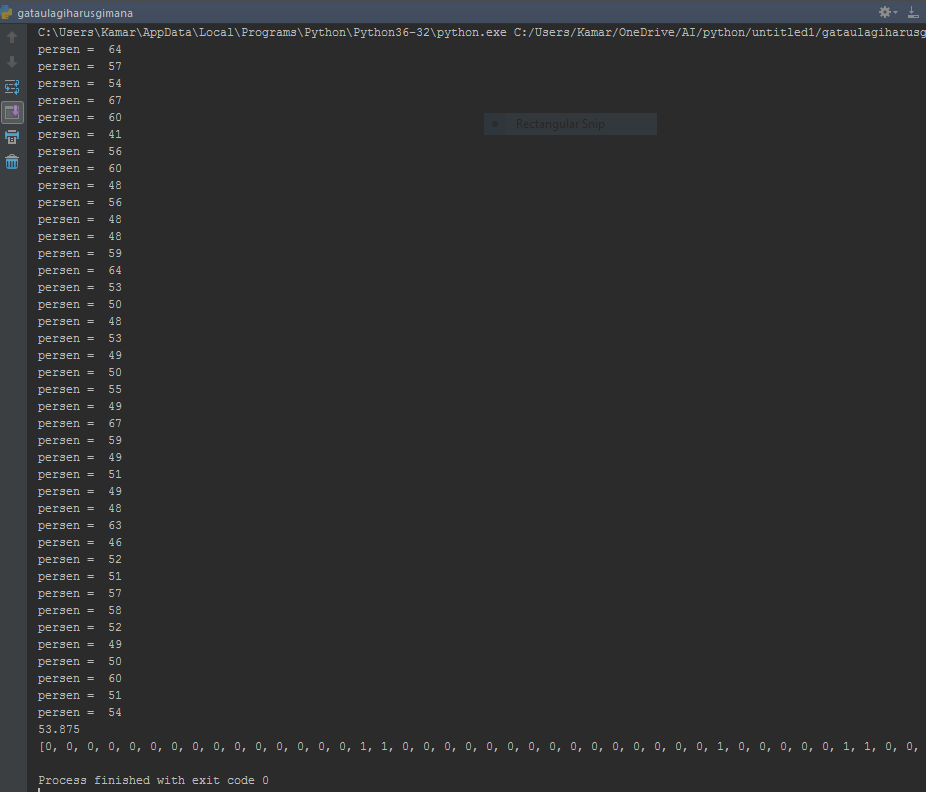
Klasifikasi untuk penentuan persentase dari program adalah 100 untuk data validasi dan 3900 untuk data training. Untuk memperoleh nilai presentase yang lebih akurat, digunakanlah cross validation (K-Fold). Sehingga program diulang sebanyak 40 kali.

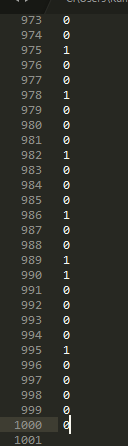
**Proses Mencari Nilai K**

Untuk mendapatkan nilai K yang optimal, dilakukan bruteforce/dilakukan satu per satu pengecekan. Nilai K yang digunakan adalah 131 dan diperoleh presentase sebesar 53.875% dengan nilai maksimal sebesar 67% dan minimal sebesar 41%

**OUTPUT YANG DIHASILKAN**

Output yang dihasilkan untuk mencari nilai HOAX pada sheet data training bernilai 0 dan 1. 0 berarti tidak HOAX dan sebaliknya, 1 bernilai HOAX. Nilai HOAX sudah terurut berdasarkan index pada saat memasukkan data ke dalam list. Berikut adalah screenshoot output yang dihasilkan



**SCREENSHOOT OUTPUT**