**LAPORAN TUGAS PROGRAM III**

**KNN (K-Nearest Neighbors)**

****

**Disusun oleh :**

**Nama : DAMAIYANTO.H**

**Nim : 1301150061**

**Kelas : IF-39-05**

**ARTIFICIAL INTELLIGENCE**

**S1 TEKNIK INFORMATIKA**

**TELKOM UNIVERSITY**

**2017**

**Deskripsi Masalah**

Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan k ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data.

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 AI 1718.xlxs”. Gunakan 4000 data dalam sheet “DataTrain”, sebagai data latih untuk mendeteksi apakah 1000 berita yang belum diketahui kelasnya, dalam sheet “DataTest, adalah berita bohong (hoax) atau bukan. Metode klasifikasi yang digunakan dalam Tugas Program 3 ini adalah k-Nearest Neighbor (kNN). Anda dapat membagi kembali data latih tersebut menjadi data latih dan validasi (dapat menggunakan metode cross validation) untuk mengukur seberapa akurat system klasifiksi yang sudah anda bangun.

**ALGORITMA PROGRAM :**

* Tentukan nilai K secara sembarang dan disarankan K adalah ganjil
* Lakukan cross validation pada data training sehingga menghasilkan beberapa model dan penentuan data latih dan datatesting
* Masukkan rumus terhadap data latih dan data testing
* Sorting dan pilih nilai terendah sebanyak K
* Lalu menentukkan HOAX/TIDAK dari nilai terendah sebanyak K tersebut
* Lakukan hingga datatesting berakhir
* Lakukan akurasi

**Detail Rancangan Metode KNN**

Metode yang digunakan dalam masalah diatas adalah metode KNN. Yaitu sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasaran atribut dan training sample.

* Rancangan Input

Nilai k yang yang diambil adalah 25

* Proses Klasifikasi

Proses klasifikasi atau cross validation di lakukan dengan cara membagi data latih menjadi dua, yaitu menjadi data latih (data train) dan data validasi (data test).

Cross validation yang di lakukan adalah dengan cara K-Fold dengan nilai K-Fold nya adalah 8.

Jadi, data train yang sejumlah 4000 dibagi menjadi 8 sehingga masing-masing terdapat

500.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| LATIH\TESTING | 0-500 | 501-100 | 101-150 | 151-200 | 201-250 | 251-300 | 301-350 | 351-400 |
| 0-500 |  |  |  |  |  |  |  |  |
| 501-100 |  |  |  |  |  |  |  |  |
| 101-150 |  |  |  |  |  |  |  |  |
| 151-200 |  |  |  |  |  |  |  |  |
| 201-250 |  |  |  |  |  |  |  |  |
| 251-300 |  |  |  |  |  |  |  |  |
| 301-350 |  |  |  |  |  |  |  |  |
| 351-400 |  |  |  |  |  |  |  |  |

= DATA LATIH

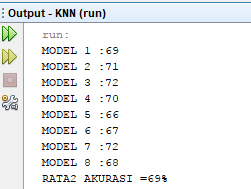
= DATA TESTING

Pronces Mencari Nilai K

Untuk mencari nilai k yang baik sehingga akan menghasilkan akurasi yang baik juga. Jadi data akan di bagi menjadi 8 untuk mengetest masing-masing k.

Dengan menggunakan cross validation, k yang digunakan di program ini adalah 25 untuk mendapatkan hasil akurasi yang bagus.

* **Output AKURASI**



Dilihat dari gambar diatas, nilai rata-rata akurasi terbesar yang di dapatkan adalah 69% dengan nilai k = 25 dan k-Fold = 8.

CATATAN:

* JAWABAN HOAX/TIDAK TERLAMPIR PADA EXCEL di folder

