**APLIKASI K-NEAREST NEIGHBOR (KNN)**



|  |
| --- |
| LAPORAN TUGAS BESAR  DIAJUKAN UNTUK MEMENUHI TUGAS PADA MATA KULIAH KECERDASAN BUATAN |



Disusun oleh:

Fiqqih Maulana Susanto 1301154178

|  |
| --- |
| PRODI S1 TEKNIK INFORMATIKA  FAKULTAS INFORMATIKA  UNIVERSITAS TELKOM  BANDUNG  2017 |

* **Deskripsi Masalah**

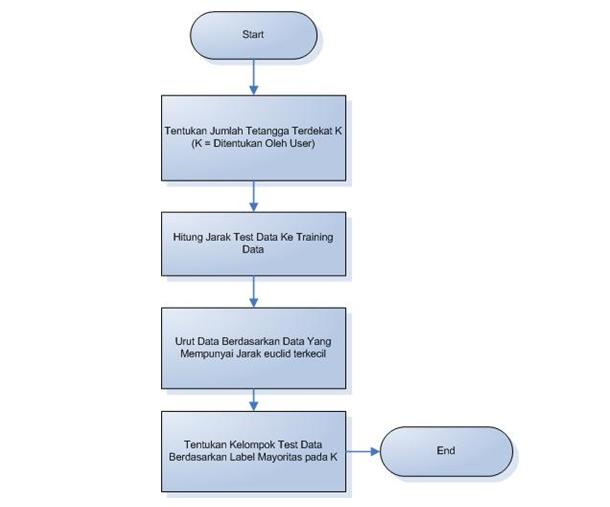
Permasalahan yang sedang dihadapi saat ini adalah bagaimana mengetahui status suatu berita dengan menggunakan metode KNN. K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru bedasarkan atribut dan training sample. Classifier tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik query, akan ditemukan sejumlah k obyek atau (titik training) yang paling dekat dengan titik query. Klasifikasi menggunakan voting terbanyak diantara klasifikasi dari k obyek.. algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari query instance yang baru.

Algoritma metode KNN sangatlah sederhana, bekerja berdasarkan jarak terpendek dari query instance ke training sample untuk menentukan KNN-nya. Training sample diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi training sample.

Pada fase training, algoritma ini hanya melakukan penyimpanan vektor – vektor fitur dan klasifikasi data training sample. Pada fase klasifikasi, fitur – fitur yang sama dihitung untuk testing data (klasifikasinya belum diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor training sample dihitung, dan sejumlah k buah yang paling dekat diambil. Titik yang baru klasifikasinya diprediksikan termasuk pada klasifikasi terbanyak dari titik – titik tersebut.

Ketepatan algoritma KNN oleh ada atau tidak adanya fitur – fitur yang tidak relevan, atau jika bobot fitur tersebut setara dengan relevansinya terhadap klasifikasi. Algoritma K Nearest Neighbor memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. Dari beberapa kelebihan tersebut, K-Nearest Neighbor juga memiliki kekurangan yaitu membutuhkan nilai K sebagai parameter, jarak dari data percobaan tidak dapat jelas dengan tipe jarak yang digunakan dan dengan atribut yang digunakan untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau hanya satu atribut yang telah pasti, dan perhitungan harga sangat tinggi karena percobaan ini membutuhkan perhitungan jarak dari beberapa query untuk semua data percobaan.

Adapun algoritma dari KNN ditunjukan pada flowchart berikut :



Adapula kelebihan dan kekurangan pada metode knn sebagai berikut :

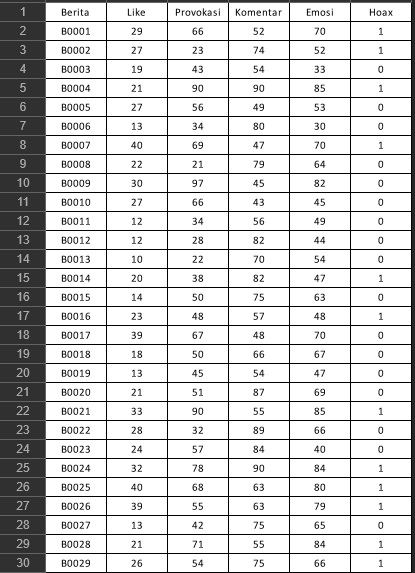
Kelebihan

* KNN memiliki beberapa kelebihan yaitu bahwa dia tangguh terhadap training data yang noisy dan efektif apabila data latih nya besar.

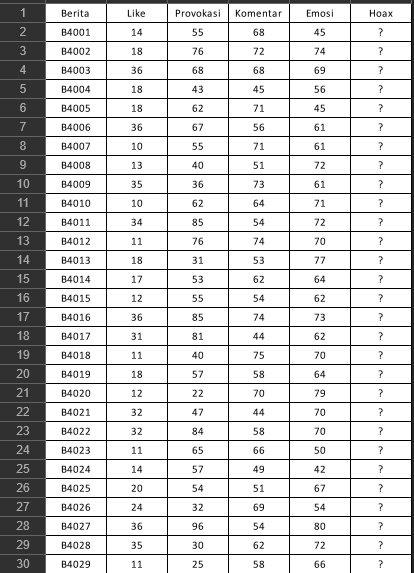
Kelemahan :

* KNN perlu menentukan nilai dari parameter K (jumlah dari tetangga terdekat).
* Pembelajaran berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil yang terbaik.
* **Rancangan Metode KNN**

Rancangan Input :



Tabel diatas adalah rancangan input dimana data diatas yang akan digunakan untuk menguji program yang kita buat, data diatas didapat dari soal yang telah ditentukan oleh pembuat, terdapat 4000 data training, gambar diatas adalah 30 teratas dari 4000 data yang diberikan oleh pembuat soal.



Tabel diatas adalah rancangan input dimana data diatas yang akan digunakan untuk menguji program yang kita buat, data diatas didapat dari soal yang telah ditentukan oleh pembuat, terdapat 1000 data testing, gambar diatas adalah 30 teratas dari 1000 data yang diberikan oleh pembuat soal.

Proses klasifikasi :

Menentukan parameter K :

Cara mencari parameter k yang saya gunakan adalah dengan mencari angka secara random lalu mencari persentase terbesar dari angka yang sudah kita pilih untuk dirandom. Angka tersebut adalah 24,94,364,924 dimana hasilnya dapat dilihat pada file training.csv.

Hitung jarak Antara data baru dengan semua data training :

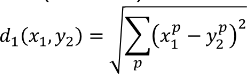
Dibawah ini adalah contoh data training :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Berita | Like | Provokasi | Komentar | Emosi |
| B0001 | 29 | 66 | 52 | 70 |
| B0002 | 27 | 23 | 74 | 52 |
| B0003 | 19 | 43 | 54 | 33 |

Dibawah ini adalah contoh data testing :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Berita | Like | Provokasi | Komentar | Emosi |
| B4001 | 14 | 55 | 68 | 45 |
| B4002 | 18 | 76 | 72 | 74 |
| B4003 | 36 | 68 | 68 | 69 |

Rumus untuk mencari jarak adalah dengan menggunakan rumus sebagai berikut :



Dimana data testing dimasukkan ke rumus diatas dan di test sebanyak data training sehingga menghasilkan jarak lalu diurutkan hasilnya mulai dari yang terkecil hingga yang terbesar, setelah itu data sebanyak k akan diambil sebagai parameter untuk menentukan data testing yang ditunjuk bernilai hoax atau tidak hoax.

**Referensi :**

<https://kuliahinformatika.wordpress.com/2010/02/13/buku-ta-k-nearest-neighbor-knn/>

<http://sharewy.blogspot.co.id/2013/04/algoritma-k-nearest-neighbor-knn_16.html>

<http://web.unair.ac.id/admin/file/f_41382_STKI-KEL-2_K-NEAREST-NEIGHBOR.pdf>

<https://drive.google.com/file/d/0B59Q6JwYvbzfMTVKN01XWjBaYlE/view>