**LAPORAN TUGAS PROGRAM**

**K-Nearest Neighbor**

*Disusun untuk memenuhi tugas mata kuliah Kecerdasan Buatan yang*

*diampu oleh :*

*Untari Novia Wisesty, S.T., M.T*



**Disusun Oleh :**

**Fero Resyanto**

**1301154318**

**IF-39-10**

**JURUSAN S1 TEKNIK INFORMATIKA**

**FAKULTAS INFORMATIKA**

**UNIVERSITAS TELKOM**

**BANDUNG**

**2017**

Daftar Isi

[**Deskripsi Masalah** 3](#_Toc500011710)

[**Rancangan Metode** 3](#_Toc500011711)

[**1.** **Rancangan Input** 3](#_Toc500011712)

[**2.** **Pemilihan Nilai K** 5](#_Toc500011713)

[**Referensi** 6](#_Toc500011714)

# **Deskripsi Masalah**

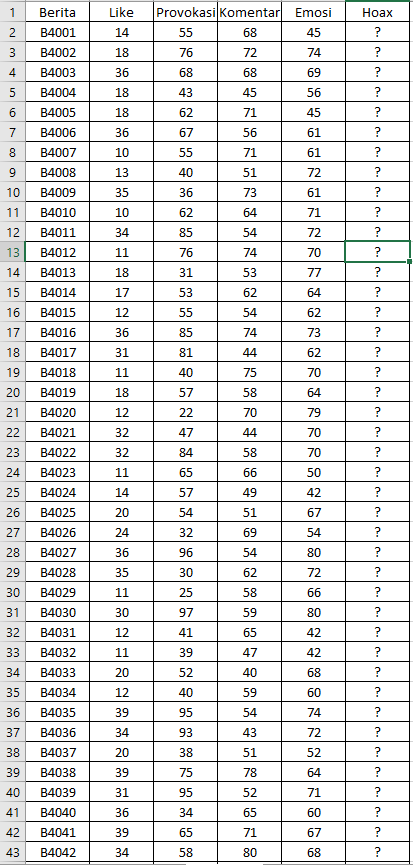
Pada kasus kali ini, kita menghadapi sebuah masalah yaitu bagaimana mengetahui bahwa suatu berita yang di publish merupakan sebuah berita hoax atau bukan. Untuk mengetahui hal tersebut kita menggunakan sistem *KNN ( K-Nearest Neighbour )* . *KNN* ialah suatu algoritma kecerdasan buatan dan suatu metode yang menggunakan algoritma *supervised* dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada *KNN*. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasakan atribut dan sampel latih. pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik uji, akan ditemukan sejumlah K objek (titik training) yang paling dekat dengan titik uji. Klasifikasi menggunakan voting terbanyak di antara klasifikasi dari K objek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sample uji yang baru.

Pada kasus ini terdapat 4 nilai masukan yaitu nilai *like*, *provokasi*, *komentar* dan *emosi*. Rentang nilai *like*, *provokasi*, *komentar* dan *emosi* yaitu 0 hingga 100. Pada sistem *KNN*  terdapat

# **Rancangan Metode**

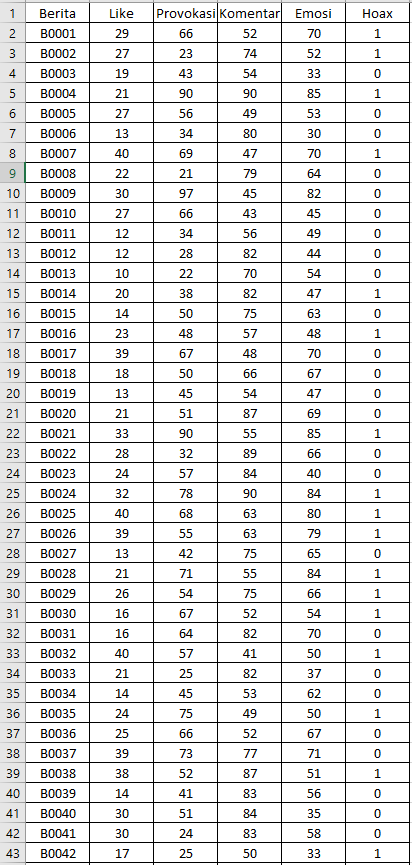
## **Rancangan Input**

Pada kasus ini digunakan table yang berisi nilai input untuk emosi, like, komentar dan provokasinya yaitu sebagai berikut :



Tabel diatas hanya menampilkan 43 data pertama. Untuk melihat lebih lengkapnya, data dapat dilihat pada file : Datatest.csv ( buka dengan ms.excel )

Sedangkan untuk tabel yang digunakan sebagai dataset, yaitu sebagai berikut :



Tabel diatas hanya menampilkan 43 data pertama. Untuk melihat lebih lengkapnya, data dapat dilihat pada file : Dataset.csv ( buka dengan ms.excel )

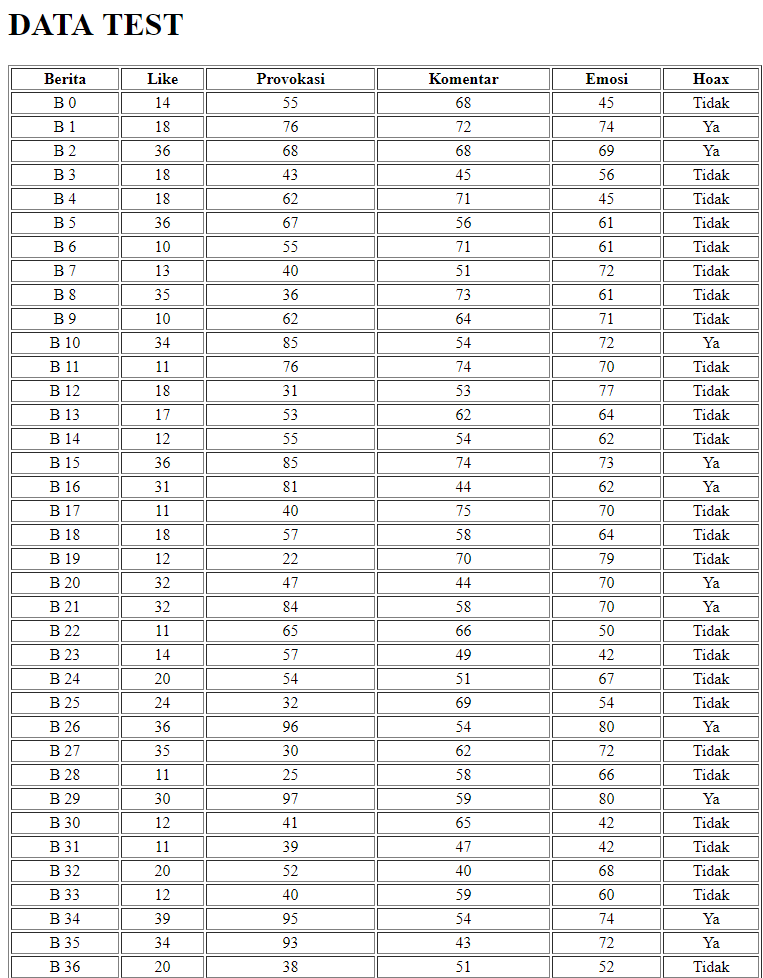
## **Pemilihan Nilai K**

Pemilihan nilai K pada kasus ini dilakukan dengan cara mencoba-coba nilai K mana yang dapat memperkecil nilai error rate. Nilai K dipilih dengan mengambil 1000 data pada dataset ( data 1 – data 1000 ) dan dijadikan datatest. Kemudian dilakukan percobaan dengan menklasifikasi data pada validation-set dengan K yang sudah ditentukan. Lalu catat score yang diperoleh. Pemilihan nilai K dilakukan sebanyak 10 kali, yaitu dengan nilai : 3 , 5, 7, 9, 11, 13, 15, 17, 19, 21. Dari 10 kali percobaan didapatkan nilai K = 21, dengan score tertinggi yaitu 703. Sehingga nilai K yang akan digunakan adalah 21 untuk menguji data tes yang sesungguhnya.

Rumus yang digunakan mencari jarak yaitu sebagai berikut :

Untuk setiap satu data testing akan dimasukan kedalam rumus tersebut dan dilakukan kalkulasi dengan seluruh data training . Hasil dari kalkulasi tersebut akan diurutkan dari yang terkecil, lalu diambil sebanyak K buah ( yang telah ditentukan sebelumnya ) dan dijadikan parameter untuk menentukan apakah data atau berita tersebut merupakan berita hoax atau bukan.

Screenshot hasil program :



Tabel diatas hanya menampilkan 37 berita, untuk lebih lengkapnya, anda dapat merunning program knn.php atau membuka file excel Hasil\_KNN.xlsx yang telah dilampirkan pada tugas ini.

# **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>