

# Pembuatan Sistem Klasifikasi Menggunakan KNN untuk Menentukan Kelas

Muh. Hasbi Abdul M – 1301160335

Program Studi Teknik Informatika, Fakultas Informatika  
Universitas Telkom Bandung  
Jl. Telekomunikasi Dayeuhkolot, Bandung 40257

## I. PENDAHULUAN

*K-Nearest Neighbors* adalah algoritma *supervised learning* dimana hasil instansi / kelas baru diklasifikasikan berdasarkan data mayoritas instansi yang ada dari k-tetangga terdekat dengan objek yang akan di klasifikasi.

Tujuan dari algoritma KNN ini untuk mengklasifikasikan objek baru berdasarkan atribut dari *data train*.

## II. MASALAH

Menentukan kategori / kelas dari sebuah file *DataTest\_Tugas3\_AI.csv* yang berisikan 200 data dengan 5 atribut yaitu  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ , dan  $X_5$ . Dengan menggunakan Algoritma KNN diminta untuk membuat sistem yang dapat menentukan kelas dari 200 data tersebut berdasarkan data pada *DataTrain\_Tugas3\_AI.csv* yang berisikan 800 *data train*.

## III. ANALISIS PERMASALAHAN

Pada metode ini terdapat beberapa tahap yang harus dilakukan, yaitu

Sebelum memulai tahap pertama hal yang harus dilakukan adalah memasukkan *data train* & *test* yang ada ke dalam program.

Tahap pertama menentukan nilai K, pada algoritma KNN nilai K diharuskan bernilai ganjil agar tidak terjadi kerancuan saat nanti menghitung mayoritas. Pada kasus ini saya menggunakan nilai  $k=11$  karena berdasarkan apa yang saya baca prediksi kelas akan bagus jika nilai k

berkisar di 11-15 walaupun pun itu belum tentu cocok dengan karakteristik data yang ada.

Tahap kedua menghitung jarak euclidian dari setiap data test ke setiap data train. Pada tahap ini mencari jarak dari 1 data test ke 800 data train dimana dalam program data tersebut ditampung di array *tmpHasil* yang mempunyai 2 atribut yaitu jarak dan kelas dari *data train*.

$$= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Gambar 1: Rumus Menghitung Jarak Euclidian

Tahap ketiga adalah mengurutkan *data train* dari yang terdekat dengan *data test* dan menentukan tetangga mana yang terdekat berdasarkan nilai K pada program yang saya buat array *tmpHasil* akan diurutkan secara *ascending* berdasarkan jarak terdekat.

Tahap terakhir yaitu menentukan prediksi kategori data test berdasarkan kategori mayoritas tetangga terdekat. Pada program saya membuat array hitung untuk menghitung mayoritas kategori k-tetangga terdekat lalu mengurutkannya secara *descending* berdasarkan kategori terbanyak. Setelah itu maka hasil kategori terbanyak akan dimasukan ke array *dataTest.y*.

Setelah hasil ditemukan maka hasil yang berupa kelas untuk setiap 200 *data test* (*dataTest.y*) dituliskan kedalam file *HasilTebakanTugas3.csv*

#### IV. REFERENSI

- [1] Ismail, Maulana. Agustus 2017. "Cara Kerja k-Nearest Neighbor (k-NN)". <https://medium.com/@asepmaulanaismail/cara-kerja-algoritma-k-nearest-neighbor-k-nn-389297de543e>
- [2] Suyanto. 2014. Artificial Intelligence. Revisi Kedua. Bandung : Penerbit Informatika