Nama : Shidqi Aqil Naufal Kelas : IF 40-08 Nim : 1301164147

**Laporan Pencarian Kelas pada suatu data dengan metode K- Nearest Neighbour (KNN)**

**Deskripsi Masalah**

Terdapat 2 buah data yang diberikan yaitu data train data test. Data train terdiri dari 5 atribut input (x1, x2, x3, x4, dan x5) dan 1 output yag memiliki 4 kelas. Sedangkan data test, mempunyai 5 atribut yang sama dengan data test, namun kelas nya belum ditentukan. Jadi diperlukan sebuah model/sistem klasifikasi untuk menentukan kelas dari data test yang telah disediakan dengan metode KNN.

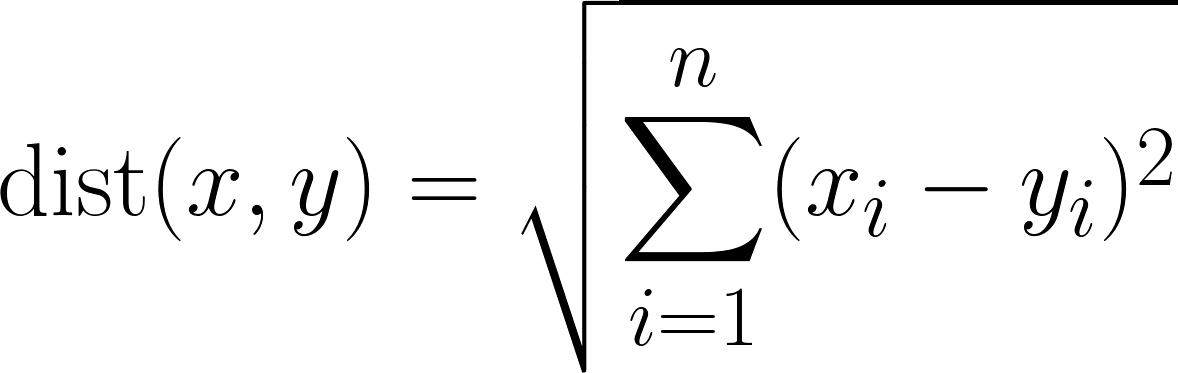
**Metode Penyelesaian**

K- Nearest neighbor (KNN) adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya (data train).

Adapaun langkah langkah dari pengimplementasian Algoritma K-Neatrest Neighbor adalh sebagai berikut**:**

1. **Mengehitung kuadrat jarak Euclidean**

Dalam menghitung jarak euclidiean, digunakan rumus sebagai berikut :



Pada perhitungan jarak eucledian, akan dilakukan perhitungan pada data testing dan data training. Sebagai contoh, xi adalah atribut pada data testing, dan yi adalh atribut pada data training. Maka akan dilakukan perhitungan seperti diatas. Perhitungan 1 data testing akan dilakukan pengulangan perhitungan sebanyak data yang terdapat di train.

1. **Melakukan Sorting secara Ascending**

Setelah mendapatkan jarak pada setiap datanya, maka akan dilakukan sorting secara ascending. Yang mana data yang memiliki jarak yang lebih kecil, akan berada diatas. Hal ini menunjukkan bahwa data tersebut memiliki jarak ketetanggaan yang dekat dengan data testing yang dihitung.

1. **Menentukan Parameter k**

Dalam menentukan parameter k, dapat ditentukan random. Namun dalam kasus ini, dilakukan pengecekan akurasi terlebih dahulu untuk menentukan nilai k.

Pengecekan akurasi dilakukan dengan membagi data train. Pembagaian dilakukan menjadi dua bagian, yaitu data train akan dibagi menjadi data train itu sendiri dan data validasi. Pembagian dua bagian dilakukan sebesar 70% dan 30 %. Yang mana 70% sebagai data training, dan 30% sebagai data validasi.

Setelah dilakukan validasi data ditemukan akurasi dari sistem yang dibuat yaitu dengan parameter k=6.

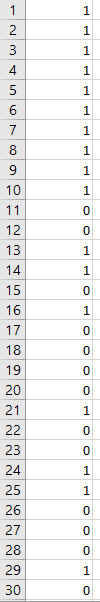
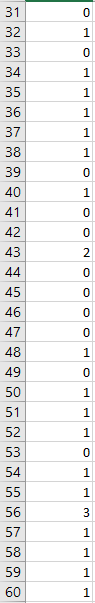
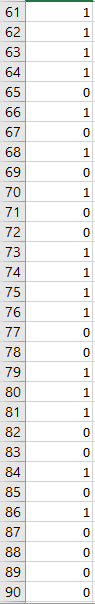
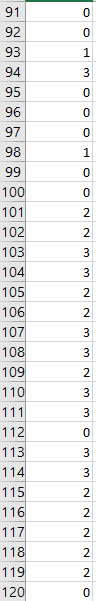
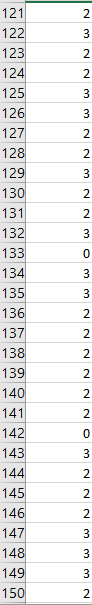
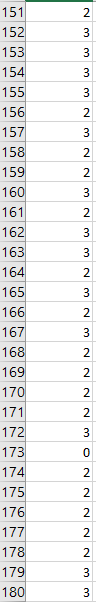
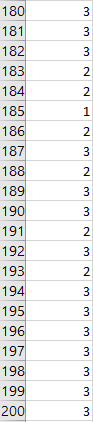
1. **Mengumpulkan kelas (label) berdasarkan k yang diambil**

Setelah menentukan k, dalam kasus ini diambil k=6. Maka 6 data awal setelah dilakukan sorting, akan diambil untuk ke langkah selanjutnya.

1. **Kategori/label yang paling mayoritas dijadikan prediksi label untuk data test**

Setelah mengambil data sebanyak k (dalamkasus ini 6 data) maka dilakukan perhitungan jumlah permasing masing label. Yang mana jumlah label yang paling banyak(mayoritas) akan menjadi label untuk data test yang diuji.

Berikut adalah hasil dari proses knn yang telah dibuat. Hasil berupa kelas pada setiap index di data test.

Berdasarkan hasil tersebit, dapat disimpulkan, salah satu parameter terpenting dalam metode KNN adalahpengambilan nilai k. karena pengambilan nilai k yang tepat, akan menghasilkan akurasi dari prediksi yang tinggi.