**Tugas 3**

**Nama : Cindy Alifia Putri**

**NIM : 1301160199**

Langkah-langkah :

1. import semua library dan juga file dataTrain.csv dan dataValidasi.csv
2. Membuat function euclidean yang akan mengeluarkan return jarak dua titik
3. Tentukan nilai k (akan dijelaskan selanjutnya)
4. Membuat ***function knn*** yang memiliki 3 parameter yaitu nilai k, tes, dan juga train. Yang mana tes akan menjadi dataTest dan train akan menjadi list dataTrain
   1. Lakukan brute force untuk 1 data tes ke semua data train. Hasil brute force akan disimpan kedalam array distance dengan kolom 1 adalah data tes ke – i, kolom 2 adalah nilai euclidnya, dan kolom ke tiga adalah nilai class dari data train ke – i. Maka list distance akan ada sebanyak panjang data tes x panjang data train. Lalu ubah tipe data list distance menjadi numpy array, tujuannya agar data dapat di bagi-bagi menjadi beberapa bagian.
   2. Lakukan pemisahan data tes ke –i di assaign kedalam arr[i]
   3. Setiap nilai pada arr[i] di urutkan, sehingga nanti akan diperoleh nilai arr[i] yang isinya sudah terurut ascending. Setiap data yang sudah terurut dimasukkan kedalam arSort[i].
   4. Lakukan brute force untuk semua data tes, lalu didalamnya lakukan perulangan sebanyak nilai k yang tujuannya adalah menambahkan k nilai pertama disetiap arr[i] yang selanjutnya akan disimpan kedalam arSort. Maka arSort[i] akan berisi data sebanyak k.
   5. Lakukan brute force kembali untuk tiap data tes yang didalamnya akan ada perulangan kembali sebanyak nilai k dengan tujuan membuat array baru yang isinya hanyalah klasifikasi class. Klasifikasi class tersebut disimpan kedalam list klasifikasiClass yang tiap indeks ke i berisi k nilai lalu berikan function mode untuk mengambil nilai modus dari nilai class sebanyak k tersebut, jadi klasifikasiClass akan memuat klasifikasi class yang paling banyak muncul sebanyak k.
   6. Lakukan return klasfikiasiClass tersebut
5. Panggil function knn dengan 3 parameter (k, dataTest, dataTrain) dan simpan kedalam klasifikasiArray. Lalu simpan array tersebut kedalam sebuah file

Langkah-langkah mencari k :

1. Membuat ***function cariNilaiK*** yang memiliki dua parameter , yaitu dataValidasi dan juga dataTrainValidasi
   1. Lakukan brute force sebanyak panjang dataValidasi yang didalamnya akan memanggil function knn dengan parameter perulangan ke-i+1, dataValidasi, dan dataTrainValidasi yang selanjutnya akan disimpan kedalam array s. Dimana dataTest dan data dataTrainValidasi akan dijelaskan di tahap selanjutnya. untuk nilai k, saya memakai data ke i+1 dikarenakan apabila data dimulai dari 0, maka numpy akan mereturn nilai 0 diman untuk numpy versi baru, return 0 sudah diganti menjadi return false. Yang kedepannya akan menyebabkan program saya error. Maka untuk mencegah hal itu terjadi, saya memakai k yang dimulai dengan i+1. Lalu selanjutnya lakukan brute force kembali sebanyak panjang dari dataValidasi yang didalamya akan membandingkan seberapa banyak klasifikasi Class pada s yang sama dengan klasifikasi class yang sebenarnya. Setelah ditemukan banyaknya data yang sama ada berapa, maka jumlah tersebut akan disimpan kedalam array arrayDataValidasi.
   2. Lalu selanjutnya akan dicari nilai maximum dari arrayDataValidasi tersebut yang selanjutnya akan disimpan kedalam nilaimaks.
   3. Return nilaimaks.
2. Lakukan split sebanyak 10 dataTrain. Trik nya adalah kita akan melakukan perulangan sebanyak 10 kali, dikarenakan dataTrain kita ada 800 data maka akan ada 800:10 = 80 data yang akan di split. Kenapa saya memakai 10 split data ? karena apabila data di split 200 maka akurasi yang saya dapatkan masih jauh dari benar.

Kemudian lakukan perulangan sebanyak 80 data. Didalam perulangan itu, akan dibuat dataValidasi dan dataTrainValidasi yang didapatkan dari data yang telah displit, untuk perulangan pertama maka akan ada 80 data sebagai dataValidasi dimulai dari dataTrain[0]sampai dataTrain[79] dan 640 dataTrainValidasi dimulai dari dataTrain[80] sampai dataTrain[799]. Perulangan selanjutnya akan ada 80 data sebagai dataValidasi dimulai dari dataTrain[80]sampai dataTrain[159] dan 640 dataTrainValidasi yang memuat semua nilai dataTrain kecuali nilai pada dataValidasi.

Setelah itu, panggil function cariNilaiK dengan parameter dataValidasi dan dataTrainValidasi sesuai dengan yang sudah saya split sebelutmnya, lalu di hasilnya akan disimpan kedalam array arrayValidasi.

Setelah kita mendapat arrayValidasi, panggil function max untuk mencari data terbesar pada kolom ke 1 kegunaannya adalah mencari data terbesar agar didapat akurasi paling banyak. Sehingga ditemukan k = 3 dengan akurasi sebesar 92,25%. Karena apabila k = 3 maka akan terdapat 74 data yang sama dari 80 data.