Laporan Tugas 3 Pengantar Kecerdasan Buatan
"k-Nearest Neighbour"

Nama : Rayhan Rahmanda
NIM : 1301184233
Kelas : IF-42-04

A. Penjelasan Tugas

Diberikan dataset (himpunan data) Pima India Diabetes Dataset (PIDD) pada file "Diabetes.csv". Dataset tersebut berisi 768 objek data (baris). Buatlah lima datasets baru menggunakan skema 5-fold cross-validation. Pertama, bagi objek data ke dalam lima subsets (sub himpunan) dengan porsi yang sama, masing-masing berisi sat per lima (20%) data. Kemudian, buat lima dataset baru dengan komposisi objek-objek data pada training set (data latih) dan testing set (data uji) seperti pada soal. Lakukan analisis, desain, dan implementasi algoritma k-nearest neighbour (kNN) ke dalam suatu program komputer. Lakukan seleksi dan estimasi model kNN tersebut menggunakan 5-fold cross-validation yang menghasilkan akurasi tertinggi.

B. Observasi

1. k-Nearest Neighbour

   k-Nearest Neighbour (kNN) merupakan salah satu metode dalam teknik Instance-Based Learning (IBL). kNN dikenal sebagai lazy learner atau pelajar malas karena kNN tidak melakukan proses belajar (dari data latih), melainkan secara langsung melakukan klasifikasi berdasarkan sejumlah tetangga, yang memiliki jarak terdekat dengan pola masukan. Karena berbasis jarak, maka bagian paling penting dalam kNN adalah pemilihan **formula jarak atau dissimilarity**. Formula jarak yang kurang tepat membuat kNN kurang akurat.

2. Fungsi/Prosedur

def createDataset(x,y)	Fungsi ini dibuat untuk membuat lima buah subsets yang berisikan Training set dan Test set.
def euclidian(row1,row2)	Fungsi ini dibuat untuk menghitung jarak Euclidian antar baris.
def prepro()	Fungsi ini digunakan untuk Pre-Processing data dengan mendapatkan informasi dari Diabetes.csv dan mengisi kolom yang bernilai 0 menjadi nilai Mean dari kolom tersebut.
def neigh(datatrain,baris_test,k)	Fungsi ini dibuat untuk menghitung serta mencari nilai ketetanggaan, yang di dalam nya menggunakan perhitungan Euclidian.
def classification(tatanggi)	Fungsi ini dilakukan untuk memprediksi nilai outcome suatu baris hasil dari prosedur def neigh(), nilai yang dihasilkan 0 atau 1 (non-diabetic dan diabetic).
def accuracy(actual,predicted)	Fungsi ini digunakan untuk menghitung tingkat akurasi antar

	baris data yang diprediksi dengan baris data yang sudah ada.
def bestK(avgsub,avg)	Fungsi dibuat untuk mencari nilai K terbaik.
def start(dataset,n_neighbor)	Prosedur untuk memulai program.

C. Output Program

Program akan menghasilkan nilai K yang terbaik beserta dengan persentase akurasi nya.

```
K for testing :  [31, 33]
Highest Accuracy : 75.2584670231729% with K =  31
>>>
```