**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS TEKNOLOGI YOGYAKARTA**

**UJIAN TENGAH SEMESTER GENAP TAHUN AKADEMIK 2022/2023**

PROGRAM STUDI : S1 INFORMATIKA (Kelas H)

Mata Ujian : Machine Learning

Hari/Tanggal : Selasa, 4 April 2023

Waktu : 13.20 – 15.00 WIB

Durasi : 100 menit

Dosen : Moh. Ali Romli, S.Kom., M.Kom.

Sifat : *Take Home*, *Open Book, Open Internet*

Tidak diperkenankan berdiskusi dengan siapapun

**PERATURAN**

1. Jawaban untuk setiap soal pada UTS ini dibuat menggunakan aplikasi pengolah kata (Ms. Word atau semacamnya) lalu hasilnya disimpan dalam format PDF. Cantumkan juga kode program yang Anda buat di dalam lembar jawaban tersebut pada setiap nomor soal.
2. Uraikanlah jawaban Anda dengan lengkap dan sistematis. Jawaban harus dibuat berurutan sesuai nomor urut soal.
3. Tidak diperkenankan berdiskusi dengan siapapun, namun Anda diperbolehkan membuka buku atau sumber belajar dari internet.
4. Jawaban UTS dalam format PDF yang telah disertai dengan kode program dicetak pada kertas A4 dan dikumpulkan sesuai jadwal ujian, yakni Hari Selasa 4 April 2023 pukul 13.20 WIB di ruangan yang telah ditentukan panitia.
5. Lembar jawaban yang tidak ditandatangani tidak akan diperiksa dan akan diberikan nilai 0.

**SOAL**

Anda diberikan sebuah dataset untuk klasifikasi tanaman Iris (Iris Spesies) yang berisi 150 record (baris) dengan 6 atribut (kolom) yang terdiri dari 5 atribut reguler dan 1 atribut label (‘Spesies’) yang dapat Anda unduh pada URL berikut : <https://www.kaggle.com/datasets/uciml/iris>. Berikut ini deskripsi atribut dari dataset yang dipakai yang terdiri dari nama atribut, dan tipe datanya:

1. **Id |** tipe : integer
2. **SepalLengthCm** | tipe : float
3. **SepalWidthCm** | tipe : float
4. **PetalLengthCm** | tipe : float
5. **PetalWidthCm** | tipe : float
6. **Species** | tipe : object **(label)**

Berbekal dataset tersebut, tuliskanlah kode program dengan Bahasa Python untuk :

1. Membaca dataset tersebut
2. Memisahkan antara fitur (x) dan label (y) dengan ketentuan mengambil semua data kolom dengan mengecualikan atribut ‘**Id**’ dan mengambil semua data baris
3. Membagi dataset menjadi data latih dan data uji dengan rasio 70 : 30 dan random\_state=1.
4. Lakukan proses untuk menormalisasi fitur
5. Lakukan proses untuk melatih model KNN dengan ketentuan nilai k=5
6. Lakukan proses untuk memprediksi pada data uji
7. Hitung akurasi model

===============SELAMAT MENGERJAKAN===============

**Lembar Jawaban UTS**

**Machine Learning**

**Nama : Farhan Mualif**

**NIM : 5210411219**

**Kelas : H**

**Pernyataan:**

Dengan mengerjakan ujian ini, maka saya menyatakan bahwa semua jawaban **SAYA** **KERJAKAN SENDIRI** tanpa bekerjasama maupun meminta bantuan siapapun. Saya bersedia diberi **nilai E** jika pernyataan ini terbukti salah.

Tanda tangan:

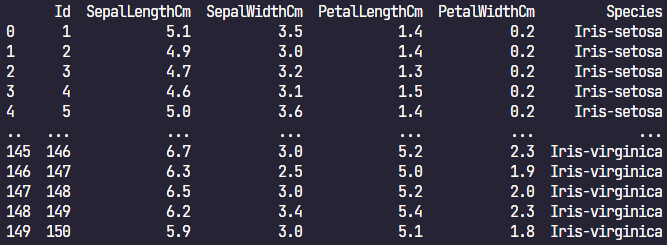
1. Untuk mengambil data dalam format csv kita menggunakan perintah read\_csv(‘nama\_file’), selanjutnya perintah tersebut kita tampung dengan variabel df, lalu untuk menampilkan data tersebut kita menggunakan perintah print(df). Berikut code dan output yang dihasilkan:

Code:

A screenshot of a computer

Description automatically generated with medium confidence

Output:



1. Untuk mengambil fitur (x) kita melakukan perintah code berikut:

Graphical user interface, text, website

Description automatically generated

Arti dari code diatas adalah menghapus kolom Id, dan Species, dan axist=1 memiliki arti bahwa yang akan kita hapus adalah kolom, setelah kita melakukan hal tersebut maka kolom yang tersisa adalah ['SepalLengthCm', 'SepalWidthCm','PetalLengthCm' ,'PetalWidthCm']. Selanjutnya untuk mengambil label (y), kita cukup menggunakan perintah code berikut:

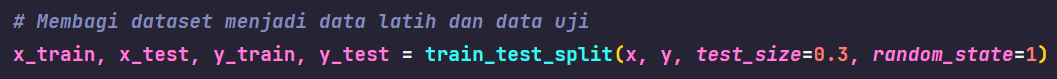


Code diatas berarti kita hanya mengambil data pada kolom Species saja

1. Sebelum kita melakukan mambagi data, kita harus mengimport module *train\_test\_split* terlebih dahulu seperti code berikut



untuk membagi data latih dengan data uji kita menggunakan perintah code berikut:

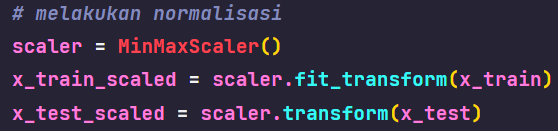


Variabel x\_train dan y\_train berguna untuk menampung data uji, sedangkan y\_train dan x\_train berguna untuk menampung data uji, selanjutnya untuk parameter test\_size berguna untuk menentukan data uji sebesar 0.3 atau 30%

1. Sebelum melakukan normalisasi kita harus import module *MinMaxScaler* terlebih dahulu



Selanjutnya kita melakukan code seperti berikut:

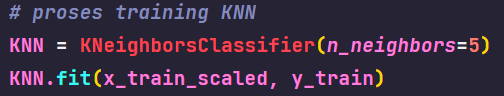


Variabel *scaler* berguna untuk menampung MinMaxScaler, variabel *x\_train\_scaled* berguna untuk menampung data latih yang akan di normalisasi, dan *x\_test\_scaled* berguna untuk menampung data uji yang akan dinormalisasi

1. Sebelum melakukan proses untuk melatih model KNN kita harus melakukan import module *KneighborsClassifier* dengan cara berikut:

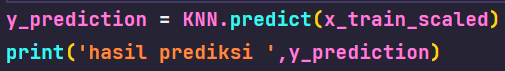


Setelah melakukan import selanjutnya lakukan coding berikut:



Variabel KNN diatas berguna untuk menampung fungsi KneighborsClassifier dengan parameter n\_neighbors=5 yang berarti saat melakukan klasifikasi, model akan mencari 5 tetangga terdekat untuk digunakan sebagai acuan dalam menentukan label kelas untuk data yang sedang diuji. Selanjutnya *KNN.fit(x\_train\_scaled, y\_train)* digunakan untuk melatih model KNN dengan parameter pertama yaitu *x\_tranin\_scaled* yang merupakan data latih yang telah dinormalisasi dan parameter kedua yaitu *y\_train* yang merupakan data kelas dari data latih.

1. Untuk melakukan prediksi label kelas, lakukan code seperti berikut:



Pada fungsi predict diatas terdapat paramater yang merupakan data latih yang sebelumnya sudah di normalisasi. Berikut merupakan output hasil prediksi yang berupa label yang telah ditentukan berdasarkan algoritmaKNN:

Graphical user interface, text

Description automatically generated

1. Sebelum melakukan penghitungan aurasi model, kita diharuskan untuk melakukan import modul accuracy\_score terlebih dahulu seperti pada code berikut:



Setelah melakukan import lakukan code seperti berikut:

Text

Description automatically generated

fungsi accuracy\_score menerima dua parameter, parameter pertama adalah variabel y\_test yang mana variabel tersebut adalah data uji, parameter kedua merupakan variable y\_prediction yang mana variable tersebut merupakan data hasil prediksi dari algoritma KNN.

Berikut merupakan code secara lengkap:

Text

Description automatically generated

Dan berikut output dari program diatas:

Text

Description automatically generated