﻿**Name: Abdulkabir Ojulari Tola**

**Assignment: Optional**

# import libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.tree import DecisionTreeClassifier

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import confusion\_matrix

from sklearn.metrics import accuracy\_score

dataset = pd.read\_csv('datasets\_19\_420\_Iris.csv')

# supervised learning algorithm

X = dataset[['SepalLengthCm', 'SepalWidthCm','PetalLengthCm', 'PetalWidthCm']]

Y = dataset.Species

# Splitting the dataset into 2 in proportion of 3:7

X\_Train, X\_Test, Y\_Train, Y\_Test = train\_test\_split(X, Y,test\_size=.3)

model = DecisionTreeClassifier() # building the model

fittedModel = model.fit(X\_Train, Y\_Train) # fitting the model with training sets for classification

# predictions for classification

predictions = fittedModel.predict(X\_Test)

print(confusion\_matrix(Y\_Test, predictions))

# accuracy score to check the efficiency of the model

print(accuracy\_score(Y\_Test, predictions)) # ascertain the percentage of the that is classfied correctly

**Libraries**

**Scikit**-learn is a free machine learning library for **Python** with several algorithms such as Decision Tree, Regression etc.

It’s one of the important libraries from which other required sub-libraries are imported.

Decision Tree Classifier was imported from Sklearn and it is a non-parametric supervised learning method used for classification and regression but in this code, we focused on classification.

Train Test Split library from Sklearn is to help split the dataset into train set and test set which also have test\_size e.g. 30:70

Confusion Matrix and Accuracy Score are essentially imported to measure the accuracy of the algorithm.