import numpy as np  
import pandas as pd  
# import seaborn as sns  
# import matplotlib.pyplot as plt

# from sklearn.linear\_model import LinearRegression  
from sklearn.datasets import make\_regression  
from sklearn.model\_selection import train\_test\_split

iris = pd.read\_csv('iris.csv')  
iris.sample(10)

iris.describe()

iris.info()

iris['Species'].value\_counts()

# perform label encoding  
from sklearn.preprocessing import LabelEncoder  
le = LabelEncoder()  
# transform the string labels to integer  
iris['Species'] = le.fit\_transform(iris['Species'].to\_numpy())  
iris.sample(10)

# input data  
X = iris.drop(columns=['Species'])  
# output data  
Y = iris['Species']  
X, Y

from sklearn.model\_selection import train\_test\_split  
## train - 80%  
## test - 20%  
# split the data for train and test  
x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, Y, test\_size=0.20, random\_state=12345) 123456

# knn - k-nearest neighbours  
# from sklearn.neighbors import KNeighborsClassifier  
# model = KNeighborsClassifier()

# decision tree  
from sklearn.tree import DecisionTreeClassifier  
model = DecisionTreeClassifier()  
model.fit(x\_train, y\_train)

# predict  
y\_pred = model.predict(x\_test)  
y\_pred, y\_test

from sklearn.metrics import confusion\_matrix  
from sklearn.metrics import precision\_score, recall\_score, f1\_score, accuracy\_score  
# print metric to get performance  
print("Accuracy: ", model.score(x\_test, y\_test))

conf\_matrix = confusion\_matrix(y\_true=y\_test, y\_pred=y\_pred)  
conf\_matrix