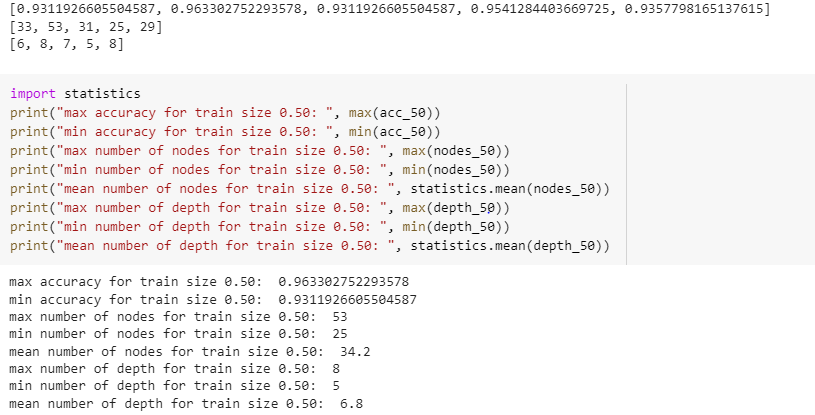
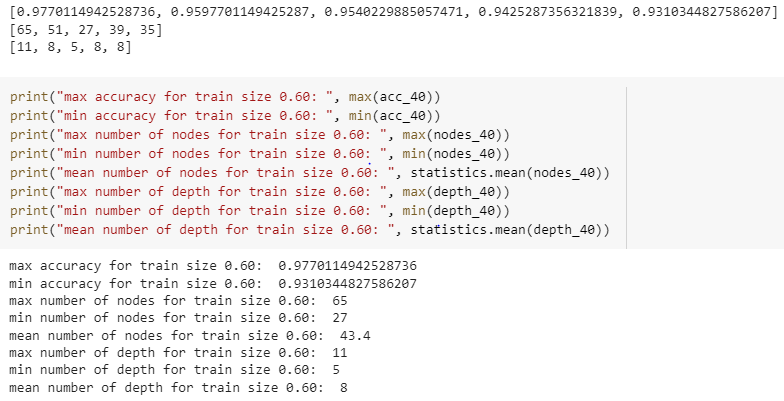
**Decision tree documentation:**

1. read data as text file using read\_csv() function and use parameters sep=’,’ to make it able making data in dataframe format, pass a list of features name to names parameter and use na\_values='?' to make any ‘?’ with null value to can handle it
2. Check for null “data contain many nulls” and handling it using mode for each feature
3. Change values of features from “n,y” to “0,1” using replace function and dictionary
4. Splitting data into inputs(x) and outputs(y)
5. Split x,y to train and test using train size [50,60,70,75] and different 5 random states for each training size and create 3 lists one for storing 5 acc for each training size ,one for storing the 5 depths for each training size another for storing 5 nodes for each training size

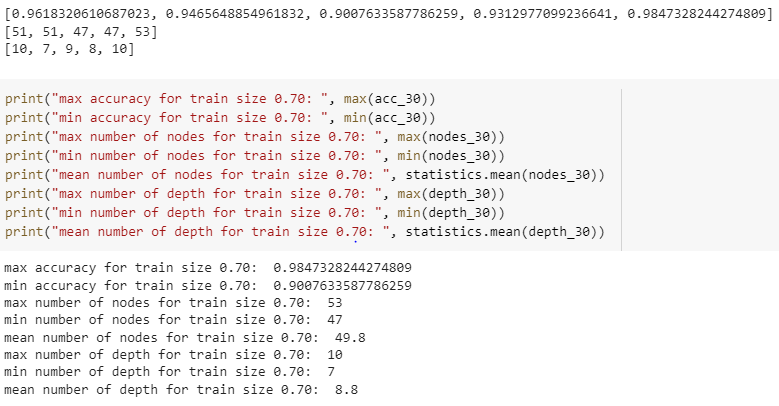
* For training size 50 and random state[1,2,3,4,5]:



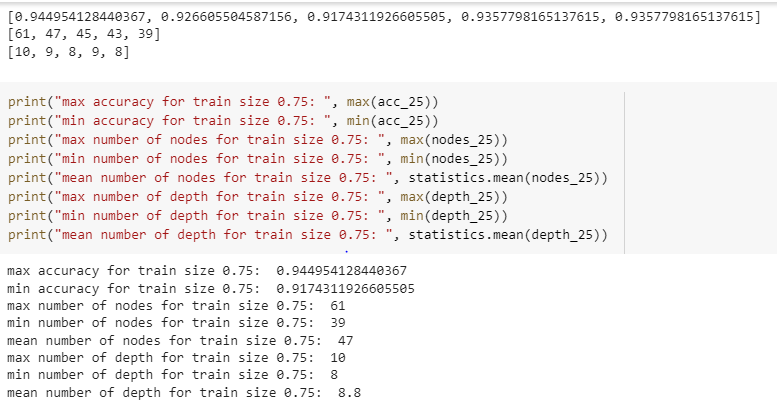
* For training size 60 and random state[3,5,7,9,11]:



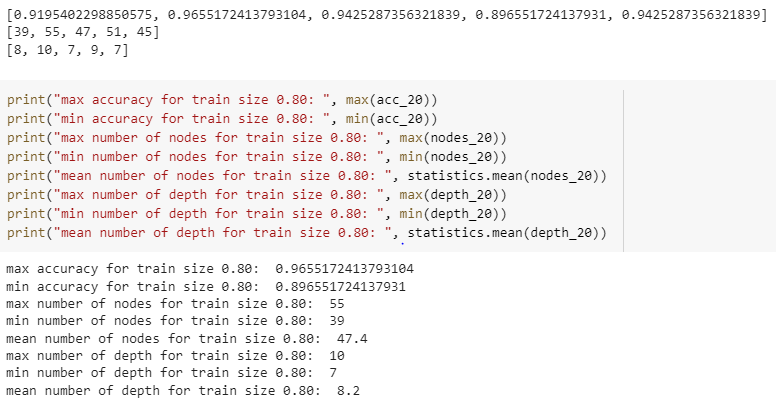
* For training size 70 and random state[9,18,27,36,45]:



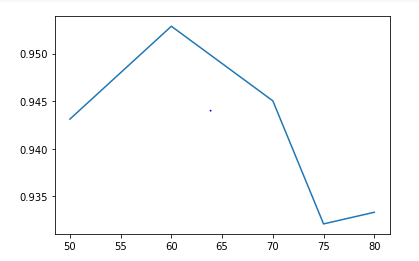
* **For training size 75 and random state[3,6,9,12,15]:**



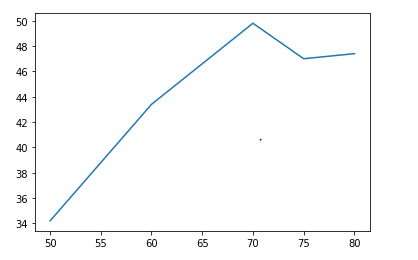
* For training size 70 and random state[9,18,27,36,45]:



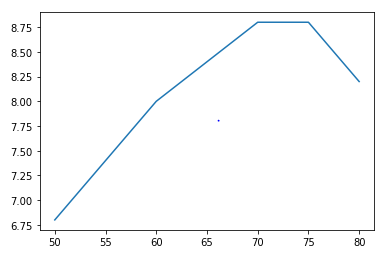
* **Ploting for avg acc and train size**



* **Ploting for avg nodes number and train size**



* **Ploting for avg depth and train size**



**Used libraries:**

import pandas as pd

import numpy as np

from matplotlib import pyplot as plt

import seaborn as sn

import warnings

warnings.filterwarnings('ignore')

import math

from google.colab import drive

import statistics