**31thjan ’21 DECISION TREE CLASSIFICATION MODEL**

**NAME:**SARIKA M

**AIM:**

To build a Decision Tree model for prediction of car quality given other attributes about the car.

**VARIABLES USED :**

buying - buying price

maintaince - price for the maintance

doors - no.of.doors

persons - no.of. persons

lug\_boot - the size of luggageboot

safety - estimated safety of the car

class - class distribution ('unacc', 'acc', 'vgood', 'good')

**MODEL:**

**IMPORTING THE LIBRARIES:**

import numpy as np

import pandas as pd

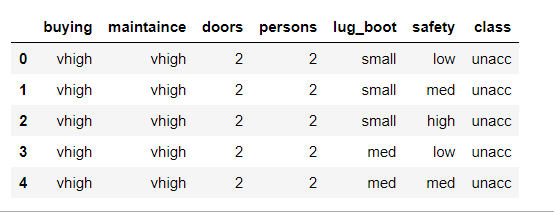
import matplotlib.pyplot as plt

import seaborn as sns

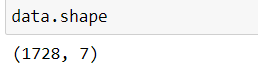
**IMPORTING DATASET:**

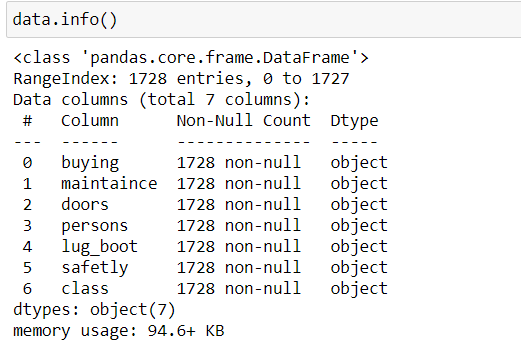
data=pd.read\_csv("D://datasets//car\_evaluation.csv")

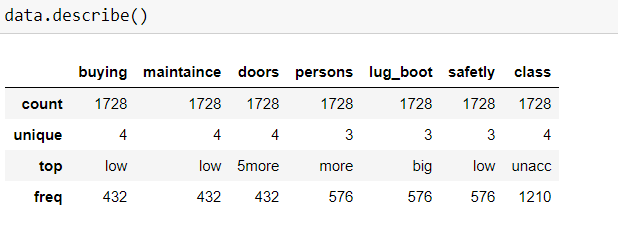
data.head(5)



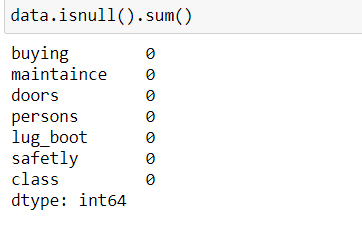
**DATA ANALYSIS:**







**DATA WRANGLING:**



data['buying'],\_ = pd.factorize(data['buying'])

data['maintaince'],\_ = pd.factorize(data['maintaince'])

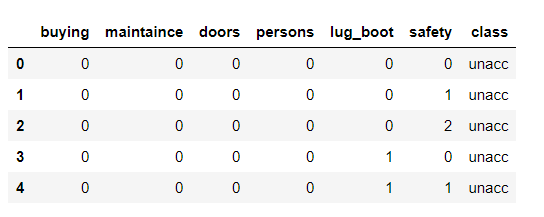
data['doors'],\_ = pd.factorize(data['doors'])

data['persons'],\_ = pd.factorize(data['persons'])

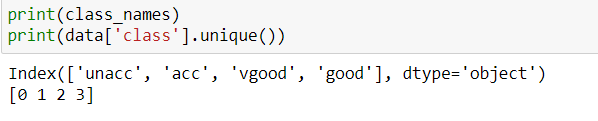
data['lug\_boot'],\_ = pd.factorize(data['lug\_boot'])

data['safety'],\_ = pd.factorize(data['safety'])

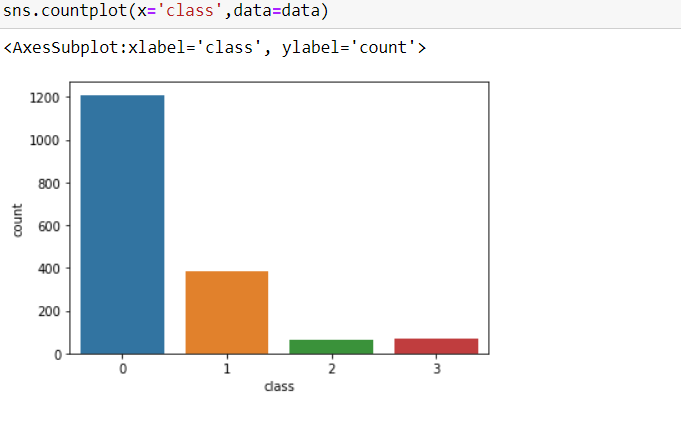
data.head()



data['class'],class\_names = pd.factorize(data['class'])

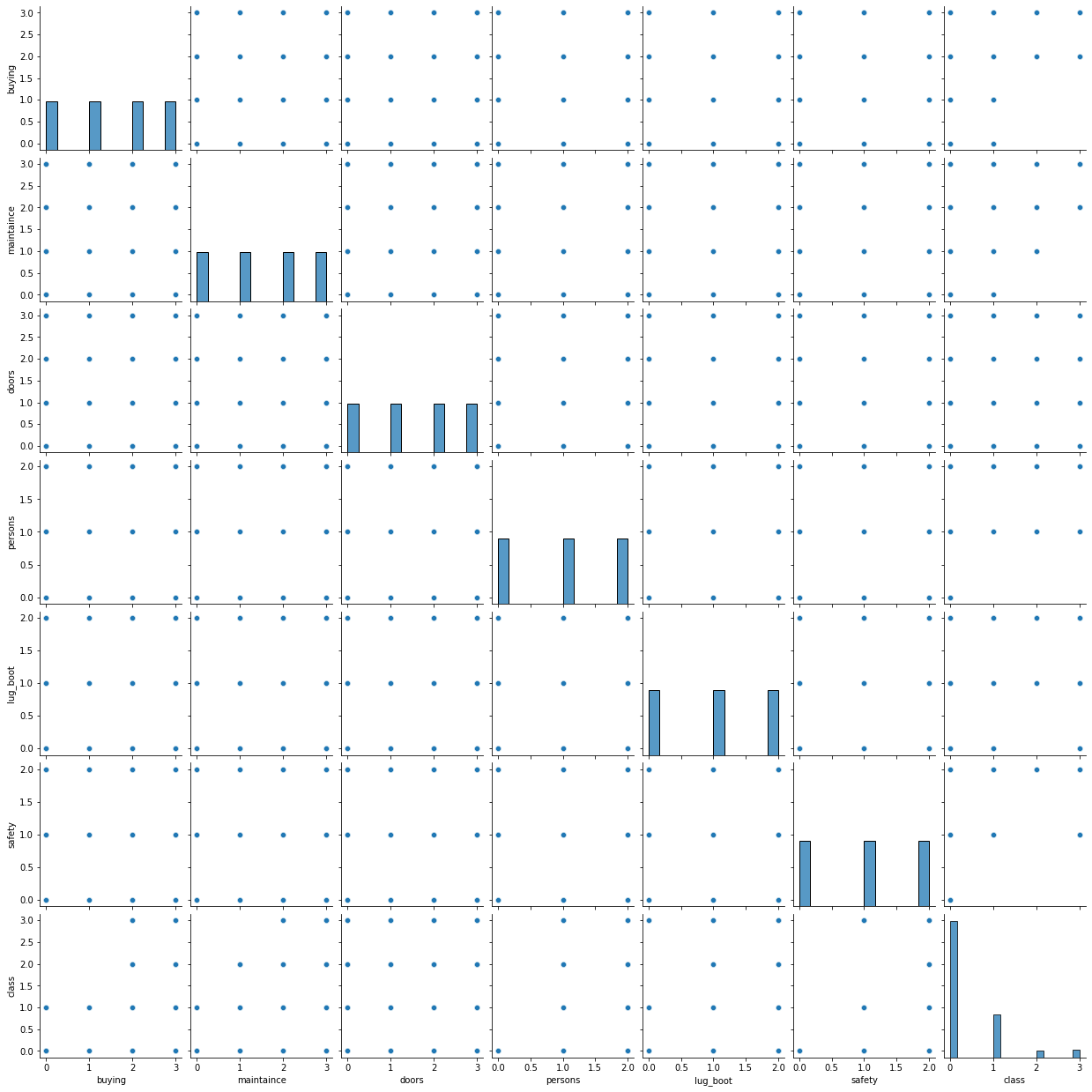


**DATA VISUALIZATION:**



From the above plot we can interpret that class 0 (unacc) has more number than any other class.

sns.pairplot(data)



**DATA PREPARATION:**



**LIBRARIES NEEDED FOR DECISION TREE CLASSIFIER**

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

from sklearn.tree import DecisionTreeClassifier

from sklearn import metrics

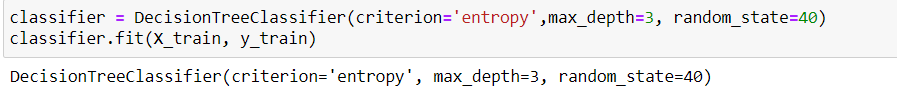
from sklearn.metrics import classification\_report, confusion\_matrix

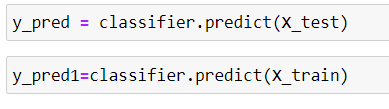
from sklearn import tree

**TEST-TRAIN SPLIT:**

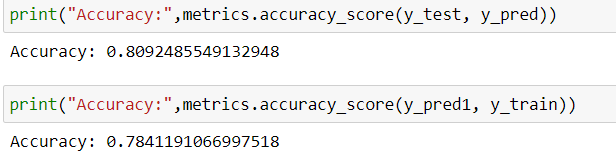


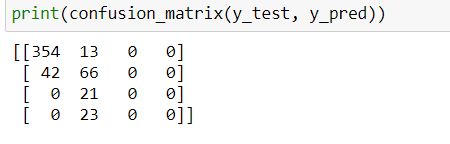
**MODEL:**

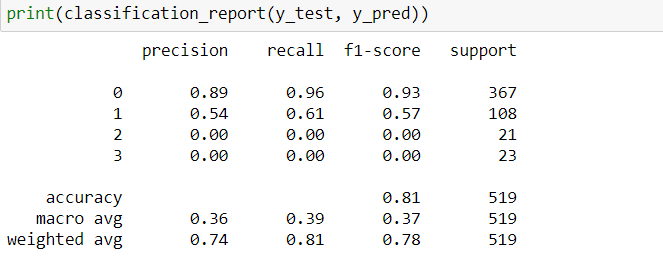




**CLASSIFICATION METRICS:**

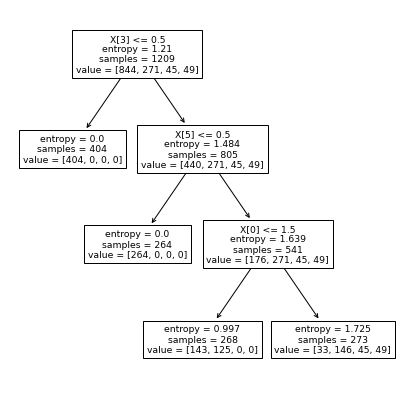




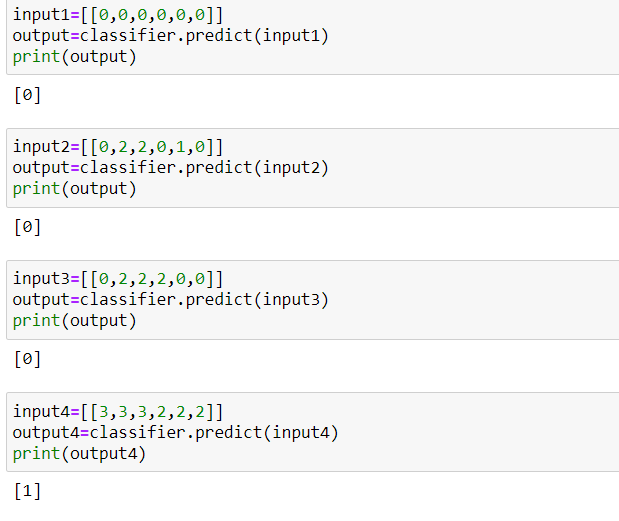


**TREE VISUALIZATION:**





**PREDICTIONS:**



**ANALYSIS:**

* Accuracy of model came out as 80.9%
* The accuracy of test data and train data looks close . so It doesn’toverfit.
* Accuracy of confusion matrix came as 80%.