Assignment #1 (5 Marks)

Submission Date: 4-2-2022

Submit a report in a single file. For the program submit a collab file.
File name enrolmentnumber_DT_Report/Program

Late submission: In every two days late submission, one mark will be deducted.

Note: You have to submit your own solution. Copy and paste from any source is prohibited, if found dealt strictly. We may use resources for understanding or may discuss with your friend, but need to write your solution and details in programs and in reports in your own language.

Data set

s.n	Price	Maintenance	Capacity	airbag	Profit (Class level)
1	low	low	2	No	YES
2	low	average	5	yes	NO
3	low	low	5	no	YES
4	low	heigh	5	no	NO
5	average	average	5	no	NO
6	average	average	5	yes	YES
7	average	heigh	2	yes	NO
8	average	heigh	7	no	YES
9	heigh	average	5	yes	YES
10	heigh	heigh	2	yes	NO
11	heigh	heigh	7	yes	YES

1. Construct a Decision tree using entropy as criterion. Write a function in python for entropy computation at each node. Using this, write a program to construct a decision tree for above data. Also, explain each step in the report. You can use only Numpy, panda and plotting libraries.

2. Use the gini index criterion and repeat all the tasks as in 1.

3. Scikit–learn based assignment.
   Use data set https://archive.ics.uci.edu/ml/datasets/car+evaluation.

   a) By using the DecisionTreeClassifier , train your model and report accuracy (F1) and confusion matrix. Plot the tree. Also, do analysis by varying parameters of DecisionTreeClassifier (at least with criterion and depth of tree). Add in a report.
   b) By using **sklearn.ensemble.RandomForestClassifier,** train your model and report accuracy (F1) and confusion matrix.Also, do analysis by varying parameters of **sklearn.ensemble.RandomForestClassifier** (at least with criterion, depth of tree and number of tree equal to 10 and 20). Add in a report.