Introduction to Machine Learning Assignment 3

Name: Vivekanand Reddy Malipatel
CWID: A20524971

Question 1 Answer:

a) (The Python code for this is in the file: 1_a.py)

Output Screenshot :

```
//////////////////////////FIRST LEVEL//////////////////////////////////

CarType :
CAR_USE      Commercial  Private    All
CAR_TYPE
Minivan             553     2141   2694
Panel Truck         853        0    853
Pickup             1068      704   1772
SUV                 555     2328   2883
Sports Car          200      979   1179
Van                 560      361    921
All                3789     6513  10302
=== Optimal Split ===
 Left Branch Set:  ['Minivan', 'Sports Car', 'SUV']
Right Branch Set:  ['Pickup', 'Panel Truck', 'Van']
   Split Entropy:  0.7684152303050842

Occupation :
CAR_USE      Commercial  Private    All
OCCUPATION
Blue Collar         1735      553   2288
Clerical             285     1305   1590
Doctor                 0      321    321
Home Maker            57      786    843
Lawyer                 0     1031   1031
Manager              308      949   1257
Professional         364     1044   1408
Student              447      452    899
Unknown              593       72    665
All                 3789     6513  10302
=== Optimal Split ===
 Left Branch Set:  ['Student', 'Unknown', 'Blue Collar']
Right Branch Set:  ['Lawyer', 'Professional', 'Home Maker', 'Clerical', 'Doctor', 'Manager']
   Split Entropy:  0.712583253573726

Education :
CAR_USE        Commercial  Private    All
EDUCATION
Bachelors            1191     1632   2823
Below High Sc         326     1189   1515
High School          1438     1514   2952
Masters               532     1546   2078
PhD                   302      632    934
All                  3789     6513  10302
=== Optimal Split ===
 Left Branch Set:  ['Below High Sc']
Right Branch Set:  ['High School', 'Bachelors', 'Masters', 'PhD']
   Split Entropy:  0.9356142508258437
//////////////////////////SECOND LEVEL//////////////////////////////////


leftBranch :

CarType :
CAR_USE      Commercial  Private   All
CAR_TYPE
Minivan             543      329   872
Panel Truck         473        0   473
Pickup              669      155   824
SUV                 542      394   936
Sports Car          193      161   354
Van                 355       38   393
All                2775     1077  3852
=== Optimal Split ===
 Left Branch Set:  ['Minivan', 'Sports Car', 'SUV']
Right Branch Set:  ['Pickup', 'Panel Truck', 'Van']
   Split Entropy:  0.7725782837913743

Occupation :
CAR_USE      Commercial  Private   All
OCCUPATION
Blue Collar         1735      553  2288
Student              447      452   899
Unknown              593       72   665
All                 2775     1077  3852
=== Optimal Split ===
 Left Branch Set:  ['Student']
Right Branch Set:  ['Unknown', 'Blue Collar']
   Split Entropy:  0.8042192219461467

Education :
CAR_USE        Commercial  Private   All
EDUCATION
Bachelors             766      122   888
Below High Sc         216      607   823
High School          1198      275  1473
Masters               371       56   427
PhD                   224       17   241
All                  2775     1077  3852
=== Optimal Split ===
 Left Branch Set:  ['Below High Sc']
Right Branch Set:  ['High School', 'Bachelors', 'Masters', 'PhD']
   Split Entropy:  0.6670194998377932
```

```
RightBranch :

CarType :
CAR_USE     Commercial  Private   All
CAR_TYPE
Minivan             10     1812  1822
Panel Truck        380        0   380
Pickup             399      549   948
SUV                 13     1934  1947
Sports Car           7      818   825
Van                205      323   528
All               1014     5436  6450
=== Optimal Split ===
 Left Branch Set:  ['Minivan', 'Sports Car', 'SUV']
Right Branch Set:  ['Pickup', 'Panel Truck', 'Van']
   Split Entropy:  0.3274450052616845

Occupation :
CAR_USE       Commercial  Private   All
OCCUPATION
Clerical             285     1305  1590
Doctor                 0      321   321
Home Maker            57      786   843
Lawyer                 0     1031  1031
Manager              308      949  1257
Professional         364     1044  1408
All                 1014     5436  6450
=== Optimal Split ===
 Left Branch Set:  ['Lawyer', 'Home Maker', 'Doctor']
Right Branch Set:  ['Professional', 'Manager', 'Clerical']
   Split Entropy:  0.5664540067183996

Education :
CAR_USE        Commercial  Private   All
EDUCATION
Bachelors             425     1510  1935
Below High Sc         110      582   692
High School           240     1239  1479
Masters               161     1490  1651
PhD                    78      615   693
All                  1014     5436  6450
=== Optimal Split ===
 Left Branch Set:  ['Below High School', 'High School', 'Bachelors']
Right Branch Set:  ['Masters', 'PhD']
   Split Entropy:  0.6187921630350406

//////////////////////Number of Observations and Probabilities////////////////////////////////

CAR_USE  Commercial  Private
Leaf
0              216      607
1             2559      470
2               30     4564
3              984      872
CAR_USE  Commercial   Private
Leaf
0         0.262454  0.737546
1         0.844833  0.155167
2         0.006530  0.993470
3         0.530172  0.469828
```

Answer :

Description of the leaf nodes of the classification tree :

LEAF	Splitting Criterion (Predictors)			Number of Observations	Predicted Probabilities		Predicted CAR_USE Category	Split Entropy Value
	Occupation	Education	CarType		Commercial	Private		
0	{'Blue Collar', 'Student', 'Unknown'}	['Below High Sc']		823	0.262454	0.737546	Private	0.6670195
1	{'Blue Collar', 'Student', 'Unknown'}	['High School', 'Bachelors', 'Masters', 'PhD']		3029	0.844833	0.155167	Commercial	
2	['Lawyer', 'Manager', 'Professional', 'Home Maker', 'Clerical', 'Doctor']		['Sports Car', 'SUV', 'Minivan']	4594	0.0065302569	0.99346974	Private	0.327445005
3	['Lawyer', 'Manager', 'Professional', 'Home Maker', 'Clerical', 'Doctor']		['Pickup', 'Van', 'Panel Truck']	1856	0.53017241	0.46982759	Commercial	

b) Answer :

As per the Above classification output, The predicted CAR_USE probability for the The person works in a *Professional* occupation, has an education level of *Doctors*, and owns a *Minivan* is **“Private”**.
Probabilities : ['Commercial' 'Private'] = [‘0.006530’ ‘0.993470’]

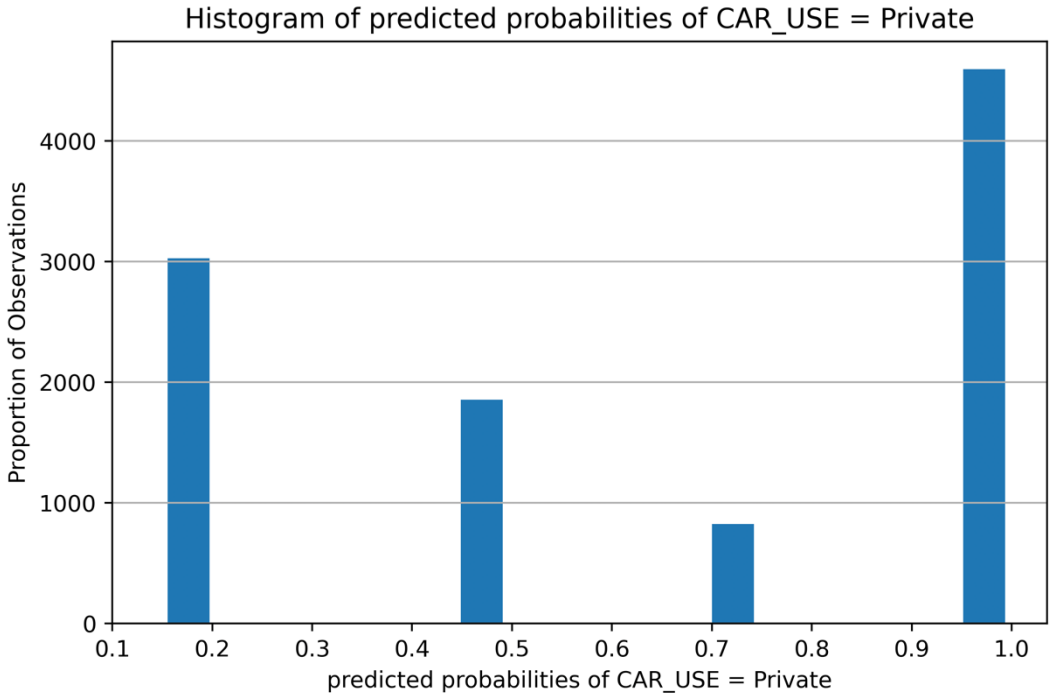
c) Answer:

As per the Above classification output, The predicted CAR_USE probability for The person is a *Student*, has a *Below High School* level of education, and owns a *Sports Car* is **"Private".**

Probabilities : ['Commercial' 'Private'] = ['0.262454' '0.737546']

d) (The Python code for this is in the file: 1_d.py)

Histogram :



e) (The Python code for this is in the file: 1_e.py)

Output Screenshot:

```
Misclassification Rate is :
15.414482624733061
```

Answer :

The misclassification rate of the Classification Tree model is 15.414482624733061.

Question 2 Answers:

a) (The Python code for this is in the file: 2_a.py)

Output Screenshot :

```
Probability of each class:
['Commercial' 'Private']
[0.36779266 0.63220734]
```

Answer:

['Commercial' 'Private']
[0.36779266 0.63220734]

b) (The Python code for this is in the file: 2_b.py)

The Below screenshot has the the frequency counts and the row probabilities in each label class.

Output Screenshot :

```
Frequency Table:
 CAR_TYPE    Minivan  Panel Truck  Pickup   SUV  Sports Car  Van
CAR_USE
Commercial      553          853    1068   555         200  560
Private        2141            0     704  2328         979  361

Row Fraction Table:
 CAR_TYPE     Minivan  Panel Truck    Pickup       SUV  Sports Car       Van
CAR_USE
Commercial  0.145949     0.225125  0.281869  0.146477    0.052784  0.147796
Private     0.328727     0.000000  0.108092  0.357439    0.150315  0.055428

Frequency Table:
 OCCUPATION  Blue Collar  Clerical  Doctor  Home Maker  Lawyer  Manager  Professional  Student  Unknown
CAR_USE
Commercial         1735       285       0          57       0      308           364      447      593
Private             553      1305     321         786    1031      949          1044      452       72

Row Fraction Table:
 OCCUPATION  Blue Collar  Clerical    Doctor  Home Maker    Lawyer   Manager  Professional   Student   Unknown
CAR_USE
Commercial     0.457904  0.075218  0.000000    0.015044  0.000000  0.081288      0.096068  0.117973  0.156506
Private        0.084907  0.200368  0.049286    0.120682  0.158299  0.145709      0.160295  0.069400  0.011055

Frequency Table:
 EDUCATION   Bachelors  Below High Sc  High School  Masters  PhD
CAR_USE
Commercial       1191            326         1438      532  302
Private          1632           1189         1514     1546  632

Row Fraction Table:
 EDUCATION   Bachelors  Below High Sc  High School   Masters       PhD
CAR_USE
Commercial   0.314331       0.086039     0.379520  0.140406  0.079704
Private      0.250576       0.182558     0.232458  0.237371  0.097037
```

c) (The Python code for this is in the file: 2_c.py)

Output Screenshot:

```
Predicted Probability:  ['Commercial' 'Private'] [[0.11280832 0.88719168]]
```

Answer:

The car usage probabilities are
['Commercial' 'Private']
[[0.11280832 0.88719168]]

d) (The Python code for this is in the file: 2_d.py)

Output Screenshot:

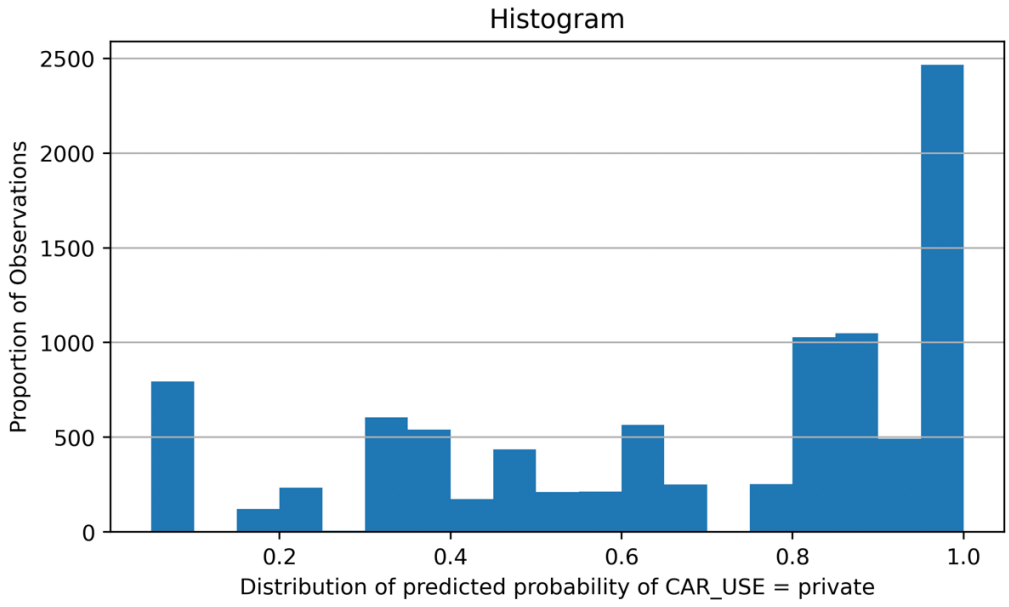
```
Predicted Probability:  ['Commercial' 'Private'] [[0.14065359 0.85934641]]
```

Answer:

The car usage probabilities are
['Commercial' 'Private']
[[0.14065359 0.85934641]]

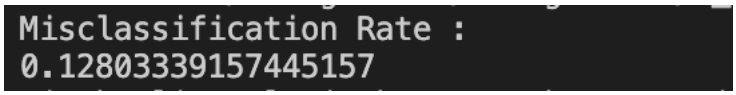
e) (The Python code for this is in the file: 2_e.py)

Histogram of the predicted probabilities of CAR_USE = *Private*:



f) (The Python code for this is in the file: 2_f.py)

Output Screenshot:



Answer :

the misclassification rate of the Naïve Bayes model is 0.12803339157445157.