Commondcode example

Prthon pythonfileAddress trainfileAddress maxDeepth testfileAddress

python C:\Users\Administrator\Desktop\DATA7703\Decision_Tree\Data7703_A1.py

C:\Users\Administrator\Desktop\DATA7703\Decision_Tree\titanic2.txt 6

C:\Users\Administrator\Desktop\DATA7703\Decision_Tree\titanic2.txt

✓ Trail_1

use two same file(tennis) and set the maxdeepth to 2

```
C:\Users\Administrator>python C:\Users\Administrator\Desktop\t\Data7703_A1.py C:\Users\Administrator\Desktop\t\tennis.txt 2 C:\Users\Administrator\Desktop\t\tennis.txt
{('sunny', 'normal'): 'yes', ('sunny', 'high'): 'no', ('rain', 'strong'): 'no', ('rain', 'weak'): 'yes', ('overcast',): 'yes'}
accuracy is 1.0

C:\Users\Administrator>
```

✓ Trail_2

use two same file(tennis) and set the maxdeepth to 1

```
C:\Users\Administrator>python C:\Users\Administrator\Desktop\t\Data7703_A1.py C:\Users\Administrator\Desktop\t\tennis.txt 1 C:\Users\Administrator\Desktop\t\tennis.txt
{('sunny',): 'no', ('rain',): 'yes', ('overcast',): 'yes'}
accuracy is 0.7142857142857143
```

✓ Trail_3

use two same file(titanic) and set the maxdeepth to 1

```
C:\Users\Administrator>python C:\Users\Administrator\Desktop\t\Data7703_A1.py C:\Users\Administrator\Desktop\t\titanic2.txt 1 C:\Users\Administrator\Desktop\t\titanic2.
txt
{('female',): 'yes', ('male',): 'no'}
accuracy is 0.7760109041344844
```

✓ Trail_4

use two same file(titanic) and set the maxdeepth to 3

```
C:\Users\Administrator>python C:\Users\Administrator\Desktop\t\Data7703_A1.py C:\Users\Administrator\Desktop\t\titanic2.txt 3 C:\Users\Administrator\Desktop\t\titanic2.
txt
{('female', '2nd', 'child'): 'yes', ('female', '2nd', 'adult'): 'yes', ('female', '1st', 'child'): 'yes', ('female', '1st', 'adult'): 'yes', ('female', '3rd', 'child'):
 'no', ('female', '3rd', 'adult'): 'no', ('female', 'crew', 'adult'): 'yes', ('male', '2nd', 'child'): 'yes', ('male', '2nd', 'adult'): 'no', ('male', '1st', 'child'):
'yes', ('male', '1st', 'adult'): 'no', ('male', '3rd', 'child'): 'no', ('male', '3rd', 'adult'): 'no', ('male', 'crew', 'adult'): 'no'}
accuracy is 0.7905497501135347
```

Basicly Code Statement:

I first package the algorithm for calculating the IG value and some other functions like that. Importantlt, during an iternation, I save the number of the column into a list L and also save the attribute of the column which is the branch of the tree. I called this function ‘’A, A1, A2 = Node_Generation(dataTrain,maxDeepth)’’ located in the 89th of the python file.

The A is an dictionary(the original form of the decesion tree, contained all the information), for instance {(‘sunny(feature)’ , 0 , ‘mild(feature)’ , 3):’yes’}, which means the feature from the 0th column is ‘’sunny’’, than the feature from the 3rd olumn is ‘’mild’’. A1 and A2 are the subset from the A, A1 is like {(‘’sunny’’, ‘’mild’’):’yes’} and A2 is like [0,3]. Blew is a small part of the branch.

When doing the calculation, for every single case I just take out all the element with the list A2, I do participation with all the items from A2 for A, and would only get one product which is one of the keys in A2. Lastly, I just make prediction.