**4) Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.**

**import** pandas **as** pd

**from** pprint **import** pprint

**from** sklearn.feature\_selection **import** mutual\_info\_classif

**from** collections **import** Counter

**def** id3(df, target\_attribute, attribute\_names, default\_class**=None**):

cnt**=**Counter(x **for** x **in** df[target\_attribute])

**if** len(cnt)**==**1:

**return** next(iter(cnt))

**elif** df**.**empty **or** (**not** attribute\_names):

**return** default\_class

**else**:

gainz **=** mutual\_info\_classif(df[attribute\_names],df[target\_attribute],discrete\_features**=True**)

index\_of\_max**=**gainz**.**tolist()**.**index(max(gainz))

best\_attr**=**attribute\_names[index\_of\_max]

tree**=**{best\_attr:{}}

remaining\_attribute\_names**=**[i **for** i **in** attribute\_names **if** i**!=**best\_attr]

**for** attr\_val, data\_subset **in** df**.**groupby(best\_attr):

subtree**=**id3(data\_subset, target\_attribute, remaining\_attribute\_names,default\_class)

tree[best\_attr][attr\_val]**=**subtree

**return** tree

df**=**pd**.**read\_csv("p-tennis.csv")

attribute\_names**=**df**.**columns**.**tolist()

print("List of attribut name")

attribute\_names**.**remove("PlayTennis")

**for** colname **in** df**.**select\_dtypes("object"):

df[colname], \_ **=** df[colname]**.**factorize()

print(df)

tree**=** id3(df,"PlayTennis", attribute\_names)

print("The tree structure")

pprint(tree)

**Output:**

List of attribut name

Outlook Temperature Humidity Windy PlayTennis

0 0 0 0 False 0

1 0 0 0 True 0

2 1 0 0 False 1

3 2 1 0 False 1

4 2 2 1 False 1

5 2 2 1 True 0

6 1 2 1 True 1

7 0 1 0 False 0

8 0 2 1 False 1

9 2 1 1 False 1

10 0 1 1 True 1

11 1 1 0 True 1

12 1 0 1 False 1

13 2 1 0 True 0

The tree structure

{'Outlook': {0: {'Humidity': {0: 0, 1: 1}},

1: 1,

2: {'Windy': {False: 1, True: 0}}}}