A02

April 29, 2022

```python
[9]: # import warnings filter
     from warnings import simplefilter
     # ignore all future warnings
     simplefilter(action='ignore', category=FutureWarning)

     import pandas as pd
     import numpy as np

     df=pd.read_csv('DTC.csv')
     df
     df_up = df
```

```python
[10]: x=df.iloc[:,1:-1]
```

```python
[11]: x
```

```
[11]:        Age  Income  Gender Marital Status
     0      <21    High    Male          Single
     1      <21    High    Male         Married
     2    21-35    High    Male          Single
     3      >35  Medium    Male          Single
     4      >35     Low  Female          Single
     5      >35     Low  Female         Married
     6    21-35     Low  Female         Married
     7      <21  Medium    Male          Single
     8      <21     Low  Female         Married
     9      >35  Medium  Female          Single
     10     <21  Medium  Female         Married
     11   21-35  Medium    Male         Married
     12   21-35    High  Female          Single
     13     >35  Medium    Male         Married
```

```python
[12]: x['Gender'].values.reshape(-1,1)
```

```
[12]: array([['Male'],
            ['Male'],
            ['Male'],
            ['Male'],
```

```
       ['Female'],
       ['Female'],
       ['Female'],
       ['Male'],
       ['Female'],
       ['Female'],
       ['Female'],
       ['Male'],
       ['Female'],
       ['Male']], dtype=object)
```

```python
[13]: from sklearn.preprocessing import OneHotEncoder
      ohe1 = OneHotEncoder()
      ohe2 = OneHotEncoder()
      ohe3 = OneHotEncoder()
      ohe4 = OneHotEncoder()

      X_age = ohe1.fit_transform(x['Age'].values.reshape(-1,1)).toarray().astype(int)
      X_income = ohe2.fit_transform(x['Income'].values.reshape(-1,1)).toarray().
       ↪astype(int)
      X_gen = ohe3.fit_transform(x['Gender'].values.reshape(-1,1)).toarray().
       ↪astype(int)
      X_ms = ohe4.fit_transform(x['Marital Status'].values.reshape(-1,1)).toarray().
       ↪astype(int)

      # print(ohe.categories_)
      dfage = pd.DataFrame(X_age,columns=ohe1.categories_)
      df = pd.concat([x,dfage],axis=1)
      df = df.drop(['Age'],axis=1)

      dfincome = pd.DataFrame(X_income,columns=ohe2.categories_)
      df = pd.concat([df, dfincome],axis=1)
      df = df.drop(['Income'],axis=1)

      dfOneHot = pd.DataFrame(X_gen, columns=ohe3.categories_)
      df = pd.concat([df, dfOneHot], axis=1)
      df = df.drop(['Gender'], axis=1)

      dfoh = pd.DataFrame(X_ms,columns=ohe4.categories_)
      df = pd.concat([df, dfoh], axis=1)
      df = df.drop(['Marital Status'], axis=1)
      df
```

```
[13]:    (21-35,)  (<21,)  (>35,)  (High,)  (Low,)  (Medium,)  (Female,)  (Male,)  \
      0         0       1       0        1       0          0          0        1
      1         0       1       0        1       0          0          0        1
      2         1       0       0        1       0          0          0        1
```

3	0	0	1	0	0	1	0	1
4	0	0	1	0	1	0	1	0
5	0	0	1	0	1	0	1	0
6	1	0	0	0	1	0	1	0
7	0	1	0	0	0	1	0	1
8	0	1	0	0	1	0	1	0
9	0	0	1	0	0	1	1	0
10	0	1	0	0	0	1	1	0
11	1	0	0	0	0	1	0	1
12	1	0	0	1	0	0	1	0
13	0	0	1	0	0	1	0	1

	(Married,)	(Single,)
0	0	1
1	1	0
2	0	1
3	0	1
4	0	1
5	1	0
6	1	0
7	0	1
8	1	0
9	0	1
10	1	0
11	1	0
12	0	1
13	1	0

```
[14]: target = df_up.iloc[:,5]
```

```
[15]: target
```

```
[15]: 0      No
      1      No
      2     Yes
      3     Yes
      4     Yes
      5      No
      6     Yes
      7      No
      8     Yes
      9     Yes
      10    Yes
      11    Yes
      12    Yes
      13     No
      Name: Buys, dtype: object
```

```
[16]: from sklearn.tree import DecisionTreeClassifier
      dtc=DecisionTreeClassifier(criterion="entropy")
      dtc.fit(df,target)
      y_pred = dtc.predict(df)
```

```
[17]: y_pred
```

```
[17]: array(['No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes',
             'Yes', 'Yes', 'Yes', 'No'], dtype=object)
```

```
[18]: # [Age < 21,Income = Low,Gender = Female, Marital Status = Married]
      text = [["<21","Low","Female","Married"]]
      text = pd.DataFrame(text,columns = df_up.columns[1:-1])
      print(text)

      a = ohe1.transform(text["Age"].values.reshape(-1,1)).toarray().astype(int)
      b = ohe2.transform(text["Income"].values.reshape(-1,1)).toarray().astype(int)
      c = ohe3.transform(text["Gender"].values.reshape(-1,1)).toarray().astype(int)
      d = ohe4.transform(text["Marital Status"].values.reshape(-1,1)).toarray().
       →astype(int)

      dfage = pd.DataFrame(a,columns=ohe1.categories_)
      df = pd.concat([text,dfage],axis=1)
      df = df.drop(['Age'],axis=1)

      dfincome = pd.DataFrame(b,columns=ohe2.categories_)
      df = pd.concat([df, dfincome],axis=1)
      df = df.drop(['Income'],axis=1)

      dfOneHot = pd.DataFrame(c, columns=ohe3.categories_)
      df = pd.concat([df, dfOneHot], axis=1)
      df = df.drop(['Gender'], axis=1)

      dfoh = pd.DataFrame(d,columns=ohe4.categories_)
      df = pd.concat([df, dfoh], axis=1)
      df = df.drop(['Marital Status'], axis=1)
      df
      # test_x=np.array([0,1,0,0,1,0,1,0,1,0])
```

```
        Age Income  Gender Marital Status
     0  <21    Low  Female        Married
```

```
[18]:    (21-35,)  (<21,)  (>35,)  (High,)  (Low,)  (Medium,)  (Female,)  (Male,)  \
     0          0       1       0        0       1          0          1        0

        (Married,)  (Single,)
     0           1          0
```

```
[19]: pred_y=dtc.predict(df)
      pred_y
```

```
[19]: array(['Yes'], dtype=object)
```
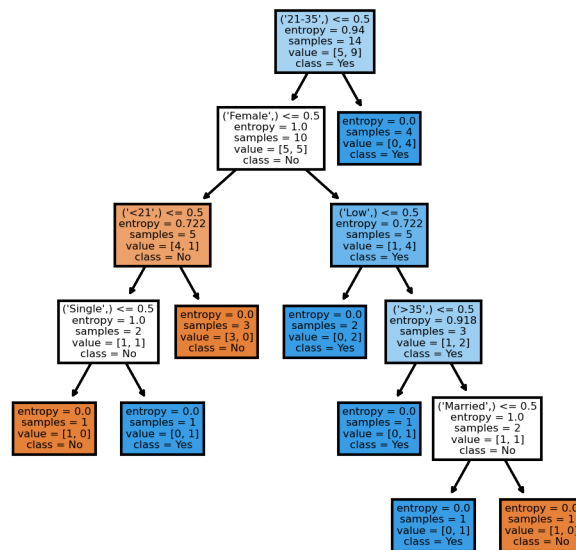
```
[20]: from sklearn.metrics import confusion_matrix

      confusion_matrix(target, y_pred)
```

```
[20]: array([[5, 0],
             [0, 9]])
```

```
[16]: from sklearn import tree
      import matplotlib.pyplot as plt

      fig, axes = plt.subplots(figsize = (4,4), dpi=300)
      tree.plot_tree(dtc, feature_names = df.columns, class_names=['No','Yes'],␣
       ↪filled = True)
      plt.show()
```



```
[ ]:
```