Assignment no 4

Name: YEOLE BHAUSAHEB

Roll No.: 176

Title: Create a machine learning model using Decision Tree

Solution:

```
import pandas as pd
import numpy as np

dataset = pd.read_csv("bill_authentication.csv")
#print dataset.shape
#print dataset.head() # prints first five enteries of the
dataset

X = dataset.iloc[:,:4]
Y = dataset.iloc[:,4]

#print X.head()
#print Y

from sklearn.cross_validation import train_test_split
X_train,X_test,Y_train,Y_test =
train_test_split(X,Y,test_size=0.2) # X and Y will be devided
into 2-2 parts each
                                                # i.e. X train Xtest
and Y train Y test

from sklearn.tree import DecisionTreeClassifier

clf= DecisionTreeClassifier()

clf.fit(X_train,Y_train)   # train the algo with predicter
variable and response variable

a = clf.predict(X_test)    # test the response
```

```
#print a.shape
#print a

from sklearn.metrics import classification_report,
confusion_matrix
print confusion_matrix(Y_test,a)



new = np.array([[0.74521,3.9357,-4.4044,-4.1414]])
npred = clf.predict(new)
print npred



from sklearn.externals.six import StringIO
from IPython.display import Image
from sklearn.tree import export_graphviz
#import pydotplus
dot_data = StringIO()

export_graphviz(clf,out_file=dot_data,filled=True,rounded=
True,special_characters=True) # describe the
characterstics of the graph
                                # round boxes, filled colors

#graph=
pydotplus.graph_from_dot_data(dot_data.getvalue()) #
creates the graph
#graph.write_png('tree.png')    # saves the graph in current
working directory
```

Output:

	Variance	Skewness	Curtosis	Entropy	Class
0	3.62160	8.6661	-2.8073	-0.44699	0
1	4.54590	8.1674	-2.4586	-1.4621	0
2	3.86600	-2.6383	1.9242	0.10645	0
3	3.45660	9.5228	-4.0112	-3.59440	0
4	0.32924	-4.4552	4.5718	-0.98880	0

```
[[142  1]
 [  0 132]]
```

	precision	recall	f1-score	support
0	0.99	1.00	1.00	142
1	1.00	0.99	1.00	133
avg / total	1.00	1.00	1.00	275