15BCE0329

Harshit Kedia

EXP 4

CSE 4020

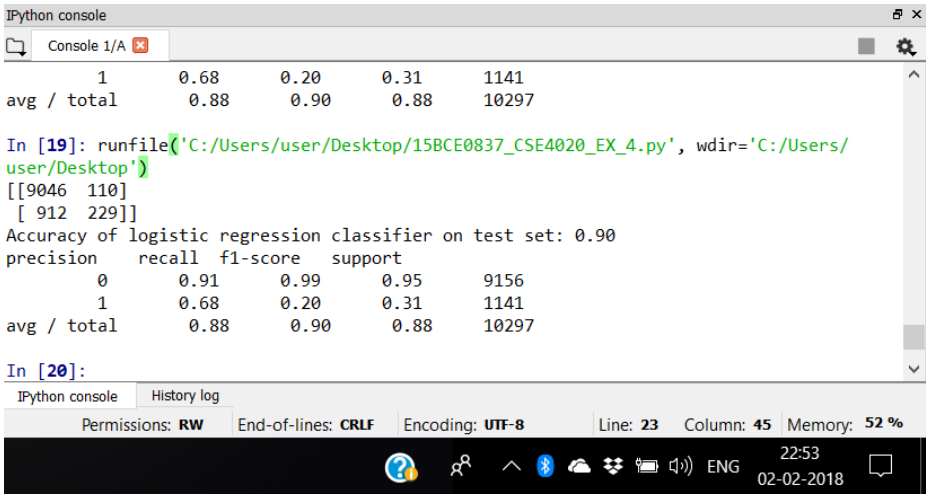
Dataset information –

Dataset link - https://archive.ics.uci.edu/ml/datasets/bank+marketing

CODE

```
import pandas as pd

import numpy as np

from sklearn import preprocessing

import matplotlib.pyplot as plt

from sklearn.linear_model import LogisticRegression

from sklearn.cross_validation import train_test_split

data = pd.read_csv('bank.csv', header=0)

data = data.dropna()

data.drop(data.columns[[0, 3, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19]], axis=1, inplace=True)

data2 = pd.get_dummies(data, columns =['job', 'marital', 'default', 'housing', 'loan', 'poutcome'])

data2.drop(data2.columns[[12, 16, 18, 21, 24]], axis=1, inplace=True)

X = data2.iloc[:,1:]

y = data2.iloc[:,0]

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)

classifier = LogisticRegression(random_state=0)

classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

from sklearn.metrics import confusion_matrix

confusion_matrix = confusion_matrix(y_test, y_pred)

print(confusion_matrix)

print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(classifier.score(X_test, y_test)))

from sklearn.metrics import classification_report

print(classification_report(y_test, y_pred))
```

OUTPUT


IPython console
Console 1/A
1 0.68 0.20 0.31 1141
avg / total 0.88 0.90 0.88 10297
In [19]: runfile('C:/Users/user/Desktop/15BCE0837_CSE4020_EX_4.py', wdir='C:/Users/user/Desktop')
[[9046 110]
[912 229]]
Accuracy of logistic regression classifier on test set: 0.90
precision recall f1-score support
0 0.91 0.99 0.95 9156
1 0.68 0.20 0.31 1141
avg / total 0.88 0.90 0.88 10297
In [20]:
IPython console
History log
Permissions: RW
End-of-lines: CRLF
Encoding: UTF-8
Line: 23
Column: 45
Memory: 52 %
ENG
22:53
02-02-2018