Task 3.1 SIT719 Sanket Thakur Scikit learn for ML

December 6, 2020

[6]:
```
#Import pandas and sklearn libraries
import pandas as pd
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier

#Importing data. the data we are using is the data of payment frauds
data = pd.read_csv('payment_fraud.csv')

#Checking the shape of the data
data.shape

#checking if there is any null values in the dataset
data[data.isnull().any(axis=1)].head()
#there is no null value in the data set

#changing catergorical variables to numeric
data_only_numbers = data.replace({'paymentMethod':{'paypal':'0', 'storecredit':
 ↪'1','creditcard':'2'}})

#creating a data set with no lables
data_no_labels = data_only_numbers.drop('label', axis = 1)

#creating a data set with only labels
data_labels_only=data_only_numbers['label']

#importing sklearn.model_selection
from sklearn.model_selection import train_test_split

#spliting  data for training and testing x is the dependent variable and y is␣
 ↪the independent variable
data_no_labels = data_only_numbers.drop('label', axis = 1)
data_no_labels
data_labels_only=data_only_numbers['label']
data_labels_only
```

```
x_train, x_test, y_train, y_test = train_test_split(data_no_labels,␣
 ↪data_labels_only, test_size=0.33, random_state=0)

#fitting a decision tree model on the data

#Creating a Decision Tree classifer object
clf = DecisionTreeClassifier()

#Train Decision Tree Classifer
clf = clf.fit(x_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(x_test)

print(accuracy_score(y_pred, y_test))
print(confusion_matrix(y_test, y_pred))
```

```
1.0
[[12771     0]
 [    0   172]]
```

[ ]: