Untitled35 (1)

January 10, 2020

[1]:
```
import pandas as pd
```

[12]:
```
data_set=pd.read_csv('C:\\Users\\IBM\\Desktop\\2010-capitalbikeshare-tripdata.
 ↪csv')
print(data_set.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 115597 entries, 0 to 115596
Data columns (total 9 columns):
Duration                115597 non-null int64
Start date              115597 non-null object
End date                115597 non-null object
Start station number    115597 non-null int64
Start station           115597 non-null object
End station number      115597 non-null int64
End station             115597 non-null object
Bike number             115597 non-null object
Member type             115597 non-null object
dtypes: int64(3), object(6)
memory usage: 7.9+ MB
None
```

[13]:
```
X = data_set.iloc[:, [0, 3, 5]].values
y = data_set.iloc[:, -1].values
print(X[:5])
print(y[:5])
```

```
[[ 1012 31208 31108]
 [   61 31209 31209]
 [ 2690 31600 31100]
 [ 1406 31600 31602]
 [ 1413 31100 31201]]
['Member' 'Member' 'Member' 'Member' 'Member']
```

[14]:
```
from sklearn.preprocessing import LabelEncoder
labelencoder_y =LabelEncoder()
y = labelencoder_y.fit_transform(y)
```

```
print("Sample y:",y[:5])
print("0 :",labelencoder_y.classes_[0])
print("1 :",labelencoder_y.classes_[1])
```

```
Sample y: [1 1 1 1 1]
0 : Casual
1 : Member
```

[15]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,
random_state=0)
```

[16]:
```
from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion='entropy',
min_samples_leaf=4, random_state=0)
classifier.fit(X_train, y_train)
```

[16]:
```
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=None,
                       max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=4, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort=False,
                       random_state=0, splitter='best')
```

[17]:
```
y_pred = classifier.predict(X_test)
```

[18]:
```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
print(cm)
```

```
[[ 3368  2721]
 [ 2641 20170]]
```

[ ]:
```

```