Day 25 - Naive Bayes

In [1]:

import pandas as pd  
import numpy as np  
  
dataset = pd.read\_csv("Dataset/train.csv")  
dataset = dataset.drop(["Name","Ticket","Cabin","PassengerId"],axis=1)  
  
dataset.head()

Out[1]:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Survived | Pclass | Sex | Age | SibSp | Parch | Fare | Embarked |
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S |

In [2]:

from sklearn import preprocessing  
  
#from sklearn.cross\_validaNtion import train\_test\_split  
from sklearn.model\_selection import train\_test\_split  
  
#from sklearn.naive\_bayes import GaussianNB  
from sklearn.naive\_bayes import \*  
  
from sklearn.metrics import accuracy\_score  
from sklearn.metrics import confusion\_matrix

In [3]:

le=preprocessing.LabelEncoder()  
  
le.fit(dataset["Sex"])  
print(le.classes\_)  
dataset["Sex"] = le.transform(dataset["Sex"])  
  
le.fit(dataset["Embarked"])  
print(le.classes\_)  
dataset["Embarked"] = le.transform(dataset["Embarked"])  
  
dataset['Age'] = np.round(dataset['Age'])  
dataset['Fare'] = np.round(dataset['Fare'])  
  
  
#changes columns with float type to int  
float\_col = dataset.select\_dtypes(include=['float64'])  
for col in float\_col.columns.values:  
 dataset[col] = dataset[col].astype('int64')  
  
  
# Changing continious to categorical using quartiles  
dataset['Age'] = np.where((dataset.Age <= 22),0,dataset.Age)  
dataset['Age'] = np.where(((dataset.Age > 22) & (dataset.Age <= 32)),1,dataset.Age)  
dataset['Age'] = np.where(((dataset.Age > 32) & (dataset.Age <= 45)),2,dataset.Age)  
dataset['Age'] = np.where((dataset.Age > 45),3,dataset.Age)  
  
#dataset['Fare'] = np.where((dataset.Fare <= 8),0,dataset.Fare)  
#dataset['Fare'] = np.where(((dataset.Fare > 8) & (dataset.Fare <= 14)),1,dataset.Fare)  
#dataset['Fare'] = np.where(((dataset.Fare > 14) & (dataset.Fare <= 31)),2,dataset.Fare)  
#dataset['Fare'] = np.where((dataset.Fare > 31),3,dataset.Fare)  
  
dataset.head(10)

['female' 'male']  
['C' 'Q' 'S']

# ==============================================================[¶](#X862e383c15c8f13a5fcea1f79fcdc84d5327c27)

In [4]:

def accuracy\_fun(feature):  
 print("======================================================================================")  
 print(" === ",feature," ===\n")  
 y=dataset[feature]  
 X=dataset.drop([feature],axis=1)  
   
 print(X.head())  
 print()  
 print("Total count of (",feature,") - dependent data : ", y.count())  
   
 X\_train,X\_test,y\_train,y\_test = train\_test\_split(X,y,test\_size=0.3,random\_state=0)  
   
 clf=BernoulliNB()  
 y\_pred=clf.fit(X\_train,y\_train).predict(X\_test)  
   
 print("\n Accuracy score : ",accuracy\_score(y\_test,y\_pred,normalize=True))  
 print("\n Confusion Matrix")  
 print(confusion\_matrix(y\_test,y\_pred))  
 print("\n======================================================================================\n\n")  
   
  
   
features = ['Survived','Pclass','Sex','Age','SibSp','Parch','Fare','Embarked']  
  
for feature in features:  
 accuracy\_fun(feature)

======================================================================================  
 === Survived ===  
  
 Pclass Sex Age SibSp Parch Fare Embarked  
0 3 1 0 1 0 7 2  
1 1 0 2 1 0 71 0  
2 3 0 1 0 0 8 2  
3 1 0 2 1 0 53 2  
4 3 1 2 0 0 8 2  
  
Total count of ( Survived ) - dependent data : 889  
  
 Accuracy score : 0.7640449438202247  
  
 Confusion Matrix  
[[131 26]  
 [ 37 73]]  
  
======================================================================================  
  
  
======================================================================================  
 === Pclass ===  
  
 Survived Sex Age SibSp Parch Fare Embarked  
0 0 1 0 1 0 7 2  
1 1 0 2 1 0 71 0  
2 1 0 1 0 0 8 2  
3 1 0 2 1 0 53 2  
4 0 1 2 0 0 8 2  
  
Total count of ( Pclass ) - dependent data : 889  
  
 Accuracy score : 0.5917602996254682  
  
 Confusion Matrix  
[[ 27 6 37]  
 [ 10 8 31]  
 [ 20 5 123]]  
  
======================================================================================  
  
  
======================================================================================  
 === Sex ===  
  
 Survived Pclass Age SibSp Parch Fare Embarked  
0 0 3 0 1 0 7 2  
1 1 1 2 1 0 71 0  
2 1 3 1 0 0 8 2  
3 1 1 2 1 0 53 2  
4 0 3 2 0 0 8 2  
  
Total count of ( Sex ) - dependent data : 889  
  
 Accuracy score : 0.7378277153558053  
  
 Confusion Matrix  
[[ 56 42]  
 [ 28 141]]  
  
======================================================================================  
  
  
======================================================================================  
 === Age ===  
  
 Survived Pclass Sex SibSp Parch Fare Embarked  
0 0 3 1 1 0 7 2  
1 1 1 0 1 0 71 0  
2 1 3 0 0 0 8 2  
3 1 1 0 1 0 53 2  
4 0 3 1 0 0 8 2  
  
Total count of ( Age ) - dependent data : 889  
  
 Accuracy score : 0.36704119850187267  
  
 Confusion Matrix  
[[30 3 35 0]  
 [10 3 58 0]  
 [30 7 65 0]  
 [ 8 4 14 0]]  
  
======================================================================================  
  
  
======================================================================================  
 === SibSp ===  
  
 Survived Pclass Sex Age Parch Fare Embarked  
0 0 3 1 0 0 7 2  
1 1 1 0 2 0 71 0  
2 1 3 0 1 0 8 2  
3 1 1 0 2 0 53 2  
4 0 3 1 2 0 8 2  
  
Total count of ( SibSp ) - dependent data : 889  
  
 Accuracy score : 0.700374531835206  
  
 Confusion Matrix  
[[160 20 0 0 2 0 0]  
 [ 38 25 0 0 2 0 0]  
 [ 6 2 0 0 0 0 0]  
 [ 2 1 0 0 4 0 0]  
 [ 0 0 0 0 2 0 0]  
 [ 0 0 0 0 1 0 0]  
 [ 2 0 0 0 0 0 0]]  
  
======================================================================================  
  
  
======================================================================================  
 === Parch ===  
  
 Survived Pclass Sex Age SibSp Fare Embarked  
0 0 3 1 0 1 7 2  
1 1 1 0 2 1 71 0  
2 1 3 0 1 0 8 2  
3 1 1 0 2 1 53 2  
4 0 3 1 2 0 8 2  
  
Total count of ( Parch ) - dependent data : 889  
  
 Accuracy score : 0.7415730337078652  
  
 Confusion Matrix  
[[189 9 0 0 0]  
 [ 31 9 0 0 0]  
 [ 22 5 0 0 0]  
 [ 1 0 0 0 0]  
 [ 1 0 0 0 0]]  
  
======================================================================================  
  
  
======================================================================================  
 === Fare ===  
  
 Survived Pclass Sex Age SibSp Parch Embarked  
0 0 3 1 0 1 0 2  
1 1 1 0 2 1 0 0  
2 1 3 0 1 0 0 2  
3 1 1 0 2 1 0 2  
4 0 3 1 2 0 0 2  
  
Total count of ( Fare ) - dependent data : 889  
  
 Accuracy score : 0.2958801498127341  
  
 Confusion Matrix  
[[0 0 0 ... 0 0 0]  
 [0 0 0 ... 0 0 0]  
 [0 0 5 ... 0 0 0]  
 ...  
 [0 0 0 ... 0 0 0]  
 [0 0 0 ... 0 0 0]  
 [0 0 0 ... 0 0 0]]  
  
======================================================================================  
  
  
======================================================================================  
 === Embarked ===  
  
 Survived Pclass Sex Age SibSp Parch Fare  
0 0 3 1 0 1 0 7  
1 1 1 0 2 1 0 71  
2 1 3 0 1 0 0 8  
3 1 1 0 2 1 0 53  
4 0 3 1 2 0 0 8  
  
Total count of ( Embarked ) - dependent data : 889  
  
 Accuracy score : 0.7340823970037453  
  
 Confusion Matrix  
[[ 0 0 49]  
 [ 0 0 22]  
 [ 0 0 196]]  
  
======================================================================================

In [ ]: