Prediction using a Model

In [3]:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score,cohen_kappa_score

#csv file
url = 'C:/Users/Prerna/Desktop/ML_jupyter_notenooks/datasets/titanic.csv'

#Creating a dataframe
dataframe = pd.read_csv(url).fillna(0)

#DATA CLEANING
#Dropping columns
dataframe = dataframe.drop('Name',axis=1)
dataframe = dataframe.drop('SexCode',axis=1)

# Create mapper
pclass_mapper = {"1st":1,"2nd":2,"3rd":3}
gender_mapper = {"male":1 ,"female":2}

# Replace feature values with scale
dataframe["PClass"] = dataframe["PClass"].replace(pclass_mapper)
dataframe["Sex"] = dataframe["Sex"].replace(gender_mapper)

#Replacing missing values of Age with mean of age
dataframe["Age"] = np.where(dataframe['Age']==0,np.mean(dataframe['Age']),dataframe['Age'])


# Input features
x = dataframe.iloc[:, :3].values
# Output class
y = dataframe.iloc[:, 3].values


'''
xtrain = training features
xtest = testing features
ytrain = classes of training data
ytest = classes of testing data
'''
```

```python
#Splitting the data into training and test
xtrain, xtest, ytrain, ytest = train_test_split(x, y, test_size = 0.3, random_state = 0)

#Create a model
gnb = GaussianNB()

#Fit a model
gnb.fit(xtrain, ytrain)

#Perform Predictions
y_pred = gnb.predict(xtest)

#Confusion Matrix
cm = confusion_matrix(ytest, y_pred)

print ("Confusion Matrix : \n", cm)
error_rate = 1 - accuracy_score(ytest, y_pred)
print ("Classification Report:\n")
print(classification_report(ytest,y_pred))
print("Accuracy: \n", accuracy_score(ytest, y_pred))
print("Error Rate: \n",error_rate )
print("Kappa Score: \n", cohen_kappa_score(ytest, y_pred))

```

```
Confusion Matrix :
 [[188  76]
 [ 33  97]]
Classification Report:

              precision    recall  f1-score   support

           0       0.85      0.71      0.78       264
           1       0.56      0.75      0.64       130

    accuracy                           0.72       394
   macro avg       0.71      0.73      0.71       394
weighted avg       0.75      0.72      0.73       394

Accuracy:
 0.7233502538071066
Error Rate:
 0.2766497461928934
```

```
Kappa Score:
 0.422784333754469
```