In [2]:
```
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import seaborn as sns
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from sklearn.model_selection import cross_val_score
%matplotlib inline
```

In [3]:
```
df= pd.read_csv("spam.csv")
df.head()
```

Out[3]:

	category	Meassage
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

In [4]:
```
df.groupby("category").describe()
```

Out[4]:

	Meassage			
	count	unique	top	freq
category				
ham	4825	4516	Sorry, I'll call later	30
spam	747	653	Please call our customer service representativ...	4

In [5]:
```
df["spam"]=df["category"].apply(lambda x:1 if x=="spam"else 0)
df.head()
```

Out[5]:

	category	Meassage	spam
0	ham	Go until jurong point, crazy.. Available only ...	0
1	ham	Ok lar... Joking wif u oni...	0
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	1
3	ham	U dun say so early hor... U c already then say...	0
4	ham	Nah I don't think he goes to usf, he lives aro...	0

In [6]:
```
inputs = df.drop('spam',axis='columns')
```

In [7]:
```
target = df['spam']
```

In [8]:
```
from sklearn.preprocessing import LabelEncoder
le_Category = LabelEncoder()
le_Message = LabelEncoder()
```

In [11]:
```
inputs['category_n'] = le_Category.fit_transform(inputs['category'])
inputs['Message_n'] = le_Message.fit_transform(inputs['Meassage'])
```

In [12]:
```
print(inputs)
```

```
     category                                           Meassage  category_n  \
0         ham  Go until jurong point, crazy.. Available only ...           0
1         ham                      Ok lar... Joking wif u oni...           0
2        spam  Free entry in 2 a wkly comp to win FA Cup fina...           1
3         ham  U dun say so early hor... U c already then say...           0
4         ham  Nah I don't think he goes to usf, he lives aro...           0
...       ...                                                ...         ...
5567     spam  This is the 2nd time we have tried 2 contact u...           1
5568      ham              Will ü b going to esplanade fr home?           0
5569      ham  Pity, * was in mood for that. So...any other s...           0
5570      ham  The guy did some bitching but I acted like i'd...           0
5571      ham                        Rofl. Its true to its name           0

      Message_n
0          1094
1          3141
2          1012
3          4137
4          2796
...         ...
5567       4041
5568       4613
5569       3328
5570       3948
5571       3452

[5572 rows x 4 columns]
```

In [14]:
```
inputs_n = inputs.drop(['category', 'Meassage'],axis='columns')
```

In [15]:
```
print(inputs_n)
```

```
      category_n  Message_n
0              0       1094
1              0       3141
2              1       1012
3              0       4137
4              0       2796
...          ...        ...
5567           1       4041
5568           0       4613
5569           0       3328
5570           0       3948
5571           0       3452

[5572 rows x 2 columns]
```

In [16]:
```
from sklearn import tree
model = tree.DecisionTreeClassifier()
```

In [17]:
```
model.fit(inputs_n,target)
```

Out[17]:
```
DecisionTreeClassifier()
```

In [18]:
```
model.score(inputs_n,target)
```

Out[18]:
```
1.0
```

In [19]:
```
model.predict([[0,4613]])
```

```
C:\Users\Windows 10\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names,
but DecisionTreeClassifier was fitted with feature names
  warnings.warn(
```

Out[19]:
```
array([0], dtype=int64)
```

In [20]:
```
model.predict([[1,5567]])
```

```
C:\Users\Windows 10\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarning: X does not have valid feature names,
but DecisionTreeClassifier was fitted with feature names
  warnings.warn(
```

Out[20]:
```
array([1], dtype=int64)
```

In [21]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(inputs_n, target, test_size=0.25,random_state=42)
```

In [22]:
```
pred = model.predict(X_test)
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test, pred)
```

Out[22]:
```
array([[1207,    0],
       [   0,  186]], dtype=int64)
```

In [23]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test, pred)
```

Out[23]:
```
1.0
```

In [24]:
```
from sklearn.metrics import precision_score
precision_score(y_test, pred)
```

Out[24]:
```
1.0
```

In [25]:
```
from sklearn.metrics import recall_score
recall_score(y_test, pred)
```

Out[25]:
```
1.0
```

In [26]:
```
from sklearn.metrics import f1_score
f1_score(y_test, pred)
```

Out[26]:
```
1.0
```

In [27]:
```
from sklearn.metrics import classification_report
print("Classification Report:\n", classification_report(y_test,pred))
```

```
Classification Report:
               precision    recall  f1-score   support

           0       1.00      1.00      1.00      1207
           1       1.00      1.00      1.00       186

    accuracy                           1.00      1393
   macro avg       1.00      1.00      1.00      1393
weighted avg       1.00      1.00      1.00      1393
```

In [28]:
```
from sklearn.metrics import mean_absolute_error
print("Mean Absolute Error(MSE)", mean_absolute_error(y_test,pred))
```

```
Mean Absolute Error(MSE) 0.0
```

In [29]:
```
from sklearn.metrics import mean_squared_error
print("Mean Squared Error(MSE)", mean_squared_error(y_test,pred))
```

```
Mean Squared Error(MSE) 0.0
```

In [30]:
```
print("RMSE=", np.sqrt(mean_squared_error(y_test,pred)))
```

```
RMSE= 0.0
```

In [32]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(df.Meassage,df.spam,test_size=0.2)
x_train.head()
```

Out[32]:
```
3413    No she didnt. I will search online and let you...
5393    All done, all handed in. Don't know if mega sh...
738     Hi. Customer Loyalty Offer:The NEW Nokia6650 M...
1381                              i dnt wnt to tlk wid u
2759              What time. I‘m out until prob 3 or so
Name: Meassage, dtype: object
```

In [33]:
```
from sklearn.feature_extraction.text import CountVectorizer
v=CountVectorizer()
x_train_count=v.fit_transform(x_train.values)
x_train_count.toarray()[:2]
```

Out[33]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [34]:
```
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(x_train_count,y_train)
```

Out[34]:
```
MultinomialNB()
```

In [35]:
```
emails = ["Hey mohan, can we get together to watch footbal game tomorrow?",
         "Upto 20% discount on parking, exclusive offer just for you.Dont miss this reward!"]
emails_count =v.transform(emails)
model.predict(emails_count)
```

Out[35]:
```
array([0, 1], dtype=int64)
```

In [36]:
```
x_test_count=v.transform(x_test)
model.score(x_test_count,y_test)
```

Out[36]:
```
0.989237668161435
```

In [37]:
```
model= svm.SVC()
accuracy =cross_val_score(model,inputs_n,target,scoring="accuracy",cv=10)
print(accuracy)
```

```
[0.8655914  0.8655914  0.86714542 0.86714542 0.86714542 0.86535009
 0.86535009 0.86535009 0.86535009 0.86535009]
```

In [39]:
```
print("Accuracy of model with cross validation is:",accuracy.mean()*100)
```

```
Accuracy of model with cross validation is: 86.59369510241115
```

In [42]:
```
categorical = [var for var in df.columns if df[var].dtype=='O']
print('There are {} categorical variables\n'.format(len(categorical)))
print("The categorical variables are :\n\n",categorical)
```

```
There are 2 categorical variables

The categorical variables are :

 ['category', 'Meassage']
```

In [44]:
```
for var in categorical:
    print(df[var].value_counts())
```

```
ham     4825
spam     747
Name: category, dtype: int64
Sorry, I'll call later
30
I cant pick the phone right now. Pls send a message
12
Ok...
10
Okie
4
Your opinion about me? 1. Over 2. Jada 3. Kusruthi 4. Lovable 5. Silent 6. Spl character 7. Not matured 8. Stylish 9. S
imple Pls reply..                        4

..
No. On the way home. So if not for the long dry spell the season would have been over
1
Urgent! Please call 09061743811 from landline. Your ABTA complimentary 4* Tenerife Holiday or £5000 cash await collecti
on SAE T&Cs Box 326 CW25WX 150ppm        1
Dear 0776xxxxxxx U've been invited to XCHAT. This is our final attempt to contact u! Txt CHAT to 86688 150p/MsgrcvdHG/S
uite342/2Lands/Row/W1J6HL LDN 18yrs      1
I think asking for a gym is the excuse for lazy people. I jog.
1
Rofl. Its true to its name
1
Name: Meassage, Length: 5169, dtype: int64
```

In [45]:
```
numerical = [var for var in df.columns if df[var].dtype !='O']
print('There are {} numerical variables\n'.format(len(numerical)))
print("The numerical variables are :\n\n",numerical)
```

```
There are 1 numerical variables

The numerical variables are :

 ['spam']
```

In [46]:
```
df[numerical].head()
```

Out[46]:

	spam
0	0
1	0
2	1
3	0
4	0

In [ ]: