

```
In [ ]:  
In [ ]:  
In [ ]:  
In [ ]:  
In [ ]: #SMS Classifier  
#develop a text classification model to classify SmS as either spam or non-spam using data science technique in python  
In [17]:  
import pandas as pd  
  
from sklearn.datasets import load_digits  
from sklearn.linear_model import LogisticRegression  
from sklearn.model_selection import train_test_split  
from sklearn.metrics import accuracy_score  
from sklearn.feature_extraction.text import TfidfVectorizer  
  
In [18]:  
print(harshi_mail_data)  
  
In [19]:  
Catergory Message Unnamed: 2 \n  
0 ham Go until jurong point, crazy.. Available only ... NaN  
1 ham Ok lar... Joking wif u oni... NaN  
2 spam Free entry in 2 a wkly comp to win FA Cup fina... NaN  
3 ham U dun say so early hor... U c already then say... NaN  
4 ham Nah I don't think he goes to usf, he lives aro... NaN  
... ... ... ... ... ...  
5567 spam This is the 2nd time we have tried 2 contact u... NaN  
5568 ham Will l^b going to esplanade fr home? NaN  
5569 ham Pity, * was in mood for that. So...any other s... NaN  
5570 ham The guy did some bitching but I acted like i'd... NaN  
5571 ham Rofl. Its true to its name NaN  
  
Unnamed: 3 Unnamed: 4 Unnamed: 5 Unnamed: 6 Unnamed: 7 Unnamed: 8 \n  
0 NaN NaN NaN NaN NaN NaN NaN  
1 NaN NaN NaN NaN NaN NaN NaN  
2 NaN NaN NaN NaN NaN NaN NaN  
3 NaN NaN NaN NaN NaN NaN NaN  
4 NaN NaN NaN NaN NaN NaN NaN  
... ... ... ... ... ...  
5567 NaN NaN NaN NaN NaN NaN NaN  
5568 NaN NaN NaN NaN NaN NaN NaN  
5569 NaN NaN NaN NaN NaN NaN NaN  
5570 NaN NaN NaN NaN NaN NaN NaN  
5571 NaN NaN NaN NaN NaN NaN NaN  
  
Unnamed: 9 ... Unnamed: 28 Unnamed: 29 Unnamed: 30 Unnamed: 31 \n  
0 NaN ... NaN NaN NaN NaN NaN  
1 NaN ... NaN NaN NaN NaN NaN  
2 NaN ... NaN NaN NaN NaN NaN  
3 NaN ... NaN NaN NaN NaN NaN  
4 NaN ... NaN NaN NaN NaN NaN  
... ... ... ... ... ...  
5567 NaN ... NaN NaN NaN NaN NaN  
5568 NaN ... NaN NaN NaN NaN NaN  
5569 NaN ... NaN NaN NaN NaN NaN  
5570 NaN ... NaN NaN NaN NaN NaN  
5571 NaN ... NaN NaN NaN NaN NaN  
  
Unnamed: 32 Unnamed: 33 Unnamed: 34 Unnamed: 35 Unnamed: 36 Unnamed: 37  
0 NaN NaN NaN NaN NaN NaN NaN  
1 NaN NaN NaN NaN NaN NaN NaN  
2 NaN NaN NaN NaN NaN NaN NaN  
3 NaN NaN NaN NaN NaN NaN NaN  
4 NaN NaN NaN NaN NaN NaN NaN  
... ... ... ... ... ...  
5567 NaN ... NaN NaN NaN NaN NaN  
5568 ... ...  
5569 ... ...  
5570 ... ...  
5571 ... ...  
  
Unnamed: 32 Unnamed: 33 Unnamed: 34 Unnamed: 35 Unnamed: 36 Unnamed: 37  
0 ...  
1 ...  
2 ...  
3 ...  
4 ...  
... ... ... ... ... ...  
5567 ...  
5568 ...  
5569 ...  
5570 ...  
5571 ...  
  
[5572 rows × 38 columns]  
In [20]:  
mail_data=harshi_mail_data.where((pd.notnull(harshi_mail_data)), "")  
print(mail_data)  
  
Catergory Message Unnamed: 2 \n  
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...  
...  
5567 spam This is the 2nd time we have tried 2 contact u...  
5568 ham Will l^b going to esplanade fr home?  
5569 ham Pity, * was in mood for that. So...any other s...  
5570 ham The guy did some bitching but I acted like i'd...  
5571 ham Rofl. Its true to its name  
  
Unnamed: 3 Unnamed: 4 Unnamed: 5 Unnamed: 6 Unnamed: 7 Unnamed: 8 \n  
0 ...  
1 ...  
2 ...  
3 ...  
4 ...  
... ... ... ... ... ...  
5567 ...  
5568 ...  
5569 ...  
5570 ...  
5571 ...  
  
[5572 rows × 38 columns]  
In [21]:  
mail_data.head()  
  
Out[21]:  
Catergory Message Unnamed: 2 Unnamed: 3 Unnamed: 4 Unnamed: 5 Unnamed: 6 Unnamed: 7 Unnamed: 8 Unnamed: 9 ... Unnamed: 28 Unnamed: 29 Unnamed: 30 Unnamed: 31 Unnamed: 32 Unnamed: 33 Unnamed: 34 Unnamed: 35 Unnamed: 36 Unnamed: 37  
0 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...  
...  
5567 ...  
5568 ...  
5569 ...  
5570 ...  
5571 ...  
  
[5572 rows × 38 columns]  
In [22]:  
mail_data.shape  
Out[22]: (5572, 38)  
  
In [23]:  
mail_data.loc[mail_data['Catergory']=="spam", 'Catergory']=0  
mail_data.loc[mail_data['Catergory']=="ham", 'Catergory']=1  
mail_data.head()  
  
Out[23]:  
Catergory Message Unnamed: 2 Unnamed: 3 Unnamed: 4 Unnamed: 5 Unnamed: 6 Unnamed: 7 Unnamed: 8 Unnamed: 9 ... Unnamed: 28 Unnamed: 29 Unnamed: 30 Unnamed: 31 Unnamed: 32 Unnamed: 33 Unnamed: 34 Unnamed: 35 Unnamed: 36 Unnamed: 37  
0 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...  
...  
5567 ...  
5568 ...  
5569 ...  
5570 ...  
5571 ...  
  
[5572 rows × 38 columns]  
In [24]:  
X=mail_data['Message']  
Y=mail_data['Catergory']  
print(X)  
print(Y)  
0 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...  
...  
5567 This is the 2nd time we have tried 2 contact u...  
5568 Will l^b going to esplanade fr home?  
5569 Pity, * was in mood for that. So...any other s...  
5570 The guy did some bitching but I acted like i'd...  
5571 ham Rofl. Its true to its name  
Name: Message, Length: 5572, dtype: object  
  
In [25]:  
X_train,X_test,Y_train,Y_test=train_test_split(X,Y,test_size=0.2,random_state=3)  
  
In [26]:  
print(X.shape)  
(5572,)  
  
In [27]:  
print(X_train.shape)  
print(X_test.shape)  
(4457,)  
(1115,)  
  
In [28]:  
from sklearn.feature_extraction.text import TfidfVectorizer  
  
feature_extraction = TfidfVectorizer(min_df=1, stop_words='english')  
X_train_features = feature_extraction.fit_transform(X_train)  
X_test_features = feature_extraction.transform(X_test)  
  
In [29]:  
Y_train=Y_train.astype('int')  
Y_test=Y_test.astype('int')  
  
In [30]:  
print(X_train)  
  
3075 Mum, hope you are having a great day. Hoping t...  
1287 Yes;)sura in sun tv.)lol.  
1614 Me sef dey laugh you. Meanwhile how's my darli...  
4384 Yo come over carlos will be here soon  
3266 Ok then i come n pick u at engin?  
...  
789 Gud mrrng dear hav a nice day  
968 Are you willing to go aptitude class.  
1687 So now my dad is gonna call after he gets out ...  
3221 OK darlin i suppose it won't ok just worry too ...  
1688 Non sonakha soladha. Why boss?  
Name: Message, Length: 4457, dtype: object  
  
In [31]:  
print(X_train_features)  
  
0.321454264242395  
(0, 3884) 0.2423597379249671  
(0, 4193) 0.38855738830457154  
(0, 4446) 0.329265563946227548  
(0, 3322) 0.321454264242395  
(0, 2891) 0.321454242127719965  
(0, 2760) 0.321454242127719965  
(0, 3202) 0.2582838623411723  
(0, 3316) 0.218726811101663  
(0, 4408) 0.29985303337317463  
(1, 3963) 0.380431198316959  
(1, 6714) 0.4306815894277422  
(1, 6266) 0.47691368959409388  
(1, 6291) 0.5652599076654626  
(1, 6276) 0.30807360532391593  
(2, 939) 0.48842623204745973  
(2, 2978) 0.426915061199681973  
(2, 3822) 0.398167475889595836  
(2, 2191) 0.426815061199681973  
(2, 5686) 0.4884262848745973  
(3, 5996) 0.4916953116975721  
(3, 1573) 0.5918958247951586  
(3, 706) 0.5194764126185371  
(3, 2497) 0.741245144160404  
...  
(4452, 2091) 0.3098883537962126  
(4452, 988) 0.67601290131282  
(4452, 7108) 0.5871739591782677  
(4452, 1735) 0.45610895640982965  
(4452, 622) 0.5871739591782677  
(4452, 2056) 0.38087360532391593  
(4454, 3824) 0.34468931682024  
(4454, 1971) 0.41658721483064375  
(4454, 1035) 0.3192403772890263  
(4454, 7181) 0.3123895049617535  
(4454, 5242) 0.4266820272147833  
(4454, 1131) 0.39966211742483932  
(4454, 5997) 0.5871739591782677  
(4454, 5999) 0.5304350313291551  
(4454, 5898) 0.5304350313291551  
(4454, 1366) 0.4469836316446079  
(4454, 4450) 0.48821933148688146  
  
In [32]:  
model=LogisticRegression()  
  
In [33]:  
model.fit(X_train_features,Y_train)  
  
Out[33]: LogisticRegression()  
LogisticRegression()  
  
In [34]:  
prediction_on_training_data=model.predict(X_train_features)  
accuracy_on_training_data=accuracy_score(Y_train,prediction_on_training_data)  
  
In [35]:  
print(accuracy_on_training_data)  
0.9861207089970832  
  
In [36]:
```