

```
In [ ]:
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
import numpy as np
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]: harsh_mail_data=pd.read_csv('spam.csv')

In [19]: print(harshi_mail_data)

      Categorty      Message  Unnamed: 2  \
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 [ ] 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  \
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: 9      ...  Unnamed: 28  Unnamed: 29  Unnamed: 30  Unnamed: 31  \
0      NaN      ...      NaN      NaN      NaN      NaN
1      NaN      ...      NaN      NaN      NaN      NaN
2      NaN      ...      NaN      NaN      NaN      NaN
3      NaN      ...      NaN      NaN      NaN      NaN
4      NaN      ...      NaN      NaN      NaN      NaN
...      ...      ...      ...      ...      ...      ...
5567      NaN      ...      NaN      NaN      NaN      NaN
5568      NaN      ...      NaN      NaN      NaN      NaN
5569      NaN      ...      NaN      NaN      NaN      NaN
5570      NaN      ...      NaN      NaN      NaN      NaN
5571      NaN      ...      NaN      NaN      NaN      NaN

      Unnamed: 32  Unnamed: 33  Unnamed: 34  Unnamed: 35  Unnamed: 36  Unnamed: 37
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

[5572 rows x 38 columns]

In [20]: mail_data=harshi_mail_data.where((pd.notnull(harshi_mail_data)), "")
print(mail_data)

      Categorty      Message  Unnamed: 2  \
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 [ ] 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  \
0      NaN      ...      NaN      NaN      NaN      NaN
1      NaN      ...      NaN      NaN      NaN      NaN
2      NaN      ...      NaN      NaN      NaN      NaN
3      NaN      ...      NaN      NaN      NaN      NaN
4      NaN      ...      NaN      NaN      NaN      NaN
...      ...      ...      ...      ...      ...      ...
5567      NaN      ...      NaN      NaN      NaN      NaN
5568      NaN      ...      NaN      NaN      NaN      NaN
5569      NaN      ...      NaN      NaN      NaN      NaN
5570      NaN      ...      NaN      NaN      NaN      NaN
5571      NaN      ...      NaN      NaN      NaN      NaN

      Unnamed: 32  Unnamed: 33  Unnamed: 34  Unnamed: 35  Unnamed: 36  Unnamed: 37
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

[5572 rows x 38 columns]

In [21]: mail_data.head()

Out[21]:      Categorty      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      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...      ...

5 rows x 38 columns

In [22]: mail_data.shape

Out[22]: (5572, 38)

In [23]: mail_data.loc[mail_data['Categorty']=='spam', 'Categorty',]=0
mail_data.loc[mail_data['Categorty']=='ham', 'Categorty',]=1
mail_data.head()

Out[23]:      Categorty      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      1      Go until jurong point, crazy.. Available only ...      ...
1      1      Ok lar... Joking wif u oni...      ...
2      0      Free entry in 2 a wkly comp to win FA Cup fina...      ...
3      1      U dun say so early hor... U c already then say...      ...
4      1      Nah I don't think he goes to usf, he lives aro...      ...

5 rows x 38 columns

In [24]: X=mail_data['Message']
Y=mail_data['Categorty']
print(X)
print(Y)

0      Go until jurong point, crazy.. Available only ...
1      Ok lar... Joking wif u oni...
2      Free entry in 2 a wkly comp to win FA Cup fina...
3      U dun say so early hor... U c already then say...
4      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 [ ] 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      Rofl. Its true to its name
Name: Message, Length: 5572, dtype: object

0      1
1      1
2      0
3      1
4      1
...
5567      0
5568      1
5569      1
5570      1
5571      1
Name: Categorty, 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...
1787      Yes:ysura in sun tv:)lol.
1614      Me sef dey laugh you. Meanwhile how's my darlin...
4304      Yo come over carlos will be here soon
3266      Ok then i come n pick u at engin?
...
789      Gud mrng dear hav a nice day
968      Are you willing to go for aptitude class.
1667      So now my dad is gonna call after he gets out ...
3221      Ok darlin I supose it was ok i just worry too ...
1688      Nan sonathaya soladha. Why boss?
Name: Message, Length: 4457, dtype: object

In [31]: print(X_train_features)

(0, 736)      0.3214544264242305
(0, 3884)      0.24235973792424671
(0, 4193)      0.38855738830457154
(0, 6446)      0.28266563946227548
(0, 3322)      0.3214544264242305
(0, 2091)      0.38607442127710905
(0, 3081)      0.46404785982379829
(0, 3262)      0.2583838622411723
(0, 3316)      0.2187268141001663
(0, 4408)      0.299530337317463
(1, 3963)      0.390431198316959
(1, 6714)      0.4396815894277422
(1, 6266)      0.47691368959540388
(1, 6291)      0.5652599076054626
(1, 7276)      0.50596913070320353
(2, 924)      0.48842628248745973
(2, 2078)      0.42681506199681973
(2, 3822)      0.39816747588955836
(2, 2191)      0.42681506199681973
(2, 5686)      0.48842628248745973
(3, 5996)      0.4916053116975721
(3, 1573)      0.5918858247851686
(3, 1815)      0.3718071728324976
(3, 7286)      0.5194704125185371
(4, 2487)      0.7412451441160404
:      :
(4452, 2091)      0.3098883537962126
(4453, 988)      0.6766129013031282
(4453, 7108)      0.5787739591782677
(4453, 1735)      0.45610005640082985
(4454, 2975)      0.4268020272147833
(4454, 2856)      0.3898736622039193
(4454, 3034)      0.34466931683024
(4454, 1971)      0.41658721403064375
(4454, 1035)      0.3192403772890263
(4454, 7181)      0.3123895049617535
(4454, 5242)      0.4268020272147833
(4455, 1131)      0.3896621174283932
(4455, 6282)      0.3896621174283932
(4455, 6213)      0.257925326934647
(4455, 2716)      0.3223083713845921
(4455, 7193)      0.29135635668405085
(4455, 7241)      0.30260038531595883
(4455, 2077)      0.3133901879162373
(4455, 4150)      0.3059160411914761
(4455, 3678)      0.1684891835057733
(4455, 4659)      0.35953417130487636
(4456, 5973)      0.5304350313291551
(4456, 5989)      0.5304350313291551
(4456, 1366)      0.4460636315446079
(4456, 4450)      0.48821933148688146

In [32]: model=LogisticRegression()

In [33]: model.fit(X_train_features,Y_train)

Out[33]: LogisticRegression
LogisticRegression()

In [ ]:

In [35]: prediction_on_training_data=model.predict(X_train_features)
accuracy_on_training_data=accuracy_score(Y_train,prediction_on_training_data)

In [36]: print(accuracy_on_training_data)

0.9661207089970832

In [ ]:
```