

TASK2:-EMAIL SPAM DETECTION WITH MACHINE LEARNIG

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df=pd.read_csv("C:/Users/HP/Desktop/oasis/spam.csv",encoding='latin')
df
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN
...
5567	spam	This is the 2nd time we have tried 2 contact u...	NaN	NaN	NaN
5568	ham	Will ò_b going to esplanade fr home?	NaN	NaN	NaN
5569	ham	Pity, * was in mood for that. So...any other s...	NaN	NaN	NaN
5570	ham	The guy did some bitching but I acted like i'd...	NaN	NaN	NaN
5571	ham	Rofl. Its true to its name	NaN	NaN	NaN

5572 rows × 5 columns

```
df.head(20)
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN
5	spam	FreeMsg Hey there darling it's been 3 week's n...	NaN	NaN	NaN
6	ham	Even my brother is not like to speak with me. ...	NaN	NaN	NaN
7	ham	As per your request 'Melle Melle (Oru Minnamin...	NaN	NaN	NaN
8	spam	WINNER!! As a valued network customer you have...	NaN	NaN	NaN
9	spam	Had your mobile 11 months or more? U R entitle...	NaN	NaN	NaN
10	ham	I'm gonna be home soon and i don't want to tal...	NaN	NaN	NaN
11	spam	SIX chances to win CASH! From 100 to 20,000 po...	NaN	NaN	NaN
12	spam	URGENT! You have won a 1 week FREE membership ...	NaN	NaN	NaN
13	ham	I've been searching for the right words to tha...	NaN	NaN	NaN
14	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	NaN	NaN	NaN
15	spam	XXXMobileMovieClub: To use your credit, click ...	NaN	NaN	NaN
16	ham	Oh k...i'm watching here:)	NaN	NaN	NaN
17	ham	Eh u remember how 2 spell his name... Yes i di...	NaN	NaN	NaN
18	ham	Fine if thatâs the way u feel. Thatâs the wa...	NaN	NaN	NaN
19	spam	England v Macedonia - dont miss the goals/team...	NaN	NaN	NaN

```
df.shape
```

```
(5572, 5)
```

```
df.isnull()
```

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	False	False	True	True	True
1	False	False	True	True	True
2	False	False	True	True	True
3	False	False	True	True	True
4	False	False	True	True	True
...
5567	False	False	True	True	True
5568	False	False	True	True	True
5569	False	False	True	True	True
5570	False	False	True	True	True
5571	False	False	True	True	True

5572 rows × 5 columns

```
df.isnull().sum()
```

```
v1          0
v2          0
Unnamed: 2   5522
Unnamed: 3   5560
Unnamed: 4   5566
dtype: int64
```

```
df.columns
```

```
Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')
```

```
df.isna().mean()
```

```
v1          0.000000
v2          0.000000
Unnamed: 2   0.991027
Unnamed: 3   0.997846
Unnamed: 4   0.998923
dtype: float64
```

```
df.drop(columns=['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis=0, inplace=True)
df
```

	v1	v2
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 Ì_ 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

5572 rows × 2 columns

```
df.shape
```

(5572, 2)

```
df.columns= ['spam/ham', 'SMS']
```

```
df.loc[df['spam/ham']=='spam', 'spam/ham',]=0
df.loc[df['spam/ham']=='ham', 'spam/ham',]=1
df
```

	spam/ham	SMS
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...
...
5567	0	This is the 2nd time we have tried 2 contact u...
5568	1	Will Ì_ b going to esplanade fr home?
5569	1	Pity, * was in mood for that. So...any other s...
5570	1	The guy did some bitching but I acted like i'd...
5571	1	Rofl. Its true to its name

5572 rows × 2 columns

```
x=df.SMS
```

```
x=df.SMS
```

```
x
```

```
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 i_ 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: SMS, Length: 5572, dtype: object
```

```
y=df['spam/ham']
```

```
y
```

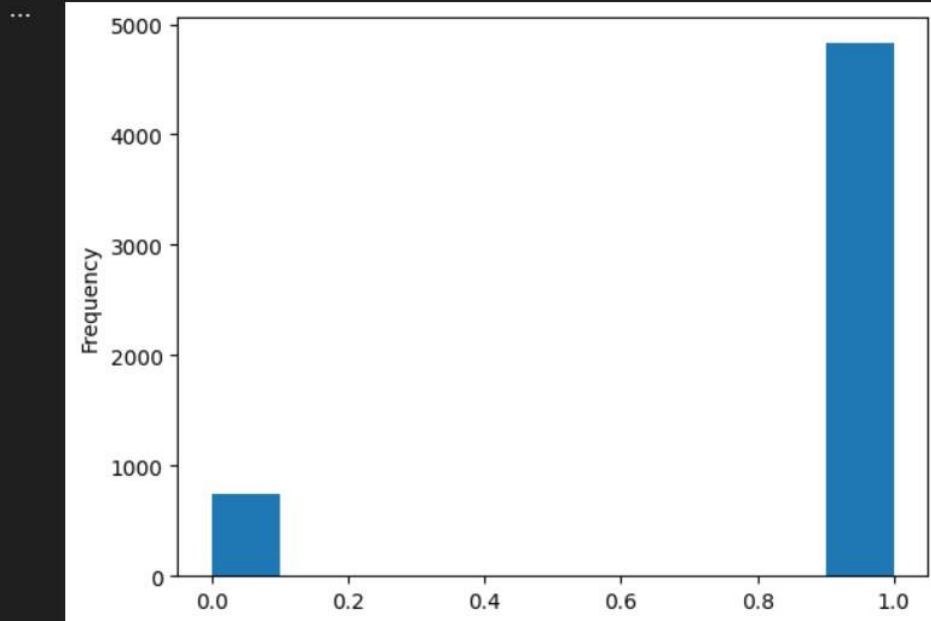
```
0      1
1      1
2      0
3      1
4      1
      ..
5567    0
5568    1
5569    1
5570    1
5571    1
```

```
Name: spam/ham, Length: 5572, dtype: object
```

```
df["spam/ham"].plot.hist()
```

[15]

```
... <AxesSubplot:ylabel='Frequency'>
```



```
> plt.bar(df["SMS"],df["spam/ham"],color='cyan')
plt.title("spam")
[16]
... Text(0.5, 1.0, 'spam')
... C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 137 (\x89) missing from current font.
fig.canvas.print_figure(bytes_io, **kw)
) missing from current font.ograms\Python\Python310\lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 13 (
fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 142 (\x8e) missing from current font.
fig.canvas.print_figure(bytes_io, **kw)
C:\Users\HP\AppData\Local\Programs\Python\Python310\lib\site-packages\IPython\core\pylabtools.py:152: UserWarning: Glyph 139 (\x8b) missing from current font.
fig.canvas.print_figure(bytes_io, **kw)
```



```
> sns.set_style("whitegrid")
plt.figure(figsize=(5,5))
sns.distplot(df["spam/ham"],color="red")
plt.show()
```

[08]

C:\Users\HP\AppData\Local\Temp\ipykernel_9224\1522660812.py:3: UserWarning:

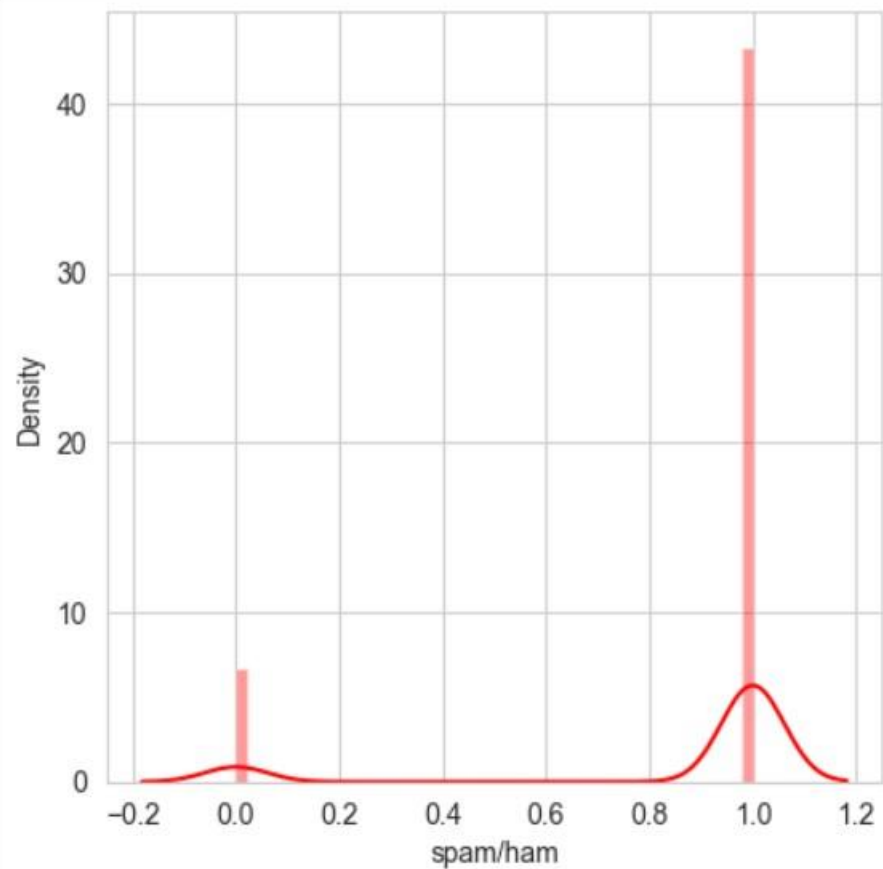
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see

<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>

```
sns.distplot(df["spam/ham"],color="red")
```



```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import confusion_matrix, accuracyscore, classification_report
```

```
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_state=3)
print(x.shape)
```

```
(5572,)
```

[+ Code](#)[+ Markdown](#)

```
print(xtrain.shape)
```

```
(4457,)
```

```
print(xtest.shape)
```

```
(1115,)
```

```
print(ytrain.shape)
```

```
(4457,)
```

```
print(ytest.shape)
```

```
(1115,)
```

```
xtrain,xtest
```

```
(3075    Mum, hope you are having a great day. Hoping t...
1787      Yes:)sura in sun tv:)lol.
1614    Me sef dey laugh you. Meanwhile how's my darli...
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 ...
3321    Ok darlin i supose it was ok i just worry too ...
1688      Nan sonathaya soladha. Why boss?
Name: SMS, Length: 4457, dtype: object,
2632      I WILL CAL YOU SIR. In meeting
454      Loan for any purpose â€500 - â€75,000. Homeown...
983      LOOK AT THE FUCKIN TIME. WHAT THE FUCK YOU THI...
1282    Ever green quote ever told by Jerry in cartoon...
4610      Wat time i_ finish?
...
4827    Lol no. Just trying to make your day a little ...
5291      Xy trying smth now. U eat already? We havent...
3325    Huh so fast... Dat means u havent finished pai...
3561    Still chance there. If you search hard you wil...
1136    Dont forget you can place as many FREE Request...
Name: SMS, Length: 1115, dtype: object)
```

```
ytrain,ytest
```

```
(3075    1
1787    1
1614    1
4304    1
3266    1
...
789     1
968     1
1667    1
3321    1
1688    1
Name: spam/ham, Length: 4457, dtype: object,
2632    1
454     0
983     1
1282    1
4610    1
...
4827    1
5291    1
3325    1
3561    1
1136    0
Name: spam/ham, Length: 1115, dtype: object)
```



```
feat_vect=TfidfVectorizer(stop_words='english',lowercase=True)
feat_vect
```

[37]

```
...
▼ TfidfVectorizer
TfidfVectorizer(stop_words='english')
```

```
ytrain=ytrain.astype('int')
ytest=ytest.astype('int')
```

[38]

```
xtrain_vec=feat_vect.fit_transform(xtrain)
```

[39]

```
xtest_vec=feat_vect.transform(xtest)
print(xtrain)
```

[40]

```
... 3075 Mum, hope you are having a great day. Hoping t...
1787 Yes:)sura in sun tv.:)lol.
1614 Me sef dey laugh you. Meanwhile how's my darli...
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 ...
3321 Ok darlin i supose it was ok i just worry too ...
1688 Nan sonathaya soladha. Why boss?
Name: SMS, Length: 4457, dtype: object
```



```
xtrain_vec
print("xtrain--->",xtrain_vec)
print("xtest--->",xtest_vec)
```

42]

```
· xtrain---> (0, 741) 0.3219352588930141
(0, 3979) 0.2410582143632299
(0, 4296) 0.3891385935794867
(0, 6599) 0.20296878731699391
(0, 3386) 0.3219352588930141
(0, 2122) 0.38613577623520473
(0, 3136) 0.440116181574609
(0, 3262) 0.25877035357606315
(0, 3380) 0.21807195185332803
(0, 4513) 0.2909649098524696
(1, 4061) 0.380431198316959
(1, 6872) 0.4306015894277422
(1, 6417) 0.4769136859540388
(1, 6442) 0.5652509076654626
(1, 7443) 0.35056971070320353
(2, 933) 0.4917598465723273
(2, 2109) 0.42972812260098503
(2, 3917) 0.40088501350982736
(2, 2226) 0.413484525934624
(2, 5825) 0.4917598465723273
(3, 6140) 0.4903863168693604
(3, 1599) 0.5927091854194291
(3, 1842) 0.3708680641487708
(3, 7453) 0.5202633571003087
(4, 2531) 0.7419319091456392
...
(1114, 3564) 0.40844238751288037
(1114, 2899) 0.2421646568502054
(1114, 2862) 0.38140394975458775
(1114, 2352) 0.270495916357943
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)

```
44] model= LogisticRegression()
```

```
45] model.fit(xtrain_vec,ytrain)
```

```
-- ▾ LogisticRegression  
LogisticRegression()
```

```
47] model.score(xtrain_vec,ytrain)
```

```
-- 0.9661207089970832
```

```
48] model.score(xtest_vec,ytest)
```

```
-- 0.9623318385650225
```

```
49] predict_model=model.predict(xtest_vec)  
predict_model
```

```
-- array([1, 1, 1, ..., 1, 1, 1])
```

```
50] accuracy_score(ytest,predict_model)
```

```
-- 0.9623318385650225
```

```
accuracy_score(ytest,predict_model)
```

```
50]
```

```
.. 0.9623318385650225
```

```
confusion_matrix(ytest,predict_model)
```

```
51]
```

```
.. array([[114,  41],  
         [  1, 959]], dtype=int64)
```

```
print(classification_report(ytest,predict_model))
```

```
53]
```

```
..
```

	precision	recall	f1-score	support
0	0.99	0.74	0.84	155
1	0.96	1.00	0.98	960
accuracy			0.96	1115
macro avg	0.98	0.87	0.91	1115
weighted avg	0.96	0.96	0.96	1115