

CO

SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

+ Code + Text

RAM Disk Colab AI

IMPORTING LIBRARIES

[1]

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

LOADING DATASET

[2]

```
df= pd.read_csv('/content/spam.csv')
df
```

| | Category | Message |
|---|----------|---|
| 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... |

0s

completed at 19:59

+ Code + Text

df.head()


| | Category | Message |
|---|----------|---|
| 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... |

Next steps: [Generate code with df](#) [View recommended plots](#)

[4] df.tail()

| | Category | Message |
|------|----------|---|
| 5567 | spam | This is the 2nd time we have tried 2 contact u... |
| 5568 | ham | Will u 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 | Definitely true to its name |

0s completed at 19:59

 SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment

Share

+ Code + Text

✓ RAM
Disk

Colab AI

df.columns

Index(['Category', 'Message'], dtype='object')

[6] df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
Column Non-Null Count Dtype
--- ---
0 Category 5572 non-null object
1 Message 5572 non-null object
dtypes: object(2)
memory usage: 87.2+ KB

[7] df.isna().sum()

Category 0
Message 0
dtype: int64

df.shape

(5572, 2)

0s completed at 19:59

+ Code + Text

RAM Disk Colab AI

↑ ↓ ↺ 🗨️ 📄 🗑️ ⋮

▼ LABEL ENCODING


+ Code + Text

[] df['Category'].replace({'spam':1, 'ham':0}, inplace=True)
df

| | Category | Message |
|------|----------|---|
| 0 | 0 | Go until jurong point, crazy.. Available only ... |
| 1 | 0 | Ok lar... Joking wif u oni... |
| 2 | 1 | Free entry in 2 a wkly comp to win FA Cup fina... |
| 3 | 0 | U dun say so early hor... U c already then say... |
| 4 | 0 | Nah I dont think he goes to usf, he lives aro... |
| ... | ... | ... |
| 5567 | 1 | This is the 2nd time we have tried 2 contact u... |
| 5568 | 0 | Will u b going to esplanade fr home? |
| 5569 | 0 | Pity, * was in mood for that. So...any other s... |
| 5570 | 0 | The guy did some bitching but I acted like I'd... |
| 5571 | 0 | D-f he said to me... |

0s completed at 19:59

● ✕

 SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings User

RAM Disk Colab AI

+ Code + Text

↑ ↓ ↺ ⌨ 📄 🗑 ⋮

▼ SPLITTING DATASET INTO X AND Y VARIABLES

```
[ ] x=df['Message']
    y=df['Category']

print(x)

0      so 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 u 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

[ ] print(y)

0      0
1      0
```

0s completed at 19:59

+ Code + Text

✓ RAM Disk Colab AI

▶ print(y)

0 0
1 0
2 1
3 0
4 0
..
5567 1
5568 0
5569 0
5570 0
5571 0
Name: Category, Length: 5572, dtype: int64

▼ SPLITTING INTO TESTING AND TRAINING DATA

[] from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=42)
x_train

708 Quite late lar... And 12 anyway i wun b drivin...
4338 on a Tuesday night r u 4 real
5029 Go chase after her and run her over while she'...
4921 6 says you never answer your texts, confirm/deny
.....

0s completed at 19:59

● ✕

+ Code + Text

✓ RAM Disk Colab AI

x_test

```
3245 Squeeeeee!! This is christmas hug.. If u lik ...
944 And also I've sorta blown him off a couple tim...
1044 Mmm thats better now i got a roast down me! i@...
2484 I'm have some kanji dont eat anything heavy ok
812 So there's a ring that comes with the guys cos...
...
2505 Hello, my boytoy! I made it home and my consta...
2525 FREE entry into our £250 weekly comp just send...
4975 Aiyo u so poor thing... Then u dun wan 2 eat? ...
660 You have won ?1,000 cash or a ?2,000 prize! To...
4463 Sorry I flaked last night, shit's seriously go...
Name: Message, Length: 1672, dtype: object
```

y_train

| | |
|------|---|
| 708 | 0 |
| 4338 | 0 |
| 5029 | 0 |
| 4921 | 0 |
| 2592 | 0 |
| .. | |
| 3772 | 0 |
| 5191 | 0 |
| 5226 | 0 |
| 5390 | 0 |
| 860 | 0 |

Name: Series, Length: 3000, dtype: int64

+ Code + Text

✓ RAM Disk Colab AI

y_train

708 0

4338 0

5029 0

4921 0

2592 0

..

3772 0

5191 0

5226 0

5390 0

860 0

Name: Category, Length: 3900, dtype: int64

y_test

3245 0

944 0

1044 0

2484 0

812 0

..

2505 0

2525 1


4975 0

650 1

4463 0

Name: Category, Length: 1073, dtype: int64

0s completed at 19:59

 SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

+ Code + Text

✓ RAM Disk Colab AI

```
print(x_train.shape)
print(x_test.shape)
```

```
(3900, 7168)
(1072, 7168)
```

```
nlk.download('wordnet')
nlk.download('stopwords')
nlk.download('punkt')
```

```
[nlk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True
```

```
[ ] variable=df.Message
variable
```

```
0      Go until junong 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  ...
```

0s completed at 19:59

+ Code + Text

RAM Disk Colab AI

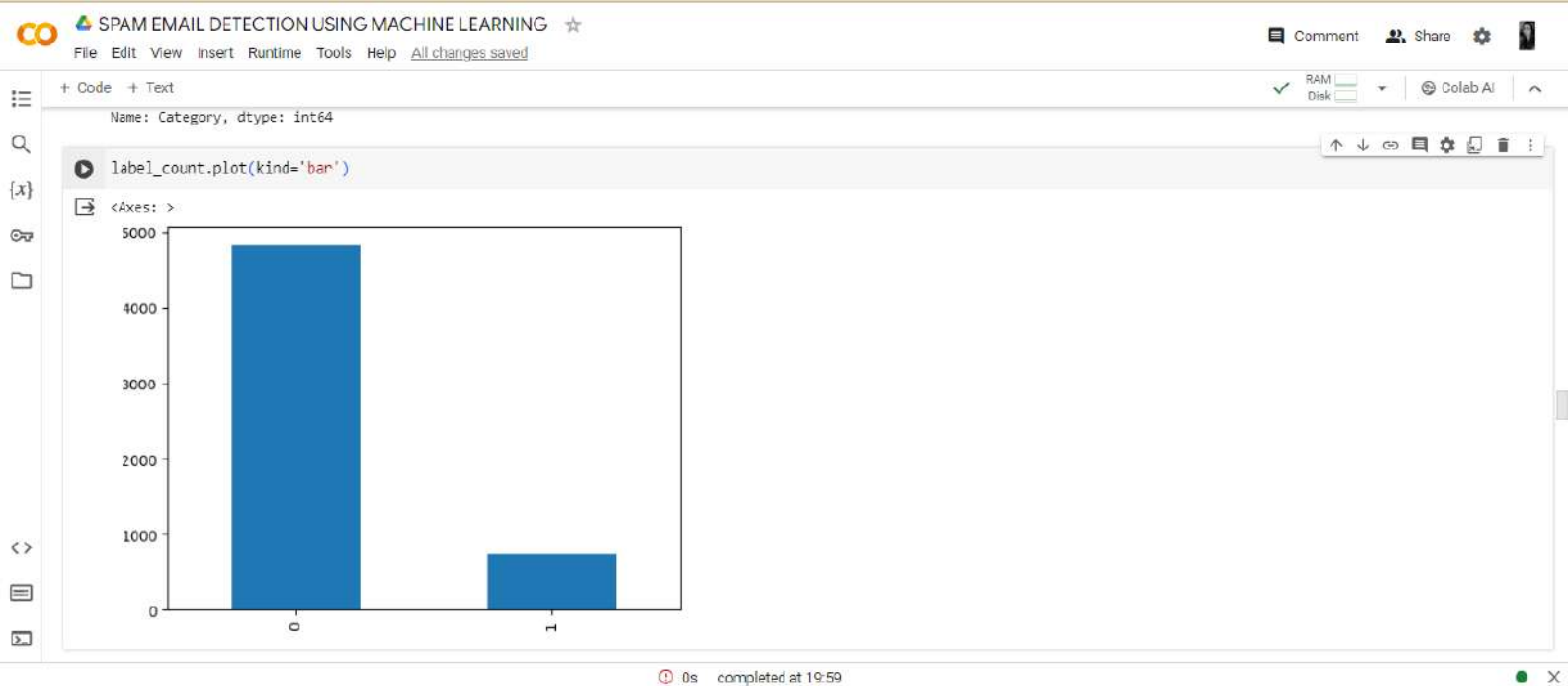
```
variable=variable.str.replace('[^a-zA-Z0-9]+',' ')
variable
```

```
<ipython-input-17-476e3d23ab92>:1: FutureWarning: The default value of regex will change from True to False in a future version.
variable=variable.str.replace('[^a-zA-Z0-9]+',' ')
0      Go until jurong point crazy Available only in ...
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 arou...
...
5567    This is the 2nd time we have tried 2 contact u...
5568          Will b going to esplanade fr home
5569    Pity wes in mood for that So any other suggest...
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
```

VALUE COUNTS

```
[ ] label_count=df['Category'].value_counts()
label_count
```

```
0      4825
1        747
Name: Category, dtype: int64
```



+ Code + Text

RAM Disk Colab AI

df.Message.value_counts()

```
Sorry, I'll call later. 30
I cant pick the phone right now. Pls send a message 12
Ok... 10
Ok. 4
Ok. 4
..
I gotta collect da car at 6 lei. 1
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 A&TA complimentary 4* Tenerife Holiday or £5000 cash await collection SAE T&Cs Box 326 CW25WX 150ppm 1
I think asking for a gym is the excuse for lazy people. I jog. 1
Rofl. Its true to its name 1
Name: Message, Length: 5157, dtype: int64
```

[] variable1=df.Message
variable1

```
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 u 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...
```

CO

SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help All changes saved

Comment Share Settings

+ Code + Text

RAM Disk Colab AI

```
from nltk import TweetTokenizer
tt=TweetTokenizer()
variable1=variable1.apply(lambda x:tt.tokenize(x)).apply(lambda x:" ".join(x))
variable1
```

```
0      Go until jurong point , crazy .. Available onl...
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 sa...
4      Nah I don't think he goes to usf , he lives ar...
...
5567     This is the 2nd time we have tried 2 contact u...
5568             Will u b going to esplanade fr home ?
5569     Pity , * was in mood for that . So ... any oth...
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
```

```
[ ] variable1=variable1.str.replace('[^a-zA-Z0-9]+',' ')
variable1
```

```
<ipython-input-24-292cb2b60ee0>:1: FutureWarning: The default value of regex will change from True to False in a future version.
  variable1=variable1.str.replace('[^a-zA-Z0-9]+',' ')
0      Go until jurong point crazy Available only in ...
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 arou...
```

0s completed at 19:59

CO

SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

+ Code + Text

RAM Disk Colab AI

STEMMING

```
from nltk.stem import SnowballStemmer
sbs=SnowballStemmer('english')
variable1=variable1.apply(lambda x:[sbs.stem(i.lower()) for i in tt.tokenize(x)]).apply(lambda x:" ".join(x))
variable1
```

```
0      go until jurong point crazi avail onli in bugi...
1              ok lor joke wif u oni
2      free entri in 2 a wkli comp to win fa cup fina...
3          u dun say so earli hor u c already then say
4      nah i don t think he goe to usf he live around...
...
5567     this is the 2nd time we have tri 2 contact u u...
5568             will b go to esplanad fr home
5569         piti was in mood for that so ani other suggest
5570     the guy did some bitch but i act like i d be i...
5571             rofl it true to it name
Name: Message, Length: 5572, dtype: object
```

```
[ ] from nltk.corpus import stopwords
sw=stopwords.words('english')
variable1=variable1.apply(lambda x:[i for i in tt.tokenize(x) if i not in sw]).apply(lambda x:" ".join(x))
variable1
```

0s completed at 19:59

 SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Colab AI

+ Code + Text

name: message, Length: 5572, dtype: object

```
from nltk.corpus import stopwords
sw=stopwords.words('english')
variable1=variable1.apply(lambda x:[i for i in tt.tokenize(x) if i not in sw]).apply(lambda x:" ".join(x))
variable1
```

```
0      go jurong point crozi avail onli bugi n great ...
1      ok lar joke wif u oni
2      free entri 2 wkli comp win fa cup final tkts 2...
3      u dun say earli hor u c already say
4      nah think goe usf live around though
...
5567    2nd time tri 2 contact u u 750 pound prize 2 c...
5568      b go esplanad fr home
5569    pitl mood ani suggest
5570    guy bitch act like interest buy someth els nex...
5571      rofl true name
Name: Message, Length: 5572, dtype: object
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
vector=TfidfVectorizer()
train_data=vector.fit_transform(variable1)
train_data
```

<5572x7168 sparse matrix of type '<class 'numpy.float64''>'
with 47260 stored elements in Compressed Sparse Row format>

0s completed at 19:59

 Colab AI

● ×

+ Code + Text

Reconnect Colab AI

▶

y = df['Category']
y

📄

0 0
1 0
2 1
3 0
4 0
..
5567 1
5568 0
5569 0
5570 0
5571 0
Name: Category, Length: 5572, dtype: int64

[]

x_train,x_test,y_train,y_test=train_test_split(train_data,y,test_size=0.30,random_state=1)

[]


print(x_train.shape)
print(x_test.shape)

<>

(3900, 7168)
(1672, 7168)

📄

IMPLEMENTATION OF MODEL

 SPAM EMAIL DETECTION USING MACHINE LEARNING ☆

File Edit View Insert Runtime Tools Help [All changes saved](#)

Comment Share Settings

+ Code + Text

Reconnect Colab AI

IMPLEMENTATION OF MODEL

```
[ ] from sklearn.linear_model import LogisticRegression
lg_model=LogisticRegression()
lg_model.fit(x_train,y_train)
y_pred=lg_model.predict(x_test)
y_pred

array([0, 0, 0, ..., 0, 0, 0])
```

```
from sklearn.ensemble import RandomForestClassifier
rf_model=RandomForestClassifier()
rf_model.fit(x_train,y_train)
y_pred=rf_model.predict(x_test)
y_pred

array([0, 0, 0, ..., 0, 0, 0])
```

```
[ ] from sklearn.ensemble import AdaBoostClassifier
ab_model=AdaBoostClassifier()
ab_model.fit(x_train,y_train)
y_pred=ab_model.predict(x_test)
y_pred
```

0s completed at 19:59

+ Code + Text

Reconnect Colab AI

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
[ ] from sklearn.ensemble import RandomForestClassifier
rf_model=RandomForestClassifier()
rf_model.fit(x_train,y_train)
y_pred=rf_model.predict(x_test)
y_pred
```

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
[ ] from sklearn.ensemble import AdaBoostClassifier
ab_model=AdaBoostClassifier()
ab_model.fit(x_train,y_train)
y_pred=ab_model.predict(x_test)
y_pred
```

```
array([0, 0, 0, ..., 0, 0, 0])
```

```
from sklearn.neighbors import KNeighborsClassifier
knn_model=KNeighborsClassifier()
knn_model.fit(x_train,y_train)
y_pred=knn_model.predict(x_test)
y_pred
```

```
array([0, 0, 0, ..., 0, 0, 0])
```