

## SMS SPAM Classification

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

### Import Libraries

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

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from keras.models import Model
from keras.layers import LSTM, Activation, Dense, Dropout, Input,
Embedding
from keras.optimizers import RMSprop
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.utils import to_categorical
from keras.callbacks import EarlyStopping
import tensorflow
%matplotlib inline
```

```
!pip install tensorflow
```

```
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: tensorflow in
/usr/local/lib/python3.7/dist-packages (2.9.2)
Requirement already satisfied: protobuf<3.20,>=3.9.2 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.17.3)
Requirement already satisfied: termcolor>=1.1.0 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.0.1)
Requirement already satisfied: gast<=0.4.0,>=0.2.1 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.4.0)
Requirement already satisfied: astunparse>=1.6.0 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.6.3)
Requirement already satisfied: tensorflow-estimator<2.10.0,>=2.9.0rc0
in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.0)
Requirement already satisfied: packaging in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (21.3)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.50.0)
Requirement already satisfied: opt-einsum>=2.3.2 in
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.3.0)
```

Requirement already satisfied: setuptools in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (57.4.0)  
Requirement already satisfied: absl-py>=1.0.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.3.0)  
Requirement already satisfied: numpy>=1.20 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.21.6)  
Requirement already satisfied: google-pasta>=0.1.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.2.0)  
Requirement already satisfied: keras<2.10.0,>=2.9.0rc0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.0)  
Requirement already satisfied: h5py>=2.9.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.1.0)  
Requirement already satisfied: wrapt>=1.11.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.14.1)  
Requirement already satisfied: tensorboard<2.10,>=2.9 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.1)  
Requirement already satisfied: flatbuffers<2,>=1.12 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.12)  
Requirement already satisfied: typing-extensions>=3.6.6 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (4.1.1)  
Requirement already satisfied: six>=1.12.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.15.0)  
Requirement already satisfied: libclang>=13.0.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (14.0.6)  
Requirement already satisfied: keras-preprocessing>=1.1.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.1.2)  
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.27.0)  
Requirement already satisfied: wheel<1.0,>=0.23.0 in  
/usr/local/lib/python3.7/dist-packages (from astunparse>=1.6.0->tensorflow) (0.37.1)  
Requirement already satisfied: cached-property in  
/usr/local/lib/python3.7/dist-packages (from h5py>=2.9.0->tensorflow) (1.5.2)  
Requirement already satisfied: werkzeug>=1.0.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9->tensorflow) (1.0.1)  
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9->tensorflow) (1.8.1)  
Requirement already satisfied: requests<3,>=2.21.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9->tensorflow) (2.23.0)  
Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0 in /usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9->tensorflow) (0.6.1)  
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9->tensorflow) (0.4.6)  
Requirement already satisfied: google-auth<3,>=1.6.3 in

```

/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-
>tensorflow) (1.35.0)
Requirement already satisfied: markdown>=2.6.8 in
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-
>tensorflow) (3.4.1)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-
>tensorboard<2.10,>=2.9->tensorflow) (0.2.8)
Requirement already satisfied: rsa<5,>=3.1.4 in
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-
>tensorboard<2.10,>=2.9->tensorflow) (4.9)
Requirement already satisfied: cachetools<5.0,>=2.0.0 in
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-
>tensorboard<2.10,>=2.9->tensorflow) (4.2.4)
Requirement already satisfied: requests-oauthlib>=0.7.0 in
/usr/local/lib/python3.7/dist-packages (from google-auth-
oauthlib<0.5,>=0.4.1->tensorboard<2.10,>=2.9->tensorflow) (1.3.1)
Requirement already satisfied: importlib-metadata>=4.4 in
/usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8-
>tensorboard<2.10,>=2.9->tensorflow) (4.13.0)
Requirement already satisfied: zipp>=0.5 in
/usr/local/lib/python3.7/dist-packages (from importlib-metadata>=4.4-
>markdown>=2.6.8->tensorboard<2.10,>=2.9->tensorflow) (3.10.0)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
/usr/local/lib/python3.7/dist-packages (from pyasn1-modules>=0.2.1-
>google-auth<3,>=1.6.3->tensorboard<2.10,>=2.9->tensorflow) (0.4.8)
Requirement already satisfied: chardet<4,>=3.0.2 in
/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (3.0.4)
Requirement already satisfied: idna<3,>=2.5 in
/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (2022.9.24)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1
in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (1.24.3)
Requirement already satisfied: oauthlib>=3.0.0 in
/usr/local/lib/python3.7/dist-packages (from requests-oauthlib>=0.7.0-
>google-auth-oauthlib<0.5,>=0.4.1->tensorboard<2.10,>=2.9->tensorflow)
(3.2.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from packaging->tensorflow)
(3.0.9)

```

```

from tensorflow.keras.preprocessing.sequence import pad_sequences

```

## Read the Dataset

```
df= pd.read_csv('/content/spam.csv',delimiter=',',encoding='latin-1')
df.head()
```

```
      v1                                                    v2 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
```

```
      Unnamed: 3 Unnamed: 4
0      NaN      NaN
1      NaN      NaN
2      NaN      NaN
3      NaN      NaN
4      NaN      NaN
```

## Pre-processing the Dataset

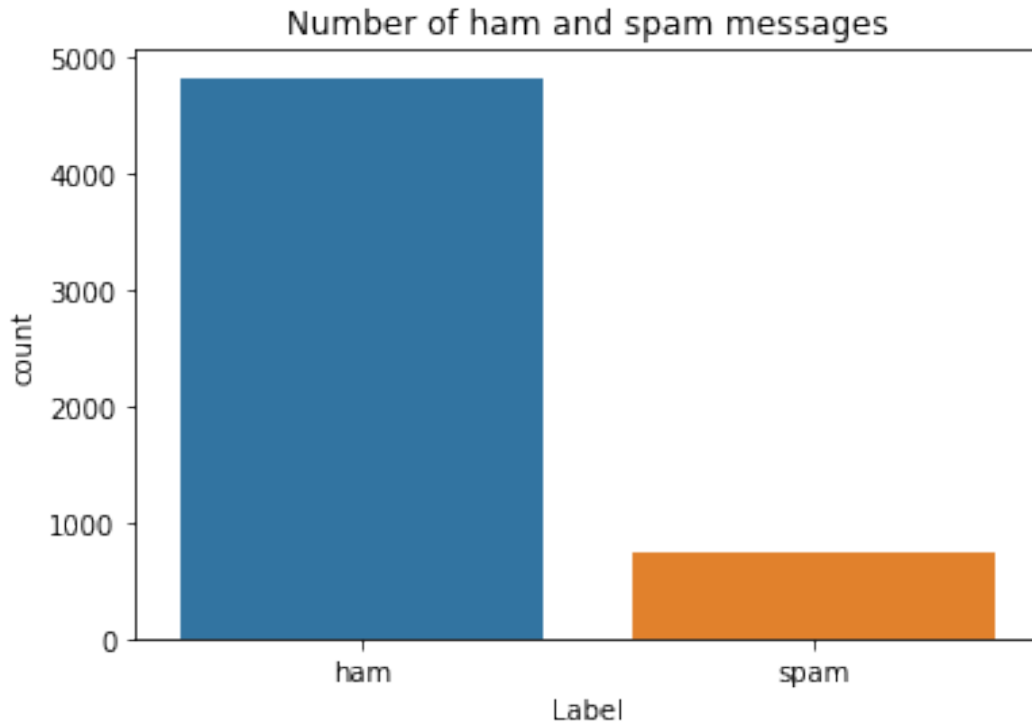
```
df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'],axis= 1,inplace=
True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0    v1      5572 non-null     object
1    v2      5572 non-null     object
dtypes: object(2)
memory usage: 87.2+ KB
```

```
sns.countplot(df.v1)
plt.xlabel('Label')
plt.title('Number of ham and spam messages')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
FutureWarning
```

```
Text(0.5, 1.0, 'Number of ham and spam messages')
```



```
x= df.v2
y= df.v1
le= LabelEncoder()
y= le.fit_transform(y)
y= y.reshape(-1,1)

x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.15)

max_words = 1000
max_len = 150
tok = Tokenizer(num_words=max_words)
tok.fit_on_texts(x_train)
sequences = tok.texts_to_sequences(x_train)
```

### Create Model & Add Layers

```
def RNN():
    inputs = Input(name='inputs',shape=[max_len])
    layer = Embedding(max_words,50,input_length=max_len)(inputs)
    layer = LSTM(64)(layer)
    layer = Dense(256,name='FC1')(layer)
    layer = Activation('relu')(layer)
    layer = Dropout(0.5)(layer)
    layer = Dense(1,name='out_layer')(layer)
    layer = Activation('sigmoid')(layer)
```

```
model = Model(inputs=inputs,outputs=layer)
return model
```

### Compile the Model

```
model = RNN()
model.summary()
model.compile(loss='binary_crossentropy',optimizer=RMSprop(),metrics=[
'accuracy'])
```

Model: "model"

Layer (type)	Output Shape	Param #
inputs (InputLayer)	[(None, 150)]	0
embedding (Embedding)	(None, 150, 50)	50000
lstm (LSTM)	(None, 64)	29440
FC1 (Dense)	(None, 256)	16640
activation (Activation)	(None, 256)	0
dropout (Dropout)	(None, 256)	0
out_layer (Dense)	(None, 1)	257
activation_1 (Activation)	(None, 1)	0
Total params: 96,337		
Trainable params: 96,337		
Non-trainable params: 0		

```
df.columns
```

```
Index(['v1', 'v2'], dtype='object')
```

```
data=df.rename(
    {
        "v1": "Category",
        "v2": "Message"
    },
    axis=1
)
```

```
df.info()
```

```

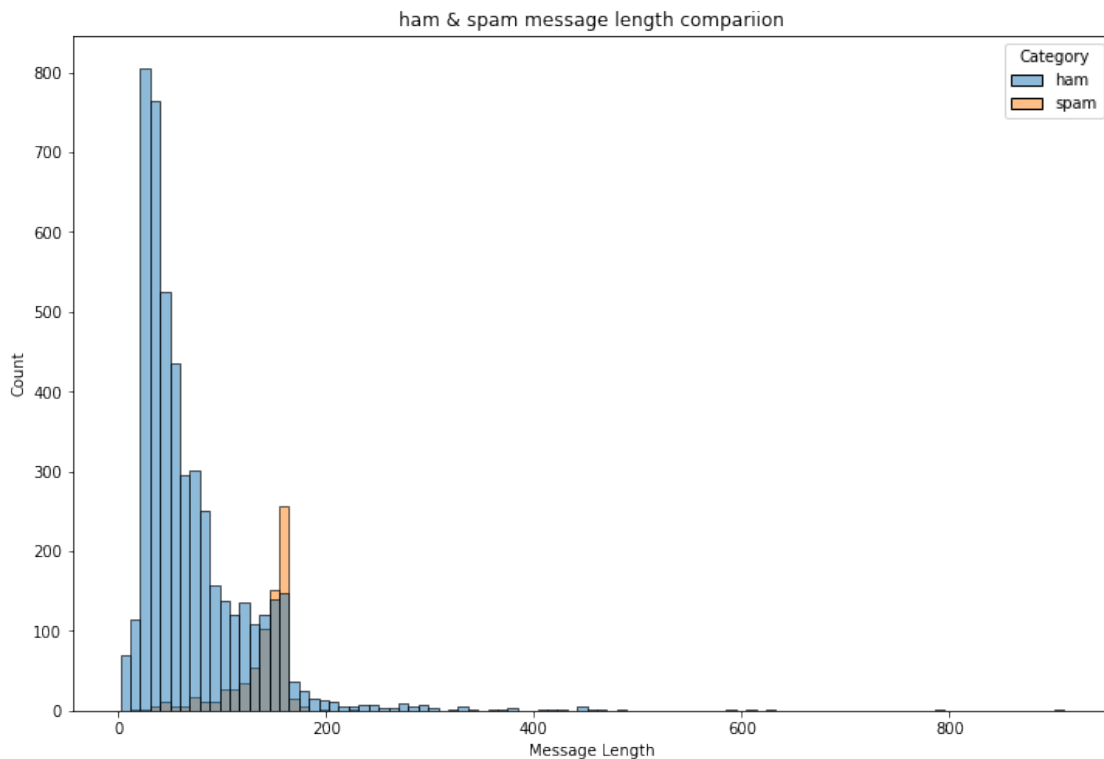
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5572 entries, 0 to 5571
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  -
 0    v1      5572 non-null    object
 1    v2      5572 non-null    object
dtypes: object(2)
memory usage: 87.2+ KB

data["Message Length"]=data["Message"].apply(len)

fig=plt.figure(figsize=(12,8))
sns.histplot(
    x=data["Message Length"],
    hue=data["Category"]
)
plt.title("ham & spam message length compariion")
plt.show

<function matplotlib.pyplot.show(*args, **kw)>

```



```

ham_desc=data[data["Category"]=="ham"]["Message Length"].describe()
spam_desc=data[data["Category"]=="spam"]["Message Length"].describe()

print("Ham Message Length Description:\n",ham_desc)
print("*****")
print("spam Message Length Description:\n",spam_desc)

```

```

Ham Message Length Description:
  count    4825.000000
  mean      71.023627
  std       58.016023
  min       2.000000
  25%      33.000000
  50%      52.000000
  75%      92.000000
  max      910.000000
Name: Message Length, dtype: float64
*****

```

```

spam Message Length Description:
  count    747.000000
  mean     138.866131
  std      29.183082
  min      13.000000
  25%     132.500000
  50%     149.000000
  75%     157.000000
  max     224.000000
Name: Message Length, dtype: float64

```

```
data.describe(include="all")
```

	Category	Message	Message Length
count	5572	5572	5572.000000
unique	2	5169	NaN
top	ham	Sorry, I'll call later	NaN
freq	4825	30	NaN
mean	NaN	NaN	80.118808
std	NaN	NaN	59.690841
min	NaN	NaN	2.000000
25%	NaN	NaN	36.000000
50%	NaN	NaN	61.000000
75%	NaN	NaN	121.000000
max	NaN	NaN	910.000000

```
data["Category"].value_counts()
```

```

ham      4825
spam     747
Name: Category, dtype: int64

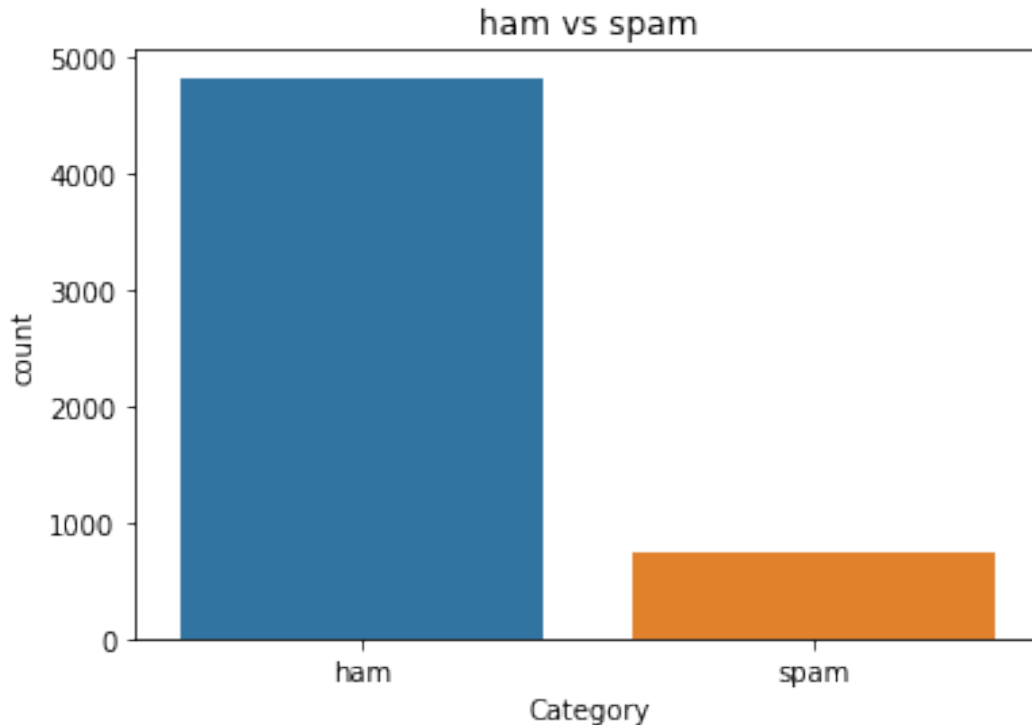
```

```

sns.countplot(
    data=data,
    x="Category"
)
plt.title("ham vs spam")
plt.show()

```





```
ham_count=data["Category"].value_counts()[0]
spam_count=data["Category"].value_counts()[1]
total_count=data.shape[0]
```

```
print("Ham Contains:{:2f}% of total
data.".format(ham_count/total_count*100))
print("Spam Contains:{:2f}% of total
data.".format(spam_count/total_count*100))
```

Ham Contains:86.593683% of total data.  
Spam Contains:13.406317% of total data.

```
minority_len=len(data[data["Category"]=="spam"])
majority_len=len(data[data["Category"]=="ham"])
minority_indices=data[data["Category"]=="spam"].index
majority_indices=data[data["Category"]=="ham"].index
random_majority_indices=np.random.choice(
    majority_indices,
    size=minority_len,
    replace=False
)
undersampled_indices=np.concatenate([minority_indices,random_majority_
indices])
df=data.loc[undersampled_indices]
df=df.sample(frac=1)

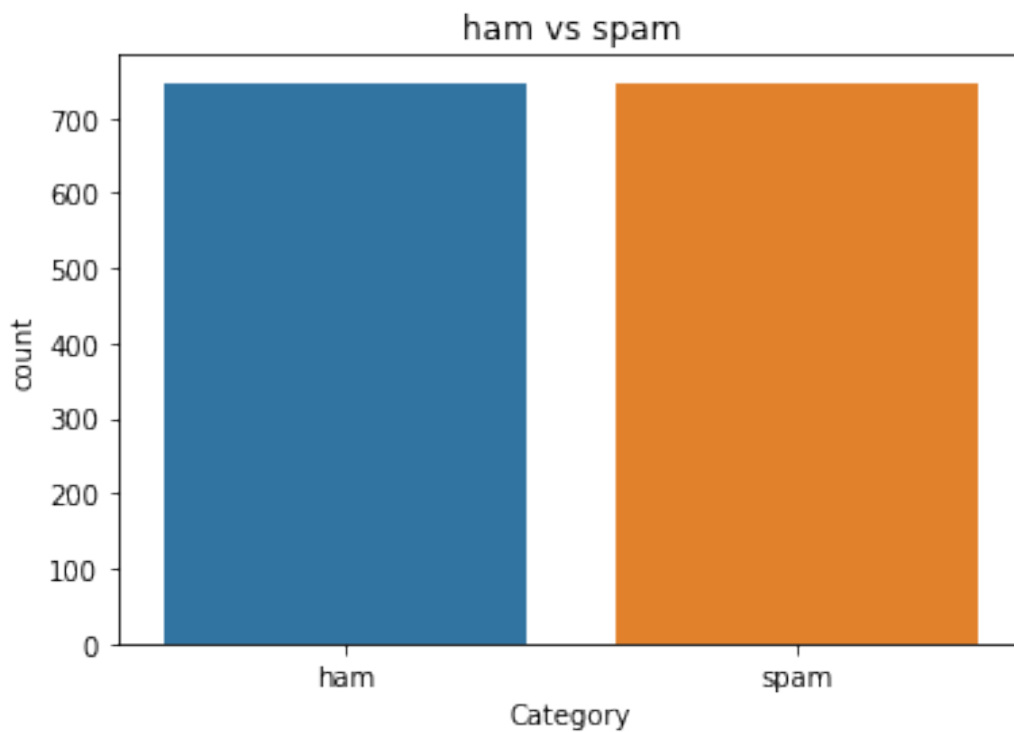
df=df.reset_index()
df=df.drop(
```

```

        columns=["index"],
    )
df.shape
(1494, 3)
df["Category"].value_counts()
ham      747
spam     747
Name: Category, dtype: int64

sns.countplot(
    data=df,
    x="Category"
)
plt.title("ham vs spam")
plt.show()

```



```

df.head()

```

	Category	Message	Message
0	ham	You should get more chicken broth if you want ...	
90			
1	ham	0. Well uv causes mutations. Sunscreen is like...	
66			
2	ham	Only saturday and sunday holiday so its very d...	

```

56
3      spam  Today's Offer! Claim ur å£150 worth of discoun...
160
4      ham   Er, hello, things didn't quite go to plan ...
114

df["label"]=df["Category"].map(
    {
        "ham":0,
        "spam":1
    }
)

df.head()

   Category      Message  Message
Length \
0      ham  You should get more chicken broth if you want ...
90
1      ham  0. Well uv causes mutations. Sunscreen is like...
66
2      ham  Only saturday and sunday holiday so its very d...
56
3      spam  Today's Offer! Claim ur å£150 worth of discoun...
160
4      ham  Er, hello, things didn't quite go to plan ...
114

   label
0      0
1      0
2      0
3      1
4      0

import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer

stemmer=PorterStemmer

corpus=[]
for message in df["Message"]:
    message=re.sub("[^a-zA-Z]", "",message)
    message=message.lower()
    message=message.split()

!pip install tensorflow

```

Looking in indexes: <https://pypi.org/simple>, <https://us-python.pkg.dev/colab-wheels/public/simple/>  
Requirement already satisfied: tensorflow in  
/usr/local/lib/python3.7/dist-packages (2.9.2)  
Requirement already satisfied: flatbuffers<2,>=1.12 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.12)  
Requirement already satisfied: keras-preprocessing>=1.1.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.1.2)  
Requirement already satisfied: protobuf<3.20,>=3.9.2 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.17.3)  
Requirement already satisfied: h5py>=2.9.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.1.0)  
Requirement already satisfied: google-pasta>=0.1.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.2.0)  
Requirement already satisfied: keras<2.10.0,>=2.9.0rc0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.0)  
Requirement already satisfied: opt-einsum>=2.3.2 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (3.3.0)  
Requirement already satisfied: wrapt>=1.11.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.14.1)  
Requirement already satisfied: six>=1.12.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.15.0)  
Requirement already satisfied: gast<=0.4.0,>=0.2.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.4.0)  
Requirement already satisfied: typing-extensions>=3.6.6 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (4.1.1)  
Requirement already satisfied: absl-py>=1.0.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.3.0)  
Requirement already satisfied: numpy>=1.20 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.21.6)  
Requirement already satisfied: libclang>=13.0.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (14.0.6)  
Requirement already satisfied: tensorboard<2.10,>=2.9 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.1)  
Requirement already satisfied: setuptools in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (57.4.0)  
Requirement already satisfied: grpcio<2.0,>=1.24.3 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.50.0)  
Requirement already satisfied: termcolor>=1.1.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (2.0.1)  
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (0.27.0)  
Requirement already satisfied: packaging in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (21.3)  
Requirement already satisfied: astunparse>=1.6.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorflow) (1.6.3)  
Requirement already satisfied: tensorflow-estimator<2.10.0,>=2.9.0rc0  
in /usr/local/lib/python3.7/dist-packages (from tensorflow) (2.9.0)  
Requirement already satisfied: wheel<1.0,>=0.23.0 in  
/usr/local/lib/python3.7/dist-packages (from astunparse>=1.6.0-

>tensorflow) (0.37.1)  
Requirement already satisfied: cached-property in  
/usr/local/lib/python3.7/dist-packages (from h5py>=2.9.0->tensorflow)  
(1.5.2)  
Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (0.4.6)  
Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (1.8.1)  
Requirement already satisfied: werkzeug>=1.0.1 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (1.0.1)  
Requirement already satisfied: tensorboard-data-server<0.7.0,>=0.6.0  
in /usr/local/lib/python3.7/dist-packages (from  
tensorboard<2.10,>=2.9->tensorflow) (0.6.1)  
Requirement already satisfied: requests<3,>=2.21.0 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (2.23.0)  
Requirement already satisfied: google-auth<3,>=1.6.3 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (1.35.0)  
Requirement already satisfied: markdown>=2.6.8 in  
/usr/local/lib/python3.7/dist-packages (from tensorboard<2.10,>=2.9-  
>tensorflow) (3.4.1)  
Requirement already satisfied: rsa<5,>=3.1.4 in  
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-  
>tensorboard<2.10,>=2.9->tensorflow) (4.9)  
Requirement already satisfied: pyasn1-modules>=0.2.1 in  
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-  
>tensorboard<2.10,>=2.9->tensorflow) (0.2.8)  
Requirement already satisfied: cachetools<5.0,>=2.0.0 in  
/usr/local/lib/python3.7/dist-packages (from google-auth<3,>=1.6.3-  
>tensorboard<2.10,>=2.9->tensorflow) (4.2.4)  
Requirement already satisfied: requests-oauthlib>=0.7.0 in  
/usr/local/lib/python3.7/dist-packages (from google-auth-  
oauthlib<0.5,>=0.4.1->tensorboard<2.10,>=2.9->tensorflow) (1.3.1)  
Requirement already satisfied: importlib-metadata>=4.4 in  
/usr/local/lib/python3.7/dist-packages (from markdown>=2.6.8-  
>tensorboard<2.10,>=2.9->tensorflow) (4.13.0)  
Requirement already satisfied: zipp>=0.5 in  
/usr/local/lib/python3.7/dist-packages (from importlib-metadata>=4.4-  
>markdown>=2.6.8->tensorboard<2.10,>=2.9->tensorflow) (3.10.0)  
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in  
/usr/local/lib/python3.7/dist-packages (from pyasn1-modules>=0.2.1-  
>google-auth<3,>=1.6.3->tensorboard<2.10,>=2.9->tensorflow) (0.4.8)  
Requirement already satisfied: chardet<4,>=3.0.2 in  
/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-  
>tensorboard<2.10,>=2.9->tensorflow) (3.0.4)  
Requirement already satisfied: idna<3,>=2.5 in

```

/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (2.10)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (2022.9.24)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1
in /usr/local/lib/python3.7/dist-packages (from requests<3,>=2.21.0-
>tensorboard<2.10,>=2.9->tensorflow) (1.24.3)
Requirement already satisfied: oauthlib>=3.0.0 in
/usr/local/lib/python3.7/dist-packages (from requests-oauthlib>=0.7.0-
>google-auth-oauthlib<0.5,>=0.4.1->tensorboard<2.10,>=2.9->tensorflow)
(3.2.2)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in
/usr/local/lib/python3.7/dist-packages (from packaging->tensorflow)
(3.0.9)

```

```

from tensorflow.keras.preprocessing.text import one_hot
vocab_size=10000

```

```

oneHot_doc=[one_hot(words,n=vocab_size)
for words in corpus
]

```

```

df["Message Length"].describe()

```

```

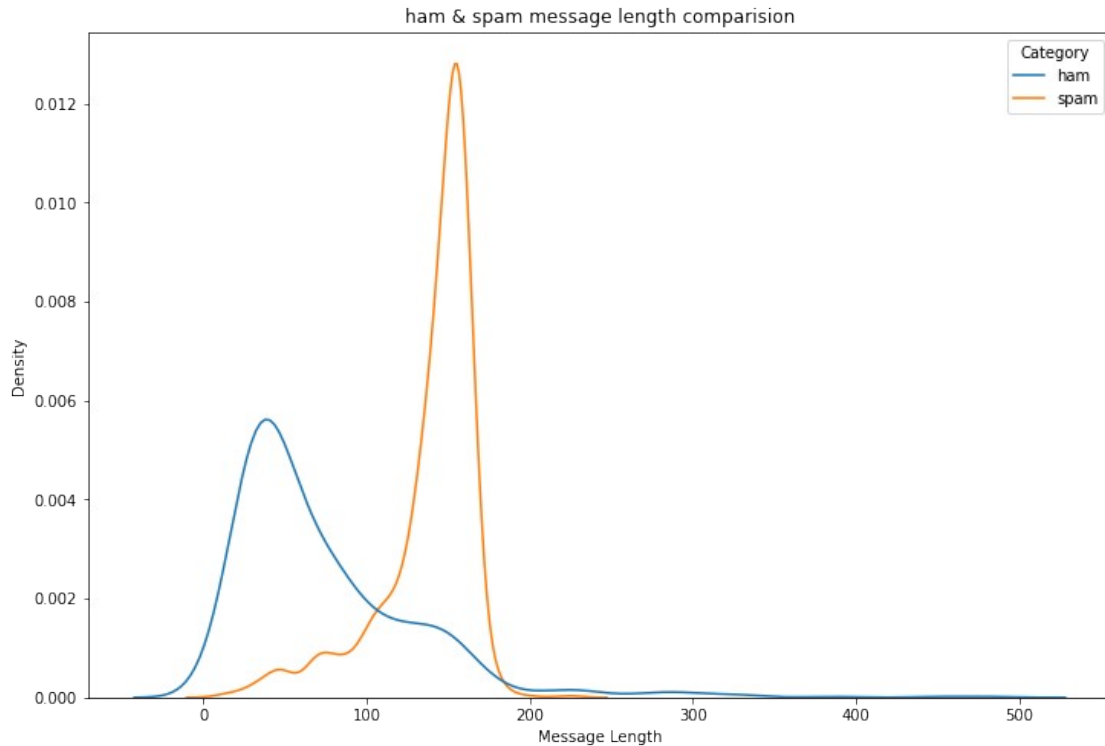
count    1494.000000
mean      106.605087
std       56.144244
min        4.000000
25%       52.000000
50%      120.000000
75%      153.000000
max      482.000000
Name: Message Length, dtype: float64

```

```

fig=plt.figure(figsize=(12,8))
sns.kdeplot(
    x=df["Message Length"],
    hue=df["Category"]
)
plt.title("ham & spam message length comparision")
plt.show()

```



```
from tensorflow.keras.preprocessing.sequence import pad_sequences
sentence_len=100
embedded_doc=pad_sequences(
    oneHot_doc,
    maxlen=sentence_len,
    padding="pre"
)
```

```
extract_features=pd.DataFrame(
    data=embedded_doc
)
target=df["label"]
```

```
df_final=pd.concat([extract_features,target],axis=1)
```

```
df_final.head()
```

```

      0   1   2   3   4   5   6   7   8   9   ...  91  92  93  94  95  96
97  \
0 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN NaN NaN NaN
NaN
1 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN NaN NaN NaN
NaN
2 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN NaN NaN NaN
NaN
3 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN NaN NaN NaN
NaN
4 NaN NaN NaN NaN NaN NaN NaN NaN NaN NaN ... NaN NaN NaN NaN NaN NaN
```

NaN

```
   98  99  label
0 NaN NaN      0
1 NaN NaN      0
2 NaN NaN      0
3 NaN NaN      1
4 NaN NaN      0
```

[5 rows x 101 columns]

```
x=df_final.drop("label",axis=1)
```

```
y=df_final["label"]
```

```
from sklearn.model_selection import train_test_split
```

```
x_trainval,x_test,y_trainval,y_test=train_test_split(
    x,
    y,
    random_state=37,
    test_size=0.2
)
```

```
x_trainval,x_val,y_trainval,y_test=train_test_split(
    x_trainval,
    y_trainval,
    random_state=37,
    test_size=0.2
)
```

```
model = RNN()
```

```
model.summary()
```

```
model.compile(loss='binary_crossentropy',optimizer=RMSprop(),metrics=[
    'accuracy'])
```

Model: "model\_1"

Layer (type)	Output Shape	Param #
inputs (InputLayer)	[(None, 150)]	0
embedding_1 (Embedding)	(None, 150, 50)	50000
lstm_1 (LSTM)	(None, 64)	29440
FC1 (Dense)	(None, 256)	16640
activation_2 (Activation)	(None, 256)	0
dropout_1 (Dropout)	(None, 256)	0



out_layer (Dense)	(None, 1)	257
activation_3 (Activation)	(None, 1)	0

=====  
Total params: 96,337  
Trainable params: 96,337  
Non-trainable params: 0  
=====