

TEAM ID	PNT2022TMID14295
ROLL NUMBER	771319EC027
DATE	08.11.2022

1. Import required library

```
import pandas as pd
import numpy as np
import nltk
import re

nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
```

2. Read dataset and do pre-processing

```
df=pd.read_csv('/content/spam.csv',encoding="ISO-8859-1")
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 2ndtime we have tried 2 contactu...	NaN	NaN	NaN

```
port=PorterStemmer()
```

```
cv = CountVectorizer()
```

```
dataset=[]
```

```
for i in range(0,len(df)):
```

```
    review = df['v2'][i] # Reading data
```

```
    review = re.sub('[^a-zA-Z]', ' ', review) # Removing special character
```

```
    review = review.lower() # Convert capital letters into small letters
```

```
    review = review.split() # Split the input
```

```
    review = [port.stem(w) for w in review if w not in set(stopwords.words('english'))] # Stemming & Stopwords
```

```
    review = ' '.join(review) # join words
```

```
    dataset.append(review)
```

```
dataset
```

```
['go jurong point crazi avail bugi n great world la e buffet cine got amor wat',
'ok lar joke wif u oni',
'free entri wkli comp win fa cup final tkt st may text fa receiv entri question std txt rate c appli',
'u dun say earli hor u c already say',
'nah think goe usf live around though',
'freemsg hey darl week word back like fun still tb ok xxx std chg send rcv',
'even brother like speak treat like aid patent',
'per request mell mell oru minnaminungint nurungu vettam set callertun caller press copi friend callertun',
'winner valu network custom select receivea prize reward claim call claim code kl valid hour',
'mobil month u r entitl updat latest colour mobil camera free call mobil updat co free',
```

'gonna home soon want talk stuff anymor tonight k cri enough today',
'six chanc win cash pound txt csh send cost p day day tsandc appli repli hl info',
'urgent week free membership prize jackpot txt word claim c www dbuk net lccltd pobox ldnw rw',
'search right word thank breather promis wont take help grant fulfil promis wonder bless time',
'date sunday',
'xxxmobilemovieclub use credit click wap link next txt messag click http wap xxxmobilemovieclub com n qjkgighjjgcb1',
'oh k watch',
'eh u rememb spell name ye v naughti make v wet',
'fine way u feel way gota b',
'england v macedonia dont miss goal team news txt ur nation team eg england tri wale scotland txt poboxox w wq',
'serious spell name',
'go tri month ha ha joke',
'pay first lar da stock comin',
'aft finish lunch go str lor ard smth lor u finish ur lunch already',
'ffffffffffff alright way meet sooner',
'forc eat slice realli hungri tho suck mark get worri know sick turn pizza lol',
'lol alway convinc',
'catch bu fri egg make tea eat mom left dinner feel love',
'back amp pack car let know room',
'ahhh work vagu rememb feel like lol',
'wait still clear sure sarcast x want live us',
'yeah got v apologet n fallen actin like spoilt child got caught till go badli cheer',
'k tell anyth',
'fear faint housework quick cuppa',
'thank subscript rington uk mobil charg month pleas confirm repli ye repli charg',
'yup ok go home look time msg xuhui go learn nd may lesson',
'oop let know roommat done',
'see letter b car',
'anyth lor u decid',
'hello saturday go text see decid anyth tomo tri invit anyth',
'pl go ahead watt want sure great weekend abiola',
'forget tell want need crave love sweet arabian steed mmmmmm yummi',
'rodger burn msg tri call repli sm free nokia mobil free camcord pleas call deliveri tomorrow',
'see',
'great hope like man well endow lt gt inch',
'call messag miss call',
'get hep b immunis nigeria',
'fair enough anyth go',
'yeah hope tyler could mayb ask around bit',
'u know stubborn even want go hospit kept tell mark weak sucker hospit weak sucker',
'think first time saw class',

```
'gram usual run like lt gt half eighth smarter though get almost whole second gram lt gt',
'k fyi x ride earli tomorrow morn crash place tonight',
'wow never realiz embarass accomod thought like sinc best could alway seem happi cave sorri give sorri offer sorri room
embarass',
'sm ac sptv new jersey devil detroit red wing play ice hockey correct incorrect end repli end sptv',
'know mallika sherawat yesterday find lt url gt',
```

```
x = cv.fit_transform(dataset).toarray()
x
```

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

```
enc=LabelEncoder()
y= enc.fit_transform(df['v1'])
y
```

```
array([0, 0, 1, ..., 0, 0, 0])
```

▼ 3. Create Model

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

▼ 4. Add Layers (LSTM, Dense-(Hidden Layers), Output)

```
model = Sequential()  
model.add(Dense(1500, activation='relu'))  
model.add(Dense(3000, activation='relu'))  
model.add(Dense(1, activation='sigmoid'))
```

▼ 5. Compile the Model

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

▼ 6. Fit the Model

```
model.fit(x,y, epochs=10)
```

```
Epoch 1/10  
175/175 [=====] - 2s 6ms/step - loss: 0.1188 - accuracy: 0.9684  
Epoch 2/10  
175/175 [=====] - 1s 6ms/step - loss: 0.0113 - accuracy: 0.9975  
Epoch 3/10  
175/175 [=====] - 1s 6ms/step - loss: 0.0018 - accuracy: 0.9995  
Epoch 4/10  
175/175 [=====] - 1s 6ms/step - loss: 4.5578e-04 - accuracy: 0.9998  
Epoch 5/10  
175/175 [=====] - 1s 6ms/step - loss: 1.3397e-04 - accuracy: 1.0000  
Epoch 6/10  
175/175 [=====] - 1s 6ms/step - loss: 6.3747e-05 - accuracy: 1.0000  
Epoch 7/10  
175/175 [=====] - 1s 6ms/step - loss: 3.7938e-05 - accuracy: 1.0000  
Epoch 8/10  
175/175 [=====] - 1s 6ms/step - loss: 2.4956e-05 - accuracy: 1.0000  
Epoch 9/10  
175/175 [=====] - 1s 6ms/step - loss: 1.7525e-05 - accuracy: 1.0000  
Epoch 10/10
```

```
175/175 [=====] - 1s 6ms/step - loss: 1.2765e-05 - accuracy: 1.0000
```

```
<keras.callbacks.History at 0x7f023a41cf50>
```

▼ 7. Save The Model

```
model.save('nlp.h5')
```

▼ 8. Test the Model

```
# Test 1
```

```
text = 'Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...'
print(text + "\n")
text = re.sub('[^a-zA-Z]', ' ', text)
text = text.lower()
text = text.split()
text = [port.stem(w) for w in text if w not in set(stopwords.words('english'))]
text = ' '.join(text)
text = cv.transform([text]).toarray()
pred = model.predict(text)
print("\nResult : ", end="")
if pred>0.5:
    print('Spam')
else: print('Ham')
```

```
Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
```

```
1/1 [=====] - 0s 18ms/step
```

```
Result : Ham
```

```
# Test 1
```

```
text = 'England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4t
print(text + "\n")
text = re.sub('[^a-zA-Z]', ' ', text)
text = text.lower()
text = text.split()
text = [port.stem(w) for w in text if w not in set(stopwords.words('english'))]
text = ' '.join(text)
text = cv.transform([text]).toarray()
pred = model.predict(text)
print("\nResult : ",end="")
if pred>0.5:
    print('Spam')
else: print('Ham')
```

England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt

1/1 [=====] - 0s 14ms/step

Result : Spam
