

Files

🔍

📁

📄

🔒

..

sample_data

spam-NLP.h5

spam.csv

<>

☰

🖨

Disk

85.01 GB available

+ Code + Text

✓ RAM

Disk

Editing

↑ ↓ 🔗 💬 ✎ 📄 🗑 ⋮

IMPORT LIBRARIES

+ Code + Text

✓ 1s

```
[1] 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

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

LOAD DATASET

✓ 0s

```
[2] a = pd.read_csv('/content/spam.csv',encoding='ISO-8859-1')
a.head()
```

Files

..

sample_data

spam-NLP.h5

spam.csv

+ Code

+ Text

✓

RAM

Disk

Editing

[nltk_data]

Unzipping corpora/stopwords.zip.

LOAD DATASET

✓

0s

[2]

a = pd.read_csv('/content/spam.csv',encoding='ISO-8859-1')

a.head()

	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

✓

0s

[3]

a=a[['v1','v2']]

a.head()

	v1	v2
--	----	----

Disk

85.01 GB available

completed at 9:46 AM

Files

sample_data

spam-NLP.h5

spam.csv

+ Code

+ Text

RAM

Disk

Editing

✓ [2]

0s

4

ham

Nah I don't think he goes to usf, he lives aro...

NaN

NaN

NaN

✓ [3]

0s

a=a[['v1', 'v2']]

a.head()

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...

✓ [4]

0s

a.shape

(5572, 2)

Files

sample_data

spam-NLP.h5

spam.csv

Text processing (NLP)

[5] ps=PorterStemmer()
message=[]
for i in range(0,5572):
 msg=a['v2'][i]
 msg=re.sub('[^a-zA-Z]', ' ',msg)
 msg=msg.lower()
 msg=msg.split(' ')
 msg = [ps.stem(word) for word in msg if word not in set(stopwords.words('english'))]
 msg=' '.join(msg)
 message.append(msg)

message[:6]

['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']

0s completed at 9:46 AM

RAM

Disk

Editing

85.01 GB available

Files



{x}

..

sample_data

spam-NLP.h5

spam.csv

<>

Disk  85.01 GB available

+ Code + Text

✓ RAM 
Disk 

Editing



✓ [6] from sklearn.feature_extraction.text import CountVectorizer

0s

```
cv = CountVectorizer()  
x = cv.fit_transform(message).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]])
```

✓ [7] #LABEL ENCODING

0s

```
from sklearn.preprocessing import LabelEncoder  
le = LabelEncoder()
```

```
a['v1']=le.fit_transform(a['v1'])  
y = a['v1'].values  
y
```

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

✓ 0s completed at 9:46 AM



Files

sample_data

spam-NLP.h5

spam.csv

MODEL BUILDING

[8] from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense

model = Sequential()
model.add(Dense(1550,activation='relu'))
model.add(Dense(3000,activation='relu'))
model.add(Dense(1,activation='sigmoid'))

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

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

Epoch 1/10
175/175 [=====] - 21s 113ms/step - loss: 0.1096 - accuracy: 0.9702
Epoch 2/10
175/175 [=====] - 19s 107ms/step - loss: 0.0083 - accuracy: 0.9971
Epoch 3/10
175/175 [=====] - 22s 125ms/step - loss: 0.0013 - accuracy: 0.9995
Epoch 4/10
175/175 [=====] - 18s 104ms/step - loss: 1.9033e-04 - accuracy: 1.0000
Epoch 5/10

RAM
Disk

85.01 GB available

0s completed at 9:46 AM

Files

sample_data

spam-NLP.h5

spam.csv

RAM

Disk

85.01 GB available

+ Code + Text

✓

3m

Epoch 3/10
[8] 175/175 [=====] - 22s 125ms/step - loss: 0.0013 - accuracy: 0.9995
Epoch 4/10
175/175 [=====] - 18s 104ms/step - loss: 1.9033e-04 - accuracy: 1.0000
Epoch 5/10
175/175 [=====] - 18s 104ms/step - loss: 7.3711e-05 - accuracy: 1.0000
Epoch 6/10
175/175 [=====] - 18s 104ms/step - loss: 4.3074e-05 - accuracy: 1.0000
Epoch 7/10
175/175 [=====] - 18s 105ms/step - loss: 2.7356e-05 - accuracy: 1.0000
Epoch 8/10
175/175 [=====] - 19s 111ms/step - loss: 1.7791e-05 - accuracy: 1.0000
Epoch 9/10
175/175 [=====] - 19s 110ms/step - loss: 1.1924e-05 - accuracy: 1.0000
Epoch 10/10
175/175 [=====] - 19s 106ms/step - loss: 8.3434e-06 - accuracy: 1.0000
<keras.callbacks.History at 0x7fa8341870d0>

SAVE THE MODEL

✓

0s

[9] model.save('spam-NLP.h5')

TEST THE MODEL

✓ 0s completed at 9:46 AM

Files

↑

↓

↺

↻

🔍

..

sample_data

spam-NLP.h5

spam.csv

+ Code + Text

✓ RAM

Disk

 Editing

TEST THE MODEL

✓ 0s

[10] msg='FREE MESSAGE Activate your 500 FREE Text Messages by replying to this message with the word FREE'

print('THE ORIGINAL MESSAGE IS: ',msg)

msg=re.sub('[^a-zA-Z]', ' ',msg)

msg=msg.lower()

msg=msg.split(' ')

msg = [ps.stem(word) for word in msg if word not in set(stopwords.words('english'))]

msg=' '.join(msg)

print('THE STEMMED MESSAGE IS: ',msg)

predict = model.predict(cv.transform([msg]))

if predict > 0.5:

pred='SPAM'

else: pred='NOT SPAM'

print('THE MESSAGE IS PREDICTED AS: ',pred)

THE ORIGINAL MESSAGE IS: FREE MESSAGE Activate your 500 FREE Text Messages by replying to this message with the word

THE STEMMED MESSAGE IS: free messag activ free text messag repli messag word free

1/1 [=====] - 0s 154ms/step

THE MESSAGE IS PREDICTED AS: SPAM

Disk 85.01 GB available

✓ 0s completed at 9:46 AM

Files

sample_data

spam-NLP.h5

spam.csv

1/1 [=====] - 0s 154ms/step

THE MESSAGE IS PREDICTED AS: SPAM

11

msg='Wishing you and your family Merry \X\" mas and HAPPY NEW Year in advance..'
print('THE ORIGINAL MESSAGE IS: ',msg)
msg=re.sub('[^a-zA-Z]', ' ',msg)
msg=msg.lower()
msg=msg.split(' ')
msg = [ps.stem(word) for word in msg if word not in set(stopwords.words('english'))]
msg= ' '.join(msg)
print('THE ORIGINAL MESSAGE IS: ',msg)

predict = model.predict(cv.transform([msg]))
if predict > 0.5:
pred='spam'
else: pred='NOT SPAM'
print('THE MESSAGE IS PREDICTED AS: ',pred)

THE ORIGINAL MESSAGE IS: Wishing you and your family Merry \X\" mas and HAPPY NEW Year in advance..
THE ORIGINAL MESSAGE IS: wish famili merri x ma happi new year advanc
1/1 [=====] - 0s 11ms/step
THE MESSAGE IS PREDICTED AS: NOT SPAM

RAM

Disk

85.01 GB available

+ Code

+ Text

RAM

Disk

Editing

0s completed at 9:46 AM