

Assignment - 4
Python Programming

Assignment Date	18-10-2022
Student Name	SODANAPALLI DEEPIKA
Student Roll Number	1 I 1519104144
Maximum Marks	2 Marks

Question 1:

Import the necessary libraries

Solution:

```
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, Dense, Dropout, Input, Embedding
from keras.optimizers import Adam
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.utils import to_categorical
from keras.callbacks import EarlyStopping
```

Question 2:

Download the Dataset

Solution:

Dataset Downloaded and uploaded to drive <https://www.kaggle.com/code/kredy10/simple-lstm-for-text-classification/data>

Question 3:

Read dataset and do pre-processing

Solution:

Read dataset

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

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until yurong 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 ie <vin FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say..	NaN	NaN	NaN
4	ham	Nah1 don'tthink hegsesto usf, he lives aro...	NaN	NaN	NaN

Pre-processing the Dataset

```
df. drop ( ['Unnamed: 2", 'unnamed: 3 ', unnamed: 4 ' ], ax1s =1, inp1ace=True)
df. 1nfo ( )
```

```
RangeIndex: 5572 entri es, 6 to 5571
Data columnns (total 2 columnns ) :
# Column non-Null Count Dtype
e •i s572 non-null object
1 v2 5572 non-null object
dtypes: object(2)
memory usage: 87.2+ KB
```

```
X df. v2
Y df. v1
1e • Labe1Encoder ( )
v • 1c. f1k_transfom(Y J
Y ¥.reshape(-1,1)
```

```
X_train,x_test,r_train,Y_test • tra1n_tcst_cptGt 'X,Y,test_s1ze•0.15 j
```

```
eex_1en - 1sa
tok • 7oken1zer (nuo_nords•eex_words `)
tok.fit_on_texts(X_train)
sequences • tok .text s_to_scquenCes ( x_traLn )
sequences_eatrl x • ped_sequencec ( sequences ,aoxLen•esx_1 en)
```

```
Inputs • InpuE {nane- " i nput s ', shape• [ cax_1cn ] j
layer • Enbcd1ng(mgx_words,50,1input_1en@h•wax_1en)(InputsI
1syer • LSTH{6A)( 1ayer)
layer • Ocnse (2S6, naee•FCI ) ( 1ayer)
1eyer - ActCvetton(*rr 1u') l 1eyer)
layer - Dropout (o.s \ (1eyer l
1ayer - Dence_1,nane•"sut_1ayer' )(1ayer j
layer • Ac t l va t1on( sig•old ) { layer)
atode1 - fJodel [ Input s•Input s , output s• 1ayer)
```

wode1: * ____1 1*

Layer (type)	&tpux Shape	Psra•s
1nput s (InputLayec)	[(None, 15e)]	0
eebedd1n\$ 1 (E iedd1ng)	(lone, 150, 59)	50000
1sM 1 (LSW!)	[None, 64]	29440
FCf (Oense)	(Mone, 2M)	1664B
activation_2 (Activation)	(None, 256)	0
drspout_l {Dcopou6}	[None, 256]	0
out_lmyer {Oense)	[None, 1]	2S7
act1vet1on_3 {Act tvst on)	(None, 1)	0

Tgte1 96,337
Trs Lnab1e pcrmis: 96, 337
□

```
[n [28]: model.fit(sequences_matrix,Y_train,batch_size=128,epochs=10,
              validation_split=0.2)

Epoch 1/10
30/30 [=====] - lbs 26A/st/ s£0p \ 055 - 6.3182 6CC uP6Cy <?.0788 - v6T_ Tc•ss : O. 15?† - \v0 J_6£.6uP dC y : 9. 9715 36 '
38 [-=====] - 7s 247•s/ step - Nos s( 6. B805 - eccur•cy) O.97bd - e1_Loss: 8.6742 - vaS_eccurecy: 8.9778
30/30 [=====) - 7 237•s/•i*$ loss: 0.$w0& - •<<••x<y. $,90$a *1_lo*x. e.08?g - •*i *cc«r <y. o.e$0i
Epoch 1/10
30/30 [=====] - 7s 245ms/step - loss: 0.0272 - accuracy: 0.9929 - val_loss: 0.0806 - val_accuracy: 0.9778
3e' 3g (=====) • 7s 242ss. step • I ass : B. az 28 - oc < ura-c y: O.e'9 47- 's I Los i: G.T8Ze • v! p•c cu r•c y* e. est
Epoch 6 'IN
36a 3-6 [=====] 7s ≥^^=≤ step Nos s. 6. B178 •c cur•cy 6i.995'S vB} loss : o.e*gs - v4} accurpc y - e. e ss
3Gy SO (-----] • 7t 241•s/ atep • loss i 6. B158 - •c curacy : 6. 995g • ma L_Lost : 0.0g69 • ve1_ac curac y : 6.99Gg
Epoch 6 'IN
30/30 [=====] - 7s 241ms/step - loss: 0.0162 - accuracy: 0.9958 - val_loss: 0.0901 - val_accuracy: 0.9768
36.' 3e (-----) ! - 7s *A6a*/ step loss < 8. 059 - secured y : 0. M68 ••s1_1oas . e. us4 - vaJ_eccur ac y : 6. 9769
Epoch 167 16
30/30 [=====] - 7s 247ms/step - loss: 0.0355 - accuracy: 0.9905 - val_loss: 0.1264 - val_accuracy: 0.9726
```

Savs Th« Uodef

```
code1.save( 'sus c1assil se<.h5' )
```

Processing the Test Data

```
test_sequences = tok.trxt.s_to_sequences|x_test>
test_sequences astr1x pad_sequences { t e st_sequences ,max len =a ax ten )
```

Testing the Modal

```
accr      t>>de1 . evat uatr i test_sequence s_aatr kx , Y_Eest )
```

27a' 27 [• • • • • - - - - - • • • • • • • • • •] - 1 s 2Ws'slep - lass: 6.66B6 - acc u'acy: e.982l

pctn% • Test set'n Lost [:6.37)*.n Accuracy: :e.*l '.fo•ost l acc r[e], ac cr |y|T)

```
Test set
cor>: e.ess
dcc u^acy : 0. 9B2
```