

Assignment Date	17 October 2022
Student Name	M.Vivek
Student Roll Number	962719106039
Maximum Marks	2Marks

1.Download the Dataset

Link: https://drive.google.com/file/d/1Sjqx5H5R86tRp2YZKzdd4_iEfjChZ3ob/view?usp=share_link

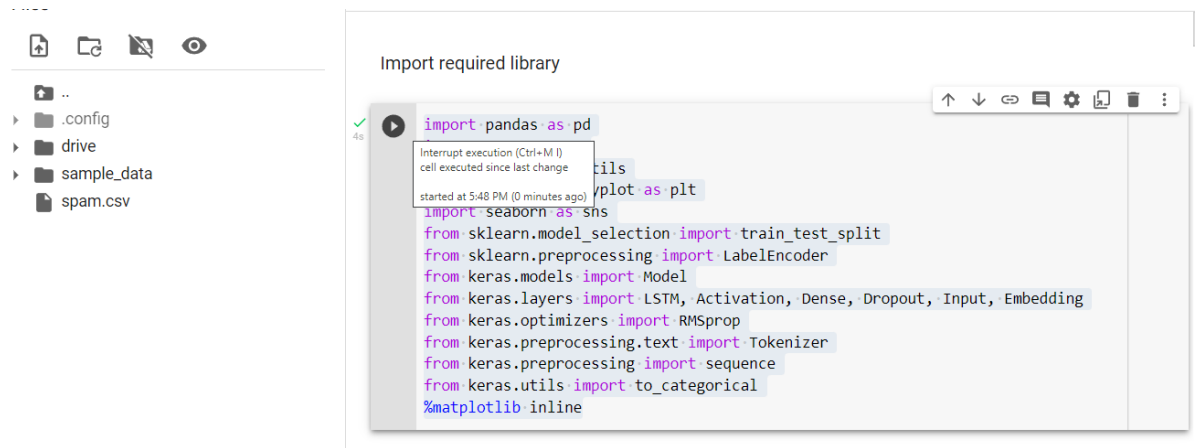
2.Import required library

Solution:

```
import pandas as pd

import numpy as np

from keras import utils
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
%matplotlib inline
```



3. Read dataset and do pre-processing

Read Data

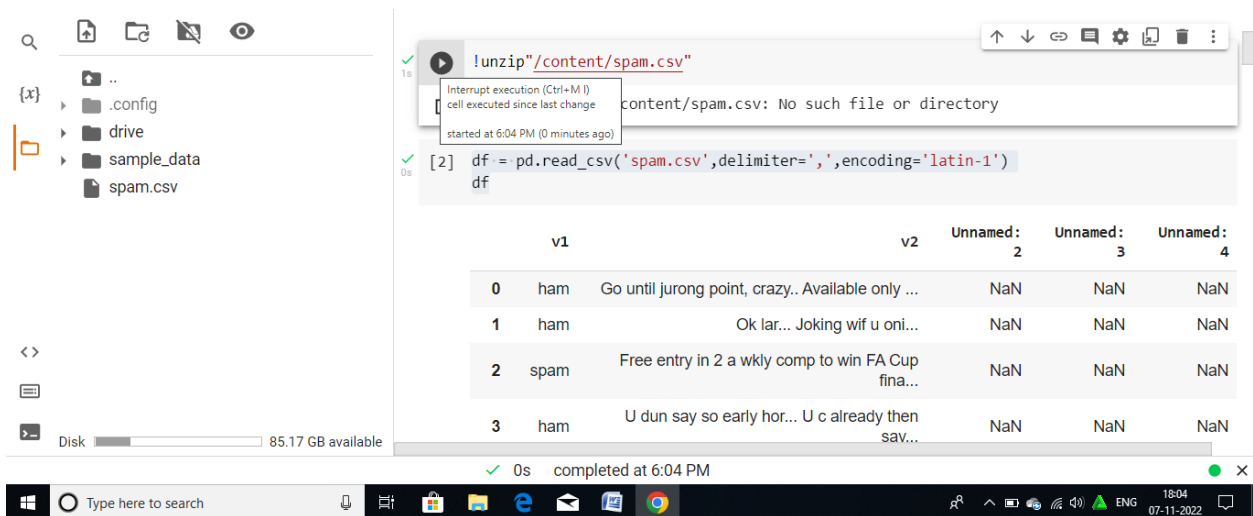
Solution:

```

!unzip"/content/spam.csv"
df = pd.read_csv('spam.csv',delimiter=',',encoding='latin-1')

```

df



Pre processing

Solution:

```

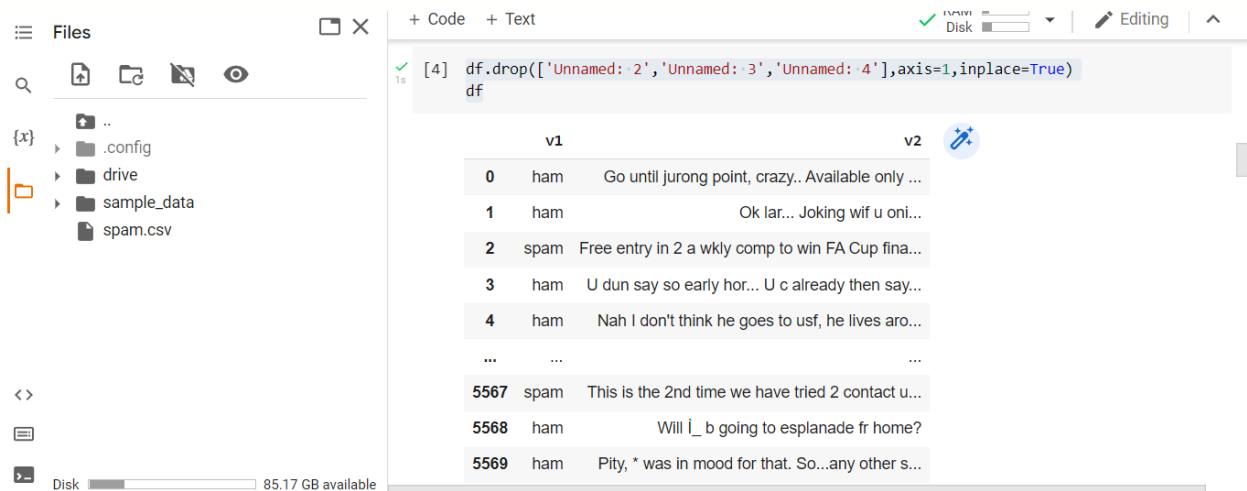
df.drop(['Unnamed: 2','Unnamed: 3','Unnamed: 4'],axis=1,inplace=True)

```

df

```
sns.countplot(df.v1,palette='Set3')

plt.xlabel('Label')
plt.title('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)
sequences_matrix = utils.pad_sequences(sequences,maxlen=max_len)
sequences_matrix.shape
sequences_matrix.ndim
sequences_matrix=np.reshape(sequences_matrix,(4736,150,1))
sequences_matrix.ndim
```



The screenshot shows a Jupyter Notebook interface. On the left is a file explorer with a search bar and icons for file operations. It shows a directory structure with files like `.config`, `drive`, `sample_data`, and `spam.csv`. The main area on the right contains a code cell with the following code:

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

Below the code cell is a preview of the DataFrame `df`. It has two columns: `v1` and `v2`. The data is as follows:

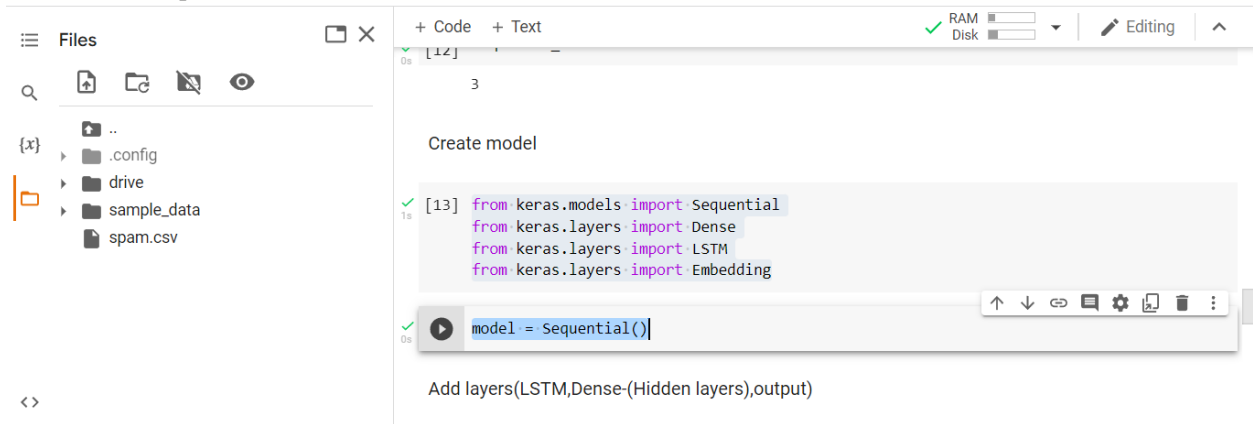
	v1	v2
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...
...
5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will I_b going to esplanade fr home?
5569	ham	Pity, * was in mood for that. So...any other s...

At the bottom of the interface, there is a status bar showing 'Disk' usage and '85.17 GB available'.


```

from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Embedding
model = Sequential()

```

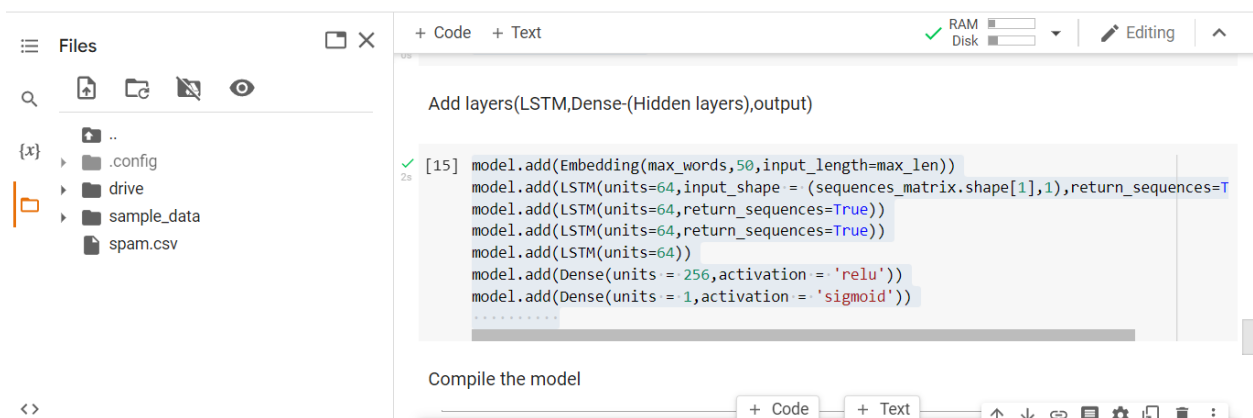


5.Add layers(LSTM,Dense-(Hidden layers),output)

```

model.add(Embedding(max_words,50,input_length=max_len))
model.add(LSTM(units=64,input_shape = (sequences_matrix.shape[1],1),return_sequences=True))
model.add(LSTM(units=64,return_sequences=True))
model.add(LSTM(units=64,return_sequences=True))
model.add(LSTM(units=64))
model.add(Dense(units = 256,activation = 'relu'))
model.add(Dense(units = 1,activation = 'sigmoid'))

```



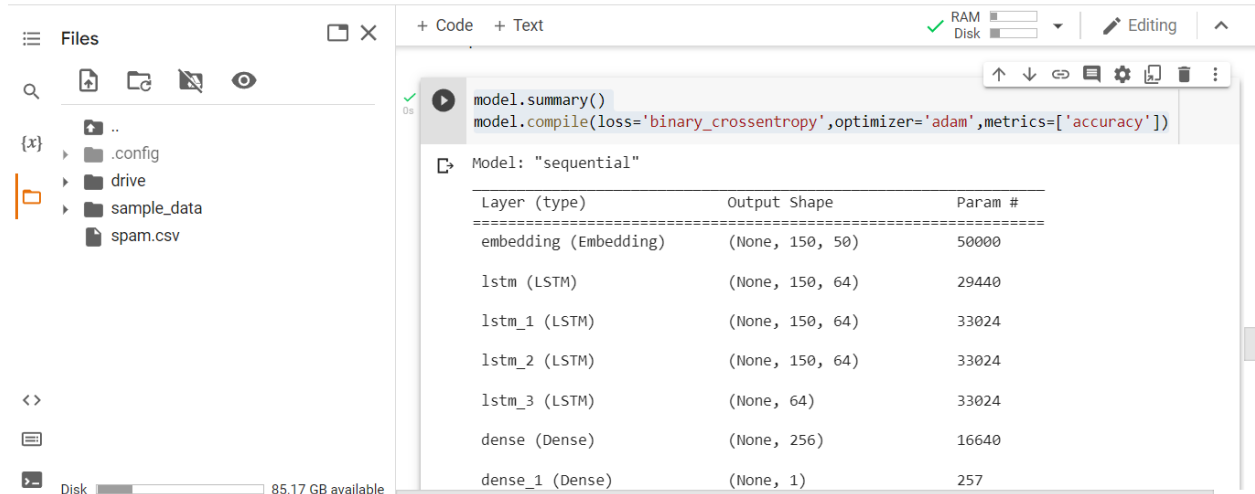
6.Compile the model

Solution:

```

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

```

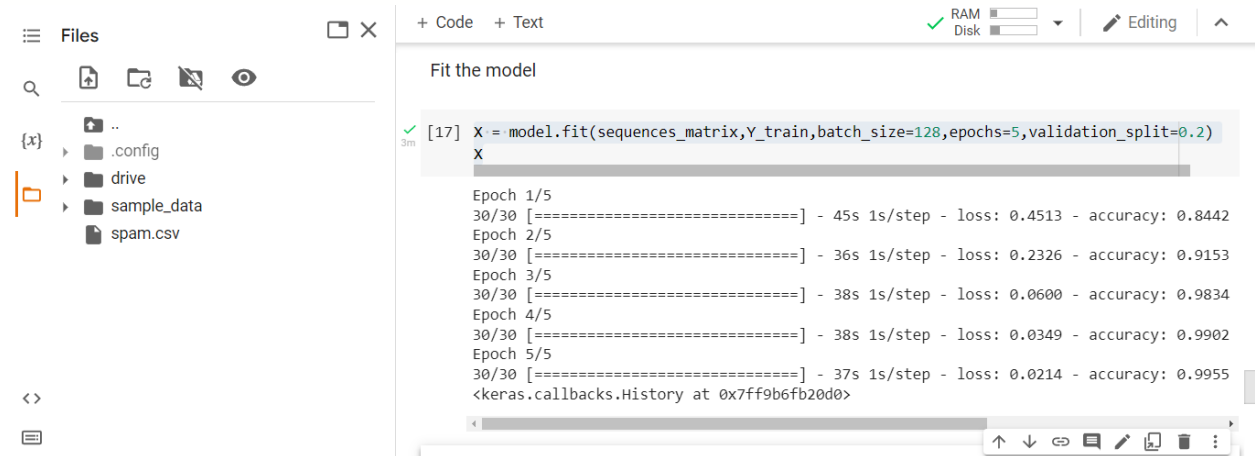


7. Fit the Model

Solution:

```
X = model.fit(sequences_matrix,Y_train,batch_size=128,epochs=5,validation_split=0.2)
```

X



8. Save the Model

Solution:

```
model.save
```