In [ ]:

```

```

Importing Model building libraries

In [ ]:

```
import pandas as pd
import numpy as np
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.models import load_model
```

Importing NLTK libraries

In [ ]:

```
import csv
import tensorflow as tf
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
STOPWORDS = set(stopwords.words('english'))
```

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

Reading dataset and preprocessing

In [ ]:

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

```
Mounted at /content/drive
```

In [ ]:

```
cd/content/drive/MyDrive/Colab Notebooks
```

```
/content/drive/MyDrive/Colab Notebooks
```

In [27]:

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

Out[27]:

	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

	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
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

In [28]:

```
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
```

In [29]:

```
df.groupby(['v1']).size()
```

Out[29]:

```
v1
ham     4825
spam     747
dtype: int64
```

In [30]:

```
#Label Encoding Required Column
X = df.v2
Y = df.v1
le = LabelEncoder()
Y = le.fit_transform(Y)
Y = Y.reshape(-1,1)
```

In [32]:

```
# Test and train data split
X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size=0.15)
```

In [33]:

```
# Tokenisation function
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 = sequence.pad_sequences(sequences,maxlen=max_len)
```

Create Model Add layers (LSTM ,Dense-(HiddenLayers),Ouput)

In [35]:

```
#LSTM model
inputs = Input(name='InputLayer',shape=[max_len])
layer = Embedding(max_words,50,input_length=max_len)(inputs)
layer = LSTM(64)(layer)
layer = Dense(256,name='FullyConnectedLayer1')(layer)
layer = Activation('relu')(layer)
layer = Dropout(0.5)(layer)
layer = Dense(1,name='OutputLayer')(layer)
layer = Activation('sigmoid')(layer)
```

In [36]:

```
model = Model(inputs=inputs,outputs=layer)
model.summary()
model.compile(loss='binary_crossentropy',optimizer=RMSprop(),metrics=['accuracy'])
```

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   
=================================================================
 InputLayer (InputLayer)     [(None, 150)]             0         
                                                                 
 embedding (Embedding)       (None, 150, 50)           50000     
                                                                 
 lstm (LSTM)                 (None, 64)                29440     
                                                                 
 FullyConnectedLayer1 (Dense  (None, 256)              16640     
 )                                                               
                                                                 
 activation (Activation)     (None, 256)               0         
                                                                 
 dropout (Dropout)           (None, 256)               0         
                                                                 
 OutputLayer (Dense)         (None, 1)                 257       
                                                                 
 activation_1 (Activation)   (None, 1)                 0         
                                                                 
=================================================================
Total params: 96,337
Trainable params: 96,337
Non-trainable params: 0
_________________________________________________________________
```

In [37]:

```
model.fit(sequences_matrix,Y_train,batch_size=128,epochs=25,validation_split=0.2)
```

```
Epoch 1/25
30/30 [==============================] - 11s 281ms/step - loss: 0.3416 - accuracy: 0.8741
- val_loss: 0.1699 - val_accuracy: 0.9378
Epoch 2/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0909 - accuracy: 0.9776
- val_loss: 0.0717 - val_accuracy: 0.9768
Epoch 3/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0470 - accuracy: 0.9871
- val_loss: 0.0740 - val_accuracy: 0.9757
Epoch 4/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0347 - accuracy: 0.9908
- val_loss: 0.0660 - val_accuracy: 0.9800
Epoch 5/25
30/30 [==============================] - 8s 261ms/step - loss: 0.0295 - accuracy: 0.9921
- val_loss: 0.0573 - val_accuracy: 0.9778
Epoch 6/25
30/30 [==============================] - 9s 297ms/step - loss: 0.0206 - accuracy: 0.9950
- val_loss: 0.0636 - val_accuracy: 0.9810
Epoch 7/25
30/30 [==============================] - 8s 262ms/step - loss: 0.0169 - accuracy: 0.9958
- val_loss: 0.0741 - val_accuracy: 0.9810
Epoch 8/25
30/30 [==============================] - 10s 319ms/step - loss: 0.0129 - accuracy: 0.9968
- val_loss: 0.0686 - val_accuracy: 0.9810
Epoch 9/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0079 - accuracy: 0.9982
- val_loss: 0.0852 - val_accuracy: 0.9800
Epoch 10/25
30/30 [==============================] - 8s 262ms/step - loss: 0.0074 - accuracy: 0.9984
- val_loss: 0.0823 - val_accuracy: 0.9789
Epoch 11/25
30/30 [==============================] - 8s 263ms/step - loss: 0.0054 - accuracy: 0.9989
- val_loss: 0.1050 - val_accuracy: 0.9800
Epoch 12/25
30/30 [==============================] - 8s 262ms/step - loss: 0.0042 - accuracy: 0.9989
```

```
- val_loss: 0.1205 - val_accuracy: 0.9778
Epoch 13/25
30/30 [==============================] - 10s 322ms/step - loss: 0.0033 - accuracy: 0.9992
- val_loss: 0.1480 - val_accuracy: 0.9800
Epoch 14/25
30/30 [==============================] - 10s 322ms/step - loss: 0.0044 - accuracy: 0.9987
- val_loss: 0.1489 - val_accuracy: 0.9778
Epoch 15/25
30/30 [==============================] - 9s 300ms/step - loss: 0.0025 - accuracy: 0.9995
- val_loss: 0.1560 - val_accuracy: 0.9768
Epoch 16/25
30/30 [==============================] - 9s 299ms/step - loss: 0.0028 - accuracy: 0.9989
- val_loss: 0.1710 - val_accuracy: 0.9778
Epoch 17/25
30/30 [==============================] - 10s 331ms/step - loss: 0.0022 - accuracy: 0.9997
- val_loss: 0.1627 - val_accuracy: 0.9768
Epoch 18/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0018 - accuracy: 0.9997
- val_loss: 0.1810 - val_accuracy: 0.9757
Epoch 19/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0026 - accuracy: 0.9995
- val_loss: 0.1547 - val_accuracy: 0.9768
Epoch 20/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0021 - accuracy: 0.9997
- val_loss: 0.1538 - val_accuracy: 0.9778
Epoch 21/25
30/30 [==============================] - 8s 262ms/step - loss: 0.0016 - accuracy: 0.9997
- val_loss: 0.1626 - val_accuracy: 0.9757
Epoch 22/25
30/30 [==============================] - 8s 261ms/step - loss: 0.0022 - accuracy: 0.9997
- val_loss: 0.1691 - val_accuracy: 0.9757
Epoch 23/25
30/30 [==============================] - 8s 260ms/step - loss: 0.0022 - accuracy: 0.9995
- val_loss: 0.1738 - val_accuracy: 0.9768
Epoch 24/25
30/30 [==============================] - 8s 261ms/step - loss: 0.0019 - accuracy: 0.9997
- val_loss: 0.1778 - val_accuracy: 0.9757
Epoch 25/25
30/30 [==============================] - 8s 261ms/step - loss: 0.0018 - accuracy: 0.9997
- val_loss: 0.1871 - val_accuracy: 0.9789
```

Out[37]:

```
<keras.callbacks.History at 0x7f4401f0ccd0>
```

In [38]:

```
model.save("Ai_Spam_Identifier")
```

```
WARNING:absl:Function `_wrapped_model` contains input name(s) InputLayer with unsupported
characters which will be renamed to inputlayer in the SavedModel.
WARNING:absl:Found untraced functions such as lstm_cell_layer_call_fn, lstm_cell_layer_ca
ll_and_return_conditional_losses while saving (showing 2 of 2). These functions will not
be directly callable after loading.
```

In [39]:

```
test_sequences = tok.texts_to_sequences(X_test)
test_sequences_matrix = sequence.pad_sequences(test_sequences,maxlen=max_len)
```

In [40]:

```
accuracy = model.evaluate(test_sequences_matrix,Y_test)
print('Accuracy: {:0.3f}'.format(accuracy[1]))
```

```
27/27 [==============================] - 1s 42ms/step - loss: 0.1433 - accuracy: 0.9844
Accuracy: 0.984
```

In [41]:

```
y_pred = model.predict(test_sequences_matrix)
print(y_pred[25:40].round(3))
```

```
27/27 [==============================] - 3s 32ms/step
[[0.   ]
 [0.051]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.002]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]
 [0.   ]]
```

In [42]:

```
print(Y_test[25:40])
```

```
[[0]
 [1]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]
 [0]]
```