Import Required Library

In [3]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

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 Adam
from keras.preprocessing.text import Tokenizer
from keras.preprocessing import sequence
from keras.utils import pad_sequences
from keras.utils import to_categorical
from keras.callbacks import EarlyStopping
```

Read The Dataset

In [4]:

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

Out[4]:

	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

In [1]:

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

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

Pre-processing The Dataset

In [5]:

```
df.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'],axis=1,inplace=True)
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
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=0.25)
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 = pad_sequences(sequences,maxlen=max_len)
```

Create Model

Create Model

In [6]:

```
inputs = Input(shape=[max_len])
layer = Embedding(max_words,50,input_length=max_len)(inputs)
```

Add Layers

In [7]:

```
layer = LSTM(128)(layer)
layer = Dense(128)(layer)
layer = Activation('relu')(layer)
layer = Dropout(0.5)(layer)
layer = Dense(1.5)(layer)
layer = Activation('sigmoid')(layer)
model = Model(inputs=inputs,outputs=layer)
```

In [8]:

```
model.summary()
```

```
Model: "model"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 150)]             0

 embedding (Embedding)       (None, 150, 50)           50000

 lstm (LSTM)                 (None, 128)               91648

 dense (Dense)               (None, 128)               16512

 activation (Activation)     (None, 128)               0

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 1)                 129

 activation_1 (Activation)   (None, 1)                 0

=================================================================
Total params: 158,289
Trainable params: 158,289
Non-trainable params: 0
_________________________________________________________________
```

Compile the Model

In [9]:

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

Fit the Model

In [10]:

```
history = model.fit(sequences_matrix,Y_train,batch_size=20,epochs=15,validation_split=0.
2)
```

```
Epoch 1/15
168/168 [==============================] - 35s 191ms/step - loss: 0.1617 - accuracy: 0.95
00 - val_loss: 0.0474 - val_accuracy: 0.9856
Epoch 2/15
168/168 [==============================] - 33s 199ms/step - loss: 0.0389 - accuracy: 0.98
80 - val_loss: 0.0585 - val_accuracy: 0.9809
Epoch 3/15
168/168 [==============================] - 31s 185ms/step - loss: 0.0220 - accuracy: 0.99
```

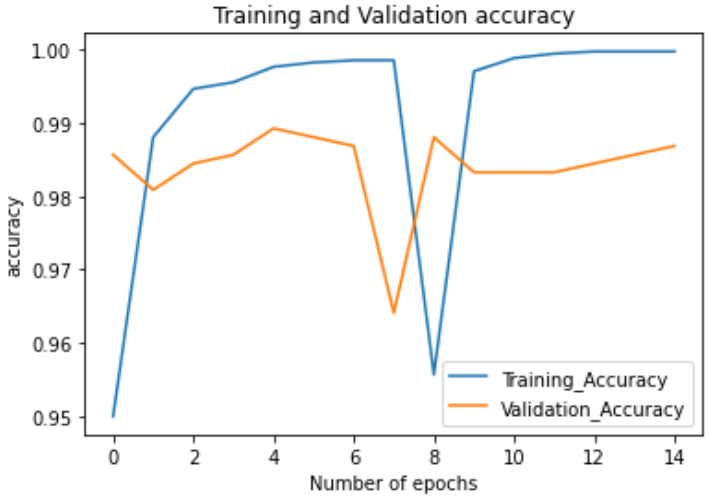
```
46 - val_loss: 0.0489 - val_accuracy: 0.9844
Epoch 4/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0144 - accuracy: 0.99
55 - val_loss: 0.0644 - val_accuracy: 0.9856
Epoch 5/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0107 - accuracy: 0.99
76 - val_loss: 0.0586 - val_accuracy: 0.9892
Epoch 6/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0070 - accuracy: 0.99
82 - val_loss: 0.0662 - val_accuracy: 0.9880
Epoch 7/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0060 - accuracy: 0.99
85 - val_loss: 0.0739 - val_accuracy: 0.9868
Epoch 8/15
168/168 [==============================] - 31s 183ms/step - loss: 0.0057 - accuracy: 0.99
85 - val_loss: 0.9505 - val_accuracy: 0.9641
Epoch 9/15
168/168 [==============================] - 30s 182ms/step - loss: 0.3441 - accuracy: 0.95
57 - val_loss: 0.0691 - val_accuracy: 0.9880
Epoch 10/15
168/168 [==============================] - 33s 197ms/step - loss: 0.0108 - accuracy: 0.99
70 - val_loss: 0.0735 - val_accuracy: 0.9833
Epoch 11/15
168/168 [==============================] - 31s 182ms/step - loss: 0.0047 - accuracy: 0.99
88 - val_loss: 0.0728 - val_accuracy: 0.9833
Epoch 12/15
168/168 [==============================] - 31s 185ms/step - loss: 0.0034 - accuracy: 0.99
94 - val_loss: 0.0759 - val_accuracy: 0.9833
Epoch 13/15
168/168 [==============================] - 31s 183ms/step - loss: 0.0029 - accuracy: 0.99
97 - val_loss: 0.0764 - val_accuracy: 0.9844
Epoch 14/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0026 - accuracy: 0.99
97 - val_loss: 0.0808 - val_accuracy: 0.9856
Epoch 15/15
168/168 [==============================] - 31s 184ms/step - loss: 0.0027 - accuracy: 0.99
97 - val_loss: 0.0739 - val_accuracy: 0.9868
```

In [11]:

```
metrics = pd.DataFrame(history.history)
metrics.rename(columns = {'loss': 'Training_Loss', 'accuracy': 'Training_Accuracy', 'val
_loss': 'Validation_Loss', 'val_accuracy': 'Validation_Accuracy'}, inplace = True)
def plot_graphs1(var1, var2, string):
    metrics[[var1, var2]].plot()
    plt.title('Training and Validation ' + string)
    plt.xlabel ('Number of epochs')
    plt.ylabel(string)
    plt.legend([var1, var2])
```

In [12]:

```
plot_graphs1('Training_Accuracy', 'Validation_Accuracy', 'accuracy')
```



Save The Model

In [13]:

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

Test The Model

In [14]:

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

In [15]:

```
accuracy1 = model.evaluate(test_sequences_matrix,Y_test)
```

```
44/44 [==============================] - 4s 82ms/step - loss: 0.1016 - accuracy: 0.9828
```

In [18]:

```
print(' Accuracy: {:0.5f}'.format(accuracy1[0],accuracy1[1]))
```

```
 Accuracy: 0.10165
```