

Import Required Library

In [30]:

```
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 [20]:

```
df = pd.read_csv('/content/spam(2).csv', delimiter=',', encoding='latin-1')
df.head()
```

Out[20]:

	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 [21]:

```
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

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)(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 [22]:

```
history = model.fit(sequences_matrix,Y_train,batch_size=15,epochs=15,validation_split=0.2)
```

```
Epoch 1/15
223/223 [=====] - 37s 164ms/step - loss: 0.1889 - accuracy: 0.94
50 - val_loss: 0.0452 - val_accuracy: 0.9844
Epoch 2/15
223/223 [=====] - 35s 155ms/step - loss: 0.0492 - accuracy: 0.98
83 - val_loss: 0.0346 - val_accuracy: 0.9904
Epoch 3/15
223/223 [=====] - 34s 154ms/step - loss: 0.0245 - accuracy: 0.99
31 - val loss: 0.0552 - val accuracv: 0.9797
```

```

Epoch 4/15
223/223 [=====] - 34s 154ms/step - loss: 0.0135 - accuracy: 0.99
64 - val_loss: 0.0388 - val_accuracy: 0.9868
Epoch 5/15
223/223 [=====] - 34s 154ms/step - loss: 0.0081 - accuracy: 0.99
70 - val_loss: 0.0440 - val_accuracy: 0.9856
Epoch 6/15
223/223 [=====] - 35s 159ms/step - loss: 0.0040 - accuracy: 0.99
94 - val_loss: 0.0424 - val_accuracy: 0.9880
Epoch 7/15
223/223 [=====] - 35s 157ms/step - loss: 0.0018 - accuracy: 0.99
94 - val_loss: 0.0478 - val_accuracy: 0.9880
Epoch 8/15
223/223 [=====] - 37s 166ms/step - loss: 6.6291e-04 - accuracy:
1.0000 - val_loss: 0.0519 - val_accuracy: 0.9904
Epoch 9/15
223/223 [=====] - 35s 157ms/step - loss: 2.1603e-04 - accuracy:
1.0000 - val_loss: 0.0599 - val_accuracy: 0.9892
Epoch 10/15
223/223 [=====] - 35s 157ms/step - loss: 1.5218e-04 - accuracy:
1.0000 - val_loss: 0.0608 - val_accuracy: 0.9880
Epoch 11/15
223/223 [=====] - 35s 156ms/step - loss: 8.5843e-05 - accuracy:
1.0000 - val_loss: 0.0628 - val_accuracy: 0.9892
Epoch 12/15
223/223 [=====] - 35s 158ms/step - loss: 7.7330e-05 - accuracy:
1.0000 - val_loss: 0.0670 - val_accuracy: 0.9868
Epoch 13/15
223/223 [=====] - 35s 155ms/step - loss: 6.2932e-05 - accuracy:
1.0000 - val_loss: 0.0709 - val_accuracy: 0.9868
Epoch 14/15
223/223 [=====] - 35s 159ms/step - loss: 5.1821e-05 - accuracy:
1.0000 - val_loss: 0.0768 - val_accuracy: 0.9868
Epoch 15/15
223/223 [=====] - 38s 169ms/step - loss: 3.9964e-05 - accuracy:
1.0000 - val_loss: 0.0758 - val_accuracy: 0.9868

```

In [23]:

```

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])

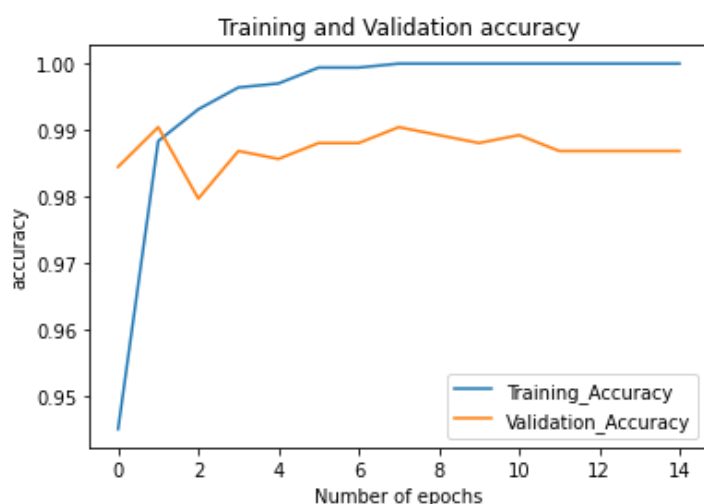
```

In [24]:

```

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

```



Save The Model

In [25]:

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

Test The Model

In [26]:

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

In [27]:

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

```
44/44 [=====] - 4s 80ms/step - loss: 0.1730 - accuracy: 0.9806
```

In [29]:

```
print(' Accuracy: {:.4f}'.format(accuracy1[0],accuracy1[1]))
```

```
Accuracy: 0.1730
```