# Importing necessary libraries

```
In [1]: import pandas as pd
        import numpy as np
        import nltk
        from nltk.tokenize import word_tokenize
        from nltk.corpus import stopwords
        nltk.download('stopwords')
        from nltk.stem import PorterStemmer,WordNetLemmatizer
        from tensorflow.keras.preprocessing.text import Tokenizer
        from tensorflow.keras.preprocessing import sequence
        from tensorflow.keras.models import Sequential
        from tensorflow.keras.layers import Dense, SimpleRNN,LSTM, Dropout, Embedding
        import warnings
        warnings.filterwarnings("ignore")
```

# Reading the dataset

```
In [2]: df=pd.read_csv('news.csv')
        df
```

Out[2]:

| | title | text | label |
|---|---|---|---|
| 0 | Donald Trump Sends Out Embarrassing New Year'... | Donald Trump just couldn t wish all Americans ... | fake |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | fake |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | fake |
| 3 | Trump Is So Obsessed He Even Has Obama's Name... | On Christmas day, Donald Trump announced that ... | fake |
| 4 | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | fake |
| ... | ... | ... | ... |
| 44893 | 'Fully committed' NATO backs new U.S. approach... | BRUSSELS (Reuters) - NATO allies on Tuesday we... | true |
| 44894 | LexisNexis withdrew two products from Chinese ... | LONDON (Reuters) - LexisNexis, a provider of l... | true |
| 44895 | Minsk cultural hub becomes haven from authorities | MINSK (Reuters) - In the shadow of disused Sov... | true |
| 44896 | Vatican upbeat on possibility of Pope Francis ... | MOSCOW (Reuters) - Vatican Secretary of State ... | true |
| 44897 | Indonesia to buy $1.14 billion worth of Russia... | JAKARTA (Reuters) - Indonesia will buy 11 Sukh... | true |

44898 rows × 3 columns

# Checking for null values

In [3]: 
```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44898 entries, 0 to 44897
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   title   44898 non-null  object
 1   text    44898 non-null  object
 2   label   44898 non-null  object
dtypes: object(3)
memory usage: 1.0+ MB
```

# Checking if there are any empty spaces in string

In [4]: 
```python
a=[]
for indx,title,text,label in df.itertuples():
    if title.isspace():
        a.append(indx)
    if text.isspace():
        a.append(indx)
```

In [5]: 
```python
a
```

Out[5]: 
```
[10923,
 11041,
 11190,
 11225,
 11236,
 11241,
 11247,
 11249,
 11267,
 11268,
 11283,
 11284,
 11285,
 11289,
 11290,
 11292,
 11295,
 11296,
 11301,
 11303
```

In [6]: 
```python
df['label'].value_counts()
```

Out[6]: 
```
fake    23481
true    21417
Name: label, dtype: int64
```

In [7]:
```python
df.drop(index=a,inplace=True)   #removing those rows which have empty spaces in string
```

In [8]:
```python
df
```

Out[8]:

|  | title | text | label |
|---|---|---|---|
| 0 | Donald Trump Sends Out Embarrassing New Year'... | Donald Trump just couldn t wish all Americans ... | fake |
| 1 | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | fake |
| 2 | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | fake |
| 3 | Trump Is So Obsessed He Even Has Obama's Name... | On Christmas day, Donald Trump announced that ... | fake |
| 4 | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | fake |
| ... | ... | ... | ... |
| 44893 | 'Fully committed' NATO backs new U.S. approach... | BRUSSELS (Reuters) - NATO allies on Tuesday we... | true |
| 44894 | LexisNexis withdrew two products from Chinese ... | LONDON (Reuters) - LexisNexis, a provider of l... | true |
| 44895 | Minsk cultural hub becomes haven from authorities | MINSK (Reuters) - In the shadow of disused Sov... | true |
| 44896 | Vatican upbeat on possibility of Pope Francis ... | MOSCOW (Reuters) - Vatican Secretary of State ... | true |
| 44897 | Indonesia to buy $1.14 billion worth of Russia... | JAKARTA (Reuters) - Indonesia will buy 11 Sukh... | true |

44267 rows × 3 columns

# Creating a function to automatically preprocess the text according to requirements

In [13]:
```python
lem=WordNetLemmatizer()
```

In [14]:
```python
stp=stopwords.words('english')
```

In [15]:
```python
def clean(x):
    x=word_tokenize(x.lower())
    x=[i for i in x if i.isalpha()]
    x=[i for i in x if i not in stp]
    x=[lem.lemmatize(i,pos='v')for i in x]
    x=' '.join(x)
    return x
```

# Applying function for preprocessing title and text columns

In [16]: `df['title']=df['title'].apply(clean)`

In [17]: `df`

Out[17]:

| | title | text | label |
|---|---|---|---|
| 0 | donald trump send embarrass new year eve messa... | Donald Trump just couldn t wish all Americans ... | fake |
| 1 | drink brag trump staffer start russian collusi... | House Intelligence Committee Chairman Devin Nu... | fake |
| 2 | sheriff david clarke become internet joke thre... | On Friday, it was revealed that former Milwauk... | fake |
| 3 | trump obsess even obama name cod website image | On Christmas day, Donald Trump announced that ... | fake |
| 4 | pope francis call donald trump christmas speech | Pope Francis used his annual Christmas Day mes... | fake |
| ... | ... | ... | ... |
| 44893 | commit nato back new approach afghanistan | BRUSSELS (Reuters) - NATO allies on Tuesday we... | true |
| 44894 | lexisnexis withdraw two products chinese market | LONDON (Reuters) - LexisNexis, a provider of l... | true |
| 44895 | minsk cultural hub become authorities | MINSK (Reuters) - In the shadow of disused Sov... | true |
| 44896 | vatican upbeat possibility pope francis visit ... | MOSCOW (Reuters) - Vatican Secretary of State ... | true |
| 44897 | indonesia buy billion worth russian jet | JAKARTA (Reuters) - Indonesia will buy 11 Sukh... | true |

44267 rows × 3 columns

In [18]: `df['text']=df['text'].apply(clean)`

In [19]: `df`

Out[19]:

| | title | text | label |
|---|---|---|---|
| 0 | donald trump send embarrass new year eve messa... | donald trump wish americans happy new year lea... | fake |
| 1 | drink brag trump staffer start russian collusi... | house intelligence committee chairman devin nu... | fake |
| 2 | sheriff david clarke become internet joke thre... | friday reveal former milwaukee sheriff david c... | fake |
| 3 | trump obsess even obama name cod website image | christmas day donald trump announce would back... | fake |
| 4 | pope francis call donald trump christmas speech | pope francis use annual christmas day message ... | fake |
| ... | ... | ... | ... |
| 44893 | commit nato back new approach afghanistan | brussels reuters nato ally tuesday welcome pre... | true |
| 44894 | lexisnexis withdraw two products chinese market | london reuters lexisnexis provider legal regul... | true |
| 44895 | minsk cultural hub become authorities | minsk reuters shadow disused factories minsk s... | true |
| 44896 | vatican upbeat possibility pope francis visit ... | moscow reuters vatican secretary state cardina... | true |
| 44897 | indonesia buy billion worth russian jet | jakarta reuters indonesia buy sukhoi fighter j... | true |

44267 rows × 3 columns

# Splitting input and output columns and encoding the output column

In [20]:
```python
x=df['text']
y=df['label']
```

In [21]:
```python
from sklearn.preprocessing import LabelEncoder
```

In [22]:
```python
le=LabelEncoder()
```

In [23]:
```python
y=le.fit_transform(df['label'])
```

In [24]: `y`

Out[24]: `array([0, 0, 0, ..., 1, 1, 1])`

# Splitting the dataset into training and testing data

```
In [25]:  from sklearn.model_selection import train_test_split
          xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.3,random_state=0)
```

# Finding out the length of tokens/words for each text and representing them into a column

```
In [26]:  senlen=[]
          for i in df['text']:
              senlen.append(len(word_tokenize(i)))
```

```
In [27]:  df['sentlen']=senlen
```

```
In [28]:  df
```

Out[28]:

| | title | text | label | sentlen |
|---|---|---|---|---|
| 0 | donald trump send embarrass new year eve messa... | donald trump wish americans happy new year lea... | fake | 242 |
| 1 | drink brag trump staffer start russian collusi... | house intelligence committee chairman devin nu... | fake | 166 |
| 2 | sheriff david clarke become internet joke thre... | friday reveal former milwaukee sheriff david c... | fake | 288 |
| 3 | trump obsess even obama name cod website image | christmas day donald trump announce would back... | fake | 226 |
| 4 | pope francis call donald trump christmas speech | pope francis use annual christmas day message ... | fake | 209 |
| ... | ... | ... | ... | ... |
| 44893 | commit nato back new approach afghanistan | brussels reuters nato ally tuesday welcome pre... | true | 248 |
| 44894 | lexisnexis withdraw two products chinese market | london reuters lexisnexis provider legal regul... | true | 75 |
| 44895 | minsk cultural hub become authorities | minsk reuters shadow disused factories minsk s... | true | 163 |
| 44896 | vatican upbeat possibility pope francis visit ... | moscow reuters vatican secretary state cardina... | true | 117 |
| 44897 | indonesia buy billion worth russian jet | jakarta reuters indonesia buy sukhoi fighter j... | true | 133 |

44267 rows × 4 columns

# Taking into account 95 percentile value of length and defining that length as the maximum length of all texts

In [29]: `max(senlen)`

Out[29]: 4606

In [30]: `np.quantile(senlen,0.95)`

Out[30]: 503.0

In [31]: `lenmax=500`

# Using the tokenizer function to analyse the text and assign a no to each word based on their importance

In [32]:
```
tok=Tokenizer()
tok.fit_on_texts(xtrain)
```

In [33]: `tok.index_word`

Out[33]:
```
{1: 'say',
 2: 'trump',
 3: 'state',
 4: 'would',
 5: 'president',
 6: 'people',
 7: 'make',
 8: 'one',
 9: 'go',
 10: 'also',
 11: 'take',
 12: 'new',
 13: 'tell',
 14: 'house',
 15: 'reuters',
 16: 'clinton',
 17: 'donald',
 18: 'time',
 19: 'get',
```

In [34]: `len(tok.index_word)`

Out[34]: 77145

In [35]: `vocablen=len(tok.index_word)`

# Converting the texts/tokens into their respective assigned numbers

```
In [36]:  trainseq=tok.texts_to_sequences(xtrain)
```

```
In [37]:  trainseq
```

```
Out[37]:  [[17,
            2,
            175,
            1,
            138,
            32,
            720,
            597,
            162,
            2386,
            2633,
            97,
            783,
            5790,
            276,
            1328,
            9148,
            852,
            1448,
```

## Inserting zeros to make all the text length same as per defined length

```
In [38]:  trainpad=sequence.pad_sequences(trainseq,maxlen=lenmax)
```

```
In [39]:  trainpad
```

```
Out[39]:  array([[    0,     0,     0, ...,  100,  877,  742],
                 [    0,     0,     0, ...,   19,  135,  398],
                 [    0,     0,     0, ...,    6,   30, 1087],
                 ...,
                 [    0,     0,     0, ..., 1048,  233,   77],
                 [    0,     0,     0, ..., 5387,  548,    1],
                 [    0,     0,     0, ...,  155,  877,  742]])
```

## Performing the same preprocess options on testing data

```
In [40]:  testseq=tok.texts_to_sequences(xtest)
          testpad=sequence.pad_sequences(testseq,maxlen=lenmax)
```

## Building the neural network to train the model

In [42]:
```python
rnn = Sequential()

rnn.add(Embedding(vocablen+1,1000, input_length=lenmax, mask_zero=True))
rnn.add(LSTM(units=32, activation="tanh"))
rnn.add(Dense(units=32, activation="relu"))
rnn.add(Dropout(0.2))

rnn.add(Dense(units=1, activation="sigmoid"))

rnn.compile(optimizer="adam", loss="binary_crossentropy")

rnn.fit(trainpad, ytrain, batch_size=512, epochs=5)

ypred = rnn.predict(testpad)

ypred = ypred>0.5
```

```
Epoch 1/5
61/61 [==============================] - 201s 3s/step - loss: 0.2953
Epoch 2/5
61/61 [==============================] - 206s 3s/step - loss: 0.0566
Epoch 3/5
61/61 [==============================] - 209s 3s/step - loss: 0.0228
Epoch 4/5
61/61 [==============================] - 208s 3s/step - loss: 0.0079
Epoch 5/5
61/61 [==============================] - 212s 3s/step - loss: 0.0057
416/416 [==============================] - 34s 79ms/step
```

# Checking the accuracy

In [43]:
```python
from sklearn.metrics import classification_report
```

In [44]:
```python
print(classification_report(ytest,ypred))
```

```
              precision    recall  f1-score   support

           0       0.99      0.98      0.99      6818
           1       0.98      0.99      0.99      6463

    accuracy                           0.99     13281
   macro avg       0.99      0.99      0.99     13281
weighted avg       0.99      0.99      0.99     13281
```

In [ ]: