

```
pip install --upgrade tensorflow

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: tensorflow in /usr/local/lib/python3.10/dist-packages (2.12.0)
Requirement already satisfied: absl-py==1.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.4.0)
Requirement already satisfied: astunparse==1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers==2.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (23.3.3)
Requirement already satisfied: gast<0.4.0,>=0.2.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.4.0)
Requirement already satisfied: google-pasta==0.1.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.2.0)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.54.0)
Requirement already satisfied: h5py==2.9.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.8.0)
Requirement already satisfied: jax==0.3.15 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.4.10)
Requirement already satisfied: keras<2.13,>=2.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.12.0)
Requirement already satisfied: libclang==13.0.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (16.0.0)
Requirement already satisfied: numpy<1.24,>=1.22 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.22.4)
Requirement already satisfied: opt-einsum==2.3.2 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (3.3.0)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from tensorflow) (23.1)
Requirement already satisfied: protobuf==4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (4.21.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from tensorflow) (67.7.2)
Requirement already satisfied: six==1.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.16.0)
Requirement already satisfied: tensorboard<2.13,>=2.12 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.12.2)
Requirement already satisfied: tensorflow-estimator<2.13,>=2.12.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.12.0)
Requirement already satisfied: termcolor==1.1.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (2.3.0)
Requirement already satisfied: typing-extensions==3.6.6 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (4.5.0)
Requirement already satisfied: wrapt<1.15,>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (1.14.1)
Requirement already satisfied: tensorflow-io-gcs-filesystem==0.23.1 in /usr/local/lib/python3.10/dist-packages (from tensorflow) (0.32.0)
Requirement already satisfied: wheel<1.0,>=0.23.0 in /usr/local/lib/python3.10/dist-packages (from astunparse==1.6.0->tensorflow) (0.40.0)
Requirement already satisfied: ml-dtypes==0.1.0 in /usr/local/lib/python3.10/dist-packages (from jax==0.3.15->tensorflow) (0.1.0)
Requirement already satisfied: scipy==1.7 in /usr/local/lib/python3.10/dist-packages (from jax==0.3.15->tensorflow) (1.10.1)
Requirement already satisfied: google-auth<3,>=1.6.3 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (2.17.3)
Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (1.0.0)
Requirement already satisfied: markdown==2.6.8 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (3.4.3)
Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (2.27.1)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (0.7.0)
Requirement already satisfied: tensorboard-plugin-wit==1.6.0 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (1.8.1)
Requirement already satisfied: werkzeug==1.0.1 in /usr/local/lib/python3.10/dist-packages (from tensorboard<2.13,>=2.12->tensorflow) (2.3.0)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow) (5.3.0)
Requirement already satisfied: pyasn1-modules==0.2.1 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow) (0.3.0)
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.10/dist-packages (from google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow) (4.9)
Requirement already satisfied: requests-oauthlib==0.7.0 in /usr/local/lib/python3.10/dist-packages (from google-auth-oauthlib<1.1,>=0.5->tensorboard<2.13,>=2.12->tensorflow) (1.0.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (1.26.15)
Requirement already satisfied: certifi==2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (2022.12.7)
Requirement already satisfied: charset-normalizer==2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (2.0.0)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.21.0->tensorboard<2.13,>=2.12->tensorflow) (3.4)
Requirement already satisfied: MarkupSafe==2.1.1 in /usr/local/lib/python3.10/dist-packages (from werkzeug==1.0.1->tensorboard<2.13,>=2.12->tensorflow) (2.1.2)
Requirement already satisfied: pyasn1<0.6.0,>=0.4.6 in /usr/local/lib/python3.10/dist-packages (from pyasn1-modules==0.2.1->google-auth<3,>=1.6.3->tensorboard<2.13,>=2.12->tensorflow) (0.5.1)
Requirement already satisfied: oauthlib==3.0.0 in /usr/local/lib/python3.10/dist-packages (from requests-oauthlib==0.7.0->google-auth-oauthlib<1.1,>=0.5->tensorboard<2.13,>=2.12->tensorflow) (3.2.2)

!pip install plotly

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: plotly in /usr/local/lib/python3.10/dist-packages (5.13.1)
Requirement already satisfied: tenacity==6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly) (8.2.2)
```

NLTK(natural language toolkit), PUNKT(parameters from a corpus in an unsupervised) is an unsupervised trainable model, which means it can be trained on unlabeled data Data that has not been tagged with information identifying its characteristics, properties, or categories is referred to as unlabeled data.

```
import nltk
nltk.download('punkt')

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
True

import tensorflow as tf
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud, STOPWORDS
import nltk
import re
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
import gensim
from gensim.utils import simple_preprocess
from gensim.parsing.preprocessing import STOPWORDS
from tensorflow.keras.preprocessing.text import one_hot, Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Embedding, Input, LSTM, Conv1D, MaxPool1D, Bidirectional
from tensorflow.keras.models import Model
```

What is stemming in NLTK? Stemming is a technique used to reduce an inflected word down to its word stem

```
df_fake = pd.read_csv("Fake.csv")
df_fake.head()
```

	title	text	subject	date
0	Donald Trump Sends Out Embarrassing New Year...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017
...	Trump Is So Obsessed He Even Has	On Christmas day, Donald Trump	..	December 29,

df_fake.tail()

		title	text	subject	date
	23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016
	23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016
	23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016
	23479	How to Blow \$700 Million: Al Jazeera	21st Century Wire says Al Jazeera	Middle-	January 14,

df_true = pd.read_csv("True.csv")
df_true.head()

		title	text	subject	date
	0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
	1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
	2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017
	3	FBI Russia probe helped by Australian	WASHINGTON (Reuters) - Trump campaign	politics	December 30,

df_true.tail()

		title	text	subject	date
	21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017
	21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017
	21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017
	21415	Vatican upbeat on possibility of Pope	MOSCOW (Reuters) - Vatican Secretary of	world	August 22,

Performing Exploratory Data Analysis Note:

- Target Column isfake is created
- isfake is set to 1 for real news and 0 for fake news

Adding A Target Class Column To Indicate Whether The News Is Real Or Fake
df_true['isfake'] = 1
df_true.head()

		title	text	subject	date	isfake
	0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1
	1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1
	2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1
	3	FBI Russia probe helped by	WASHINGTON (Reuters) - Trump	politics	December 30,	1

df_fake['isfake'] = 0
df_fake.head()

		title	text	subject	date	isfake
	0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
	1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
	2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
	3	Trump Is So Obsessed He Even Has	On Christmas dav. Donald Trumo	News	December 29,	0

- 💡 Notes:
- By using Concat() function two dataframe df_true and df_fake convert into df dataframe
 - reset_index() method resets the index back to the default 0, 1, 2 etc indexes. By default, this method will keep the "old" indexes in a column named "index". To avoid this, drop parameter is used.

Concatenating Real And Fake News
df = pd.concat([df_true, df_fake]).reset_index(drop = True)
df

```

# Loading the dataset
df = pd.read_csv('data/fake_news.csv')

# Displaying the first few rows of the dataset
df.head()

# Checking if dataset is class-imbalanced or not
df['isfake'].value_counts()

0    23481
1    21417
Name: isfake, dtype: int64
```

- Notes:
- No of samples in 2 classes are 23481 and 21417
 - So, dataset is almost balanced
 - used drop() function to remove unnecessary column

```
df.drop(columns = ['date'], inplace = True)

df['original'] = df['title'] + ' ' + df['text']
df.head()
```

	title	text	subject	isfake	original
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	1	As U.S. budget fight looms, Republicans flip t...
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	1	U.S. military to accept transgender recruits o...
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	1	Senior U.S. Republican senator: 'Let Mr. Muell...
3	FBI Russia probe helped	WASHINGTON (Reuters) - Trump	1	FBI Russia probe helped

Performing Data Cleaning Notes:

- stopwords imported from nltk.corpus and extending using some words
- STOPWORDS imported from gensim.parsing.preprocessing is also used to remove stopwords
- NLTK library has 179 words in the stopword collection, whereas Gensim has 337 words
- Any tokens shorter than min_len=2 characters and greater than max_len=15 characters are discarded using simple_preprocess()
- 9276947 words are found by traversing through each word of each sentence
- 108704 unique words are obtained using set keyword
- join() method is used to combine words to sentence

```

nltk.download("stopwords")

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

# Obtaining Additional Stopwords From nltk
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
stop_words.extend(['from', 'subject', 're', 'edu', 'use'])

# Removing Stopwords And Remove Words With 2 Or Less Characters
def preprocess(text):
    result = []
    for token in gensim.utils.simple_preprocess(text):
        if token not in gensim.parsing.preprocessing.STOPWORDS and len(token) > 3 and token not in stop_words:
            result.append(token)

    return result

# Applying The Function To The Dataframe
df['clean'] = df['original'].apply(preprocess)
# Showing Original News
df['original'][0]

'As U.S. budget fight looms, Republicans flip their fiscal script WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who voted t
his month for a huge expansion of the national debt to pay for tax cuts, called himself a "fiscal conservative" on Sunday and urged budget restraint in 2018. In keeping
with a sharp pivot under way among Republicans, U.S. Representative Mark Meadows, speaking on CBS' "Face the Nation," drew a hard line on federal spending, which lawmak
ers are bracing to do battle over in January. When they return from the holidays on Wednesday, lawmakers will begin trying to pass a federal budget in a fight likely to
be linked to other issues, such as immigration policy, even as the November congressional election campaigns approach in which Republicans will seek to keep control of
Congress. President Donald Trump and his Republicans want a big budget increase in military spending, while Democrats also want proportional increases for non-...'

# Showing Cleaned Up News After Removing Stopwords
print(df['clean'][0])

['budget', 'fight', 'looms', 'republicans', 'flip', 'fiscal', 'script', 'washington', 'reuters', 'head', 'conservative', 'republican', 'faction', 'congress', 'voted', 'r
df
```

	title	text	subject	isfake	original	clean
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	1	As U.S. budget fight looms, Republicans flip t...	[budget, fight, looms, republicans, flip, fisc...
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	1	U.S. military to accept transgender recruits o...	[military, accept, transgender, recruits, mond...
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	1	Senior U.S. Republican senator: 'Let Mr. Muell...	[senior, republican, senator, mueller, washing...
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	1	FBI Russia probe helped by Australian diplomat...	[russia, probe, helped, australian, diplomat, ...
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	1	Trump wants Postal Service to charge 'much mor...	[trump, wants, postal, service, charge, amazon...
...
44893	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	0	McPain: John McCain Furious That Iran Treated ...	[mcpain, john, mccain, furious, iran, treated,...
44894	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	0	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	[justice, yahoo, settles, mail, privacy, class...
44895	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	0	Sunnistan: US and Allied 'Safe Zone' Plan to T...	[sunnistan, allied, safe, zone, plan, territor...

```
# Obtaining The Total Words Present In The Dataset
list_of_words = []
for i in df.clean:
    for j in i:
        list_of_words.append(j)
list_of_words

'transgender',
'advocacy',
'group',
'glad',
'called',
'decision',
'appeal',
'great',
'news',
'hoping',
'means',
'government',
'come',
'justify',
'good',
'military',
'country',
'levi',
'said',
'glad',
'american',
'civil',
'liberties',
'union',
'represent',
'plaintiffs',
'lawsuits',
'filed',
'administration',
'appealed',
'hard',
'line',
'conservative',
'supporters',
'trump',
'announced',
'july',
'prohibit',
'transgender',
'people',
'serving',
'military',
'reversing',
'democratic',
'president',
'barack',
'obama',
'policy',
'accepting',
'trump',
'said',
'twitter',
'time',
'military',
'burdened',
'tremendous',
'medical',
'costs',
'disruption',

len(list_of_words)

9276947

# Obtaining The Total Number Of Unique Words
total_words = len(list(set(list_of_words)))
total_words

108704

# Joining The Words Into A String
df['clean_joined'] = df['clean'].apply(lambda x: " ".join(x))
df
```

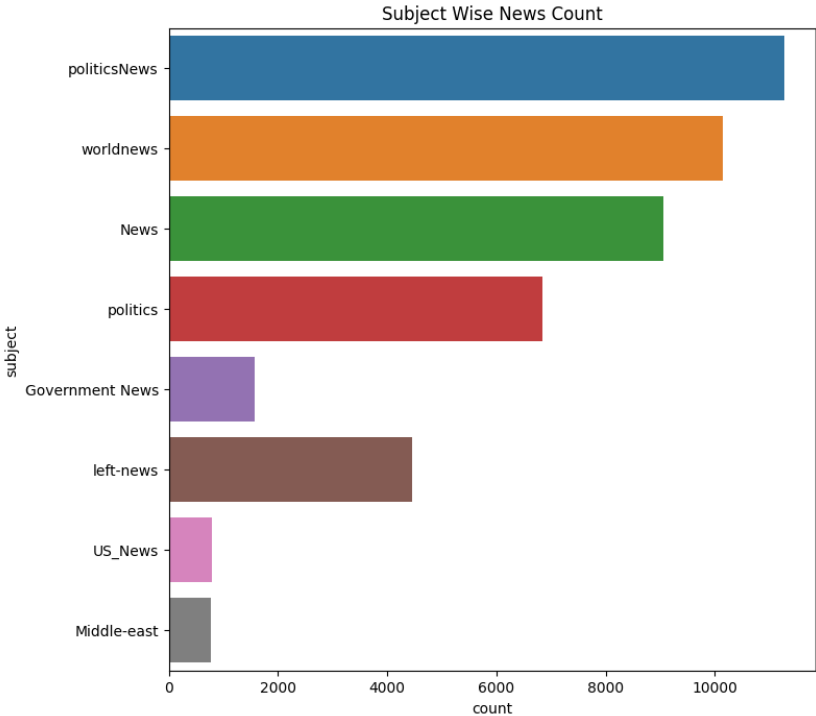
	title	text	subject	isfake	original	clean	clean_joined
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	1	As U.S. budget fight looms, Republicans flip t...	[budget, fight, looms, republicans, flip, fisc...	budget fight looms republicans flip fiscal scr...
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	1	U.S. military to accept transgender recruits o...	[military, accept, transgender, recruits, mond...	military accept transgender recruits monday pe...
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	1	Senior U.S. Republican senator: 'Let Mr. Muell...	[senior, republican, senator, mueller, washing...	senior republican senator mueller washington r...
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	1	FBI Russia probe helped by Australian diplomat...	[russia, probe, helped, australian, diplomat, ...	russia probe helped australian diplomat washin...
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	1	Trump wants Postal Service to charge 'much mor...	[trump, wants, postal, service, charge, amazon...	trump wants postal service charge amazon shipm...
...

df['clean_joined'][0]

'budget fight looms republicans flip fiscal script washington reuters head conservative republican faction congress voted month huge expansion national debt cuts called fiscal conservative sunday urged budget restraint keeping sharp pivot republicans representative mark meadows speaking face nation drew hard line federal spending lawma kers bracing battle january return holidays wednesday lawmakers begin trying pass federal budget fight likely linked issues immigration policy november congressional el ection campaigns approach republicans seek control congress president donald trump republicans want budget increase military spending democrats want proportional increa ses defense discretionary spending programs support education scientific research infrastructure public health environmental protection trump administration willing goi ng increase defense discretionary spending percent meadows chairman small influential house freedom caucus said program democrats saying need government raise p...'

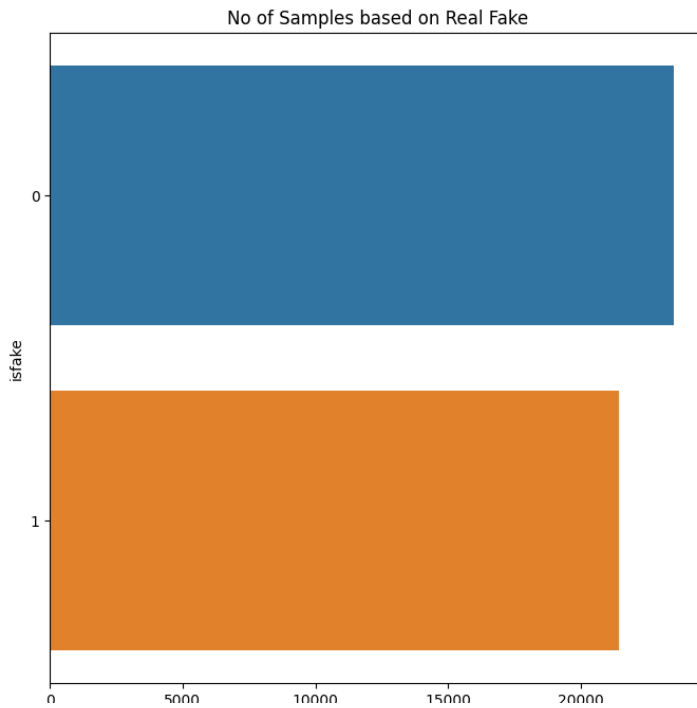
- Visualizing Cleaned Up Dataset 📌 Notes:
- Two countplot are plotted, one for Subject wise News Count, another for Real-Fake wise News Count
 - The most frequent words are among the given text are visualized using WordCloud

```
# Plotting The Number Of Samples In 'subject'
plt.figure(figsize = (8, 8))
plt.title("Subject Wise News Count")
sns.countplot(y = "subject", data = df)
plt.show()
```

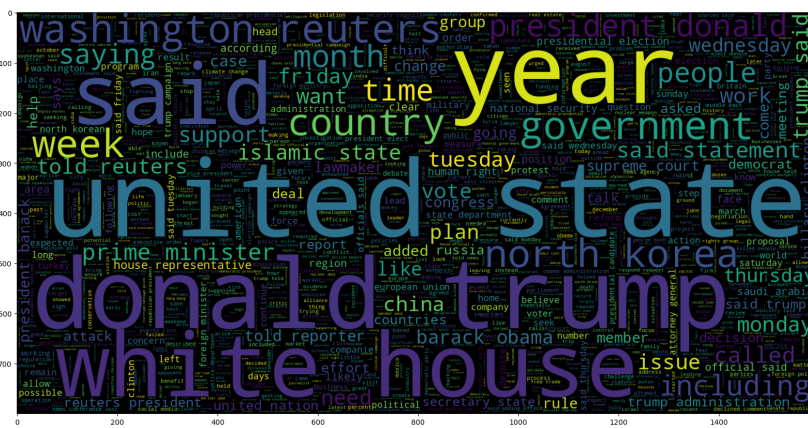


- Notes:
- politicsNews is in largest amount
 - US_News, Middle-east are in shortest amount

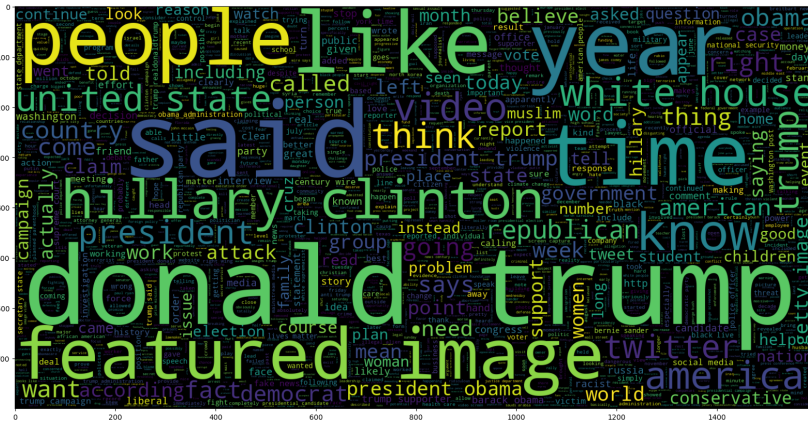
```
plt.figure(figsize = (8, 8))
plt.title("No of Samples based on Real Fake")
sns.countplot(y = "isfake", data = df)
plt.show()
```



```
plt.figure(figsize = (20,20))
wc = WordCloud(max_words = 2000 , width = 1600 , height = 800 , stopwords = stop_words).generate(" ".join(df[df.isfake == 1].clean_joined))
plt.imshow(wc, interpolation = 'bilinear')
plt.show()
```



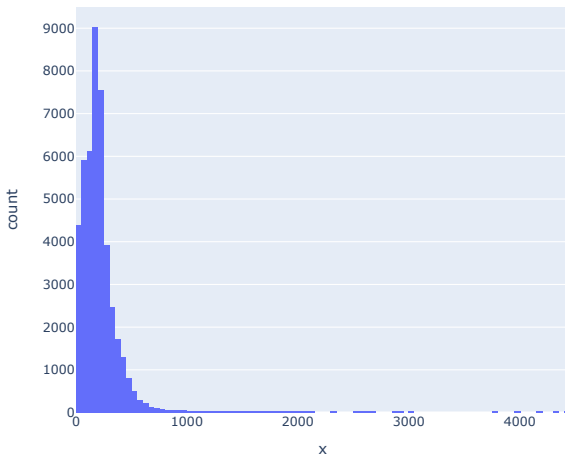
```
# Plotting The Word Cloud For Text That Is Fake
plt.figure(figsize = (20,20))
wc = WordCloud(max_words = 2000 , width = 1600 , height = 800 , stopwords = stop_words).generate(" ".join(df[df.isfake == 0].clean_joined))
plt.imshow(wc, interpolation = 'bilinear')
plt.show()
```



```
# Determining The Maximum Number Of Words In Any Document Required To Create Word Embeddings
maxlen = -1
for doc in df.clean_joined:
    tokens = nltk.word_tokenize(doc)
    if(maxlen < len(tokens)):
        maxlen = len(tokens)
print("The maximum number of words in any document is =", maxlen)
```

The maximum number of words in any document is = 4405

```
# Visualizing The Distribution Of Number Of Words In A Text
import plotly.express as px
fig = px.histogram(x = [len(nltk.word_tokenize(x)) for x in df.clean_joined], nbins = 100)
fig.show()
```



Preparing The Data By Performing Tokenization And Padding Notes:

- A text string is tokenized into individual words using word_tokenize
- maxlen is set to the largest no of tokens among texts
- Text Data is splitted into train-test as 80:20 ratio
- A set of text data is converted into a sequence of integers using texts_to_sequences
- All sequences have been converted to the same length using pad_sequences

```
# Splitting Data Into Test And Train
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(df.clean_joined, df.isfake, test_size = 0.2)
```

```
from nltk import word_tokenize
```

```
# Creating A Tokenizer To Tokenize The Words And Create Sequences Of Tokenized Words
tokenizer = Tokenizer(num_words = total_words)
tokenizer.fit_on_texts(x_train)
train_sequences = tokenizer.texts_to_sequences(x_train)
test_sequences = tokenizer.texts_to_sequences(x_test)
```

```
print("The encoding for document\n",df.clean_joined[0],"\n is : \n\n",train_sequences[0])
```

The encoding for document
budget fight looms republicans flip fiscal script washington reuters head conservative republican faction congress voted month huge expansion national debt cuts called
is :

[228, 4664, 7362, 475, 2866, 372, 9, 1862, 54, 228, 1797, 2518, 475, 1677, 1411, 372, 35, 1138, 2751, 475, 2866, 528, 753, 110, 48, 8, 457, 3443, 79, 69, 3991, 7362, 2

```
# Adding Padding
padded_train = pad_sequences(train_sequences,maxlen = 40, padding = 'post', truncating = 'post')
padded_test = pad_sequences(test_sequences,maxlen = 40, truncating = 'post')
```

```
for i,doc in enumerate(padded_train[:2]):
    print("The padded encoding for document",i+1," is : ",doc)

    The padded encoding for document 1 is : [ 228 4664 7362 475 2866 372 9 1862 54 228 1797 2518
    475 1677 1411 372 35 1138 2751 475 2866 528 753 110
    48 8 457 3443 79 69 3991 7362 2518 676 75 3388
    23254 2866 372 2033]
    The padded encoding for document 2 is : [ 497 151 17479 4186 111 497 1272 2415 25 497 151 2345
    173 1791 318 406 1324 931 52 32 93 54 4843 868
    8 2828 711 1327 406 931 45 19 695 975 376 64
    1171 499 1562 957]
```

Building And Training The Model Notes:

- Sequential model is built
- Integer inputs, which represent the index of words, are converted to dense vector of fixed size using Embedding
- Input Sequences are processed in both directions using Bidirectional
- A Bidirectional LSTM layer is created with 128 units

```
# Sequential Model
model = Sequential()

# Embedding layer
model.add(Embedding(total_words, output_dim = 128))

# Bi-Directional RNN and LSTM
model.add(Bidirectional(LSTM(128)))

# Dense layers
model.add(Dense(128, activation = 'relu'))
model.add(Dense(1,activation= 'sigmoid'))
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['acc'])
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
=====		
embedding (Embedding)	(None, None, 128)	13914112
bidirectional (Bidirectional LSTM)	(None, 256)	263168
dense (Dense)	(None, 128)	32896
dense_1 (Dense)	(None, 1)	129
=====		
Total params: 14,210,305		
Trainable params: 14,210,305		
Non-trainable params: 0		

```
total_words

108704
```

```
y_train = np.asarray(y_train)
# Training the model
model.fit(padded_train, y_train, batch_size = 64, validation_split = 0.1, epochs = 5)

Epoch 1/5
506/506 [=====] - 55s 91ms/step - loss: 0.0382 - acc: 0.9869 - val_loss: 0.0047 - val_acc: 0.9983
Epoch 2/5
506/506 [=====] - 15s 29ms/step - loss: 0.0043 - acc: 0.9993 - val_loss: 0.0023 - val_acc: 0.9992
Epoch 3/5
506/506 [=====] - 9s 18ms/step - loss: 8.5890e-04 - acc: 0.9998 - val_loss: 0.0065 - val_acc: 0.9989
Epoch 4/5
506/506 [=====] - 8s 16ms/step - loss: 9.6040e-04 - acc: 0.9997 - val_loss: 0.0072 - val_acc: 0.9983
Epoch 5/5
506/506 [=====] - 8s 16ms/step - loss: 7.9959e-04 - acc: 0.9998 - val_loss: 0.0053 - val_acc: 0.9992
<keras.callbacks.History at 0x7f2e345214b0>
```

Assessing Trained Model Performance Notes: Model Performance is assessed using-

- accuracy_score
- confusion_matrix
- classification_report

```
# Making prediction
pred = model.predict(padded_test)

281/281 [=====] - 2s 5ms/step

# If The Predicted Value Is >0.95 (i.e., More Than 95%), It Is Real Else It Is Fake
prediction = []
for i in range(len(pred)):
    if pred[i].item() > 0.95:
        prediction.append(1)
    else:
        prediction.append(0)

# Getting The Accuracy
from sklearn.metrics import accuracy_score

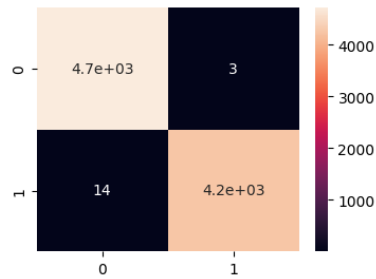
accuracy = accuracy_score(list(y_test), prediction)

print("Model Accuracy : ", accuracy)

Model Accuracy : 0.9981069042316258
```



```
# Getting The Confusion Matrix
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(list(y_test), prediction)
plt.figure(figsize = (4, 3))
sns.heatmap(cm, annot = True)
plt.show()
```



Notes:

- False Positive and False Negative is very less compare to True Positive and True Negative So, confusion_matix indicates model is performing good

```
# Getting The classification_report
from sklearn.metrics import classification_report
print(classification_report(list(y_test), prediction))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	4717
1	1.00	1.00	1.00	4263
accuracy			1.00	8980
macro avg	1.00	1.00	1.00	8980
weighted avg	1.00	1.00	1.00	8980

Notes:

- precision, recall, f1-score indicates model's performance is pretty well