

FAKE NEWS DETECTION USING NLP

TEAM MEMBER:822621104307 –
P.THOOYAVAN

Phase-3 Document submission



INTRODUCTION:



Induction for fake news detection using NLP is a process of training a machine learning model to identify fake news articles by analyzing their text. The model is

trained on a labeled dataset of news articles, where each article is labeled as either real or fake. The model learns to identify patterns in the text that are associated with fake news, such as certain words or phrases, grammatical errors, and inconsistencies in the story.

Once the model is trained, it can be used to predict whether a new news article is real or fake. To do this, the model extracts features from the article's text, such as the words used, the grammatical structure of the sentences, and the length of the article. The model then uses these features to predict whether the article is real or fake.

Induction for fake news detection using NLP is a powerful technique that can be used to identify fake news articles with high accuracy. However, it is important to note that the model is only as good as the data it is trained on. If the training data is biased or incomplete, the model will not be able to accurately identify fake news articles.

Here is a simple example of how induction can be used for fake news detection using NLP:

1. Collect a dataset of labeled news articles, where each article is labeled as either real or fake.
2. Split the dataset into a training set and a test set.
3. Train a machine learning model on the training set.
4. Evaluate the model's performance on the test set.
5. Use the trained model to predict whether new news articles are real or fake.

The following are some of the NLP techniques that can be used for fake news detection:

- **Feature extraction:** This involves extracting features from the text of a news article that are relevant to fake news detection. Some common features include the words used, the grammatical structure of the sentences, and the length of the article.
- **Feature selection:** This involves selecting the features that are most informative for fake news detection. This can be done using a variety of methods, such as statistical analysis or machine learning.
- **Classification:** This involves training a machine learning model to classify news articles as real or fake. Some common classification algorithms include logistic regression, support vector machines, and decision trees.

Induction for fake news detection using NLP is a rapidly evolving field, and new techniques are being developed all the time. As more data becomes available and NLP techniques continue to improve, we can expect to see even more accurate and effective fake news detection models in the future.

What is "Fake News"?

"Fake news" is a term that has come to mean different things to different people. At its core, we are defining "fake news" as those news stories that are false: the story itself is fabricated, with no verifiable facts, sources or quotes. Sometimes these stories may be propaganda that is intentionally designed to mislead the reader, or may be designed as "clickbait" written for economic incentives (the writer profits on the number of people who click on the story). In recent years, fake news stories have proliferated via social media, in part because they are so easily and quickly shared online.

About Dataset

This data set consists of 40000 fake and real news. Our goal is to train our model to accurately predict whether a particular piece of news is real or fake. Fake and real news data are given in two separate data sets, with each data set consisting of approximately 20000 articles.

Content:

1. [Import Libraries](#)
2. [Load and Check Data](#)
3. [Visualization](#)

1. [Data Cleaning](#)
 - [Removal of HTML Contents](#)
 - [Removal of Punctuation Marks and Special Characters](#)
 - [Removal of Stopwords](#)
 - [Lemmatization](#)
 - [Perform it for all the examples](#)
2. [N-Gram Analysis](#)
 - [Unigram Analysis](#)
 - [Bigram Analysis](#)
 - [Trigram Analysis](#)
3. [Modeling](#)
 - [Train - Test Split](#)
 - [Tokenizing](#)
 - [Training LSTM Model](#)
 - [Analysis After Training](#)

Import Libraries

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

import nltk
import re
import string

from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report

import keras
from keras.preprocessing import text, sequence
from keras.models import Sequential
from keras.layers import Dense, Embedding, LSTM, Dropout

import warnings
warnings.filterwarnings('ignore')

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
/kaggle/input/fake-and-real-news-dataset/True.csv
/kaggle/input/fake-and-real-news-dataset/Fake.csv
```

Load and Check Data

```
real_data = pd.read_csv('/kaggle/input/fake-and-real-news-dataset/True.csv')
fake_data = pd.read_csv('/kaggle/input/fake-and-real-news-dataset/Fake.csv')
```

linkcode

```
real_data.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 diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

	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
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

	title	text	subject	date	target
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

#Merging the 2 datasets

```
data = pd.concat([real_data, fake_data], ignore_index=True, sort=False)
data.head()
```

	title	text	subject	date	target
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 Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1

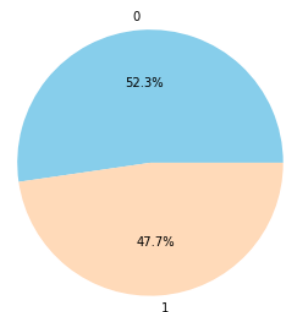
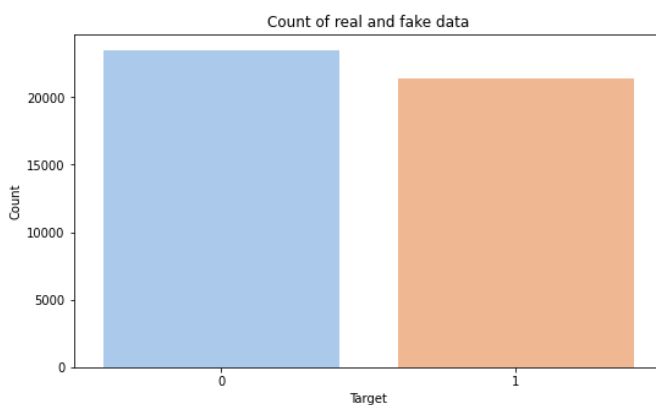
```
data.isnull().sum()
```

```
title      0
text       0
subject    0
date       0
target     0
dtype: int64
linkcode
```

Visualization

1.Count of Fake and Real Data

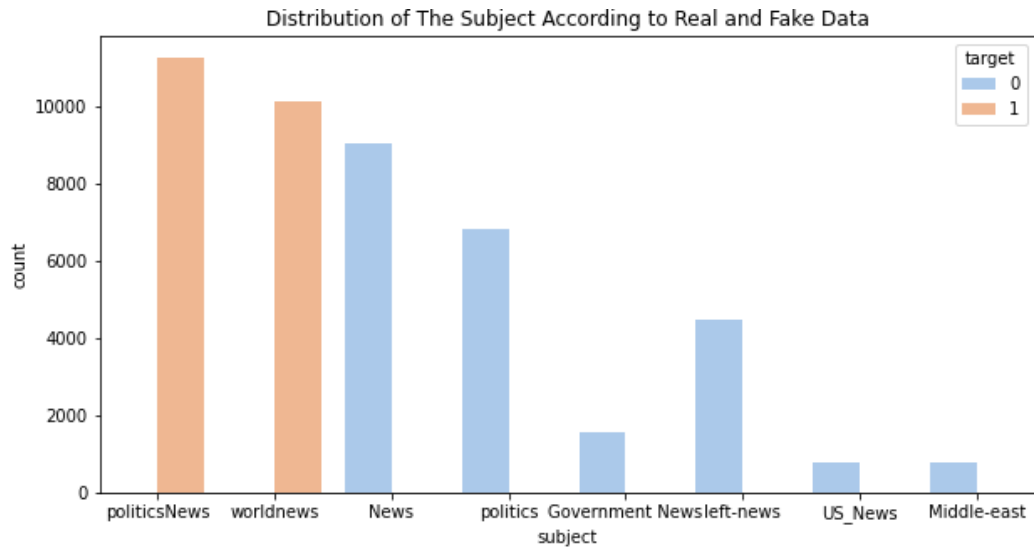
```
print(data["target"].value_counts())
fig, ax = plt.subplots(1,2, figsize=(19, 5))
g1 = sns.countplot(data.target,ax=ax[0],palette="pastel");
g1.set_title("Count of real and fake data")
g1.set_ylabel("Count")
g1.set_xlabel("Target")
g2 = plt.pie(data["target"].value_counts().values,explode=[0,0],labels=data.target.value_counts().index, autopct='%1.1f%%',colors=['SkyBlue', 'PeachPuff'])
fig.show()
0    23481
1    21417
Name: target, dtype: int64
```



2.Distribution of The Subject According to Real and Fake Data

```
print(data.subject.value_counts())
plt.figure(figsize=(10, 5))

ax = sns.countplot(x="subject", hue='target', data=data, palette="pastel")
plt.title("Distribution of The Subject According to Real and Fake Data")
politicsNews    11272
worldnews       10145
News            9050
politics        6841
left-news       4459
Government News 1570
US_News         783
Middle-east     778
Name: subject, dtype: int64
Text(0.5, 1.0, 'Distribution of The Subject According to Real and Fake Data')
```



Data Cleaning

```
data['text'] = data['subject'] + " " + data['title'] + " " + data['text']
del data['title']
del data['subject']
del data['date']
data.head()
```

	text	target
0	politicsNews As U.S. budget fight looms, Repub...	1
1	politicsNews U.S. military to accept transgend...	1
2	politicsNews Senior U.S. Republican senator: '...	1
3	politicsNews FBI Russia probe helped by Austra...	1
4	politicsNews Trump wants Postal Service to cha...	1

```
first_text = data.text[10]
first_text
```

Out[12]:

'politicsNews Jones certified U.S. Senate winner despite Moore challenge (Reuters)
 - Alabama officials on Thursday certified Democrat Doug Jones the winner of the state's U.S. Senate race, after a state judge denied a challenge by Republican Roy Moore, whose campaign was derailed by accusations of sexual misconduct with teenage girls. Jones won the vacant seat by about 22,000 votes, or 1.6 percentage points, election officials said. That made him the first Democrat in a quarter of a century to win a Senate seat in Alabama. The seat was previously held by Republican Jeff Sessions, who was tapped by U.S. President Donald Trump as attorney general. A state canvassing board composed of Alabama Secretary of State John Merrill, Governor Kay I

vey and Attorney General Steve Marshall certified the election results. Seating Jones will narrow the Republican majority in the Senate to 51 of 100 seats. In a statement, Jones called his victory "a new chapter" and pledged to work with both parties. Moore declined to concede defeat even after Trump urged him to do so. He stood by claims of a fraudulent election in a statement released after the certification and said he had no regrets, media outlets reported. An Alabama judge denied Moore's request to block certification of the results of the Dec. 12 election in a decision shortly before the canvassing board met. Moore's challenge alleged there had been potential voter fraud that denied him a chance of victory. His filing on Wednesday in the Montgomery Circuit Court sought to halt the meeting scheduled to ratify Jones' win on Thursday. Moore could ask for a recount, in addition to possible other court challenges, Merrill said in an interview with Fox News Channel. He would have to complete paperwork "within a timed period" and show he has the money for a challenge, Merrill said. "We've not been notified yet of their intention to do that," Merrill said. Regarding the claim of voter fraud, Merrill told CNN that more than 100 cases had been reported. "We've adjudicated more than 60 of those. We will continue to do that," he said. Republican lawmakers in Washington had distanced themselves from Moore and called for him to drop out of the race after several women accused him of sexual assault or misconduct dating back to when they were teenagers and he was in his early 30s. Moore has denied wrongdoing and Reuters has not been able to independently verify the allegations. '

Removal of HTML Contents

First, let's remove HTML content.

```
In [13]:
pip install bs4
Collecting bs4
  Downloading bs4-0.0.1.tar.gz (1.1 kB)
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.9.3-py3-none-any.whl (115 kB)
    |████████████████████████████████████████| 115 kB 1.3 MB/s
Collecting soupsieve>1.2
  Downloading soupsieve-2.2.1-py3-none-any.whl (33 kB)
Building wheels for collected packages: bs4
  Building wheel for bs4 (setup.py) ... -# #\# #done
  Created wheel for bs4: filename=bs4-0.0.1-py3-none-any.whl size=1273 sha256=2bea095cbbbc5fb6fc44736f40fce54b119a54eba4fa1dbedd43deddc70fda9b
  Stored in directory: /root/.cache/pip/wheels/0a/9e/ba/20e5bbc1afef3a491f0b3bb74d508f99403aabe76eda2167ca
Successfully built bs4
Installing collected packages: soupsieve, beautifulsoup4, bs4
Successfully installed beautifulsoup4-4.9.3 bs4-0.0.1 soupsieve-2.2.1
Note: you may need to restart the kernel to use updated packages.
```

```
In [14]:
from bs4 import BeautifulSoup

soup = BeautifulSoup(first_text, "html.parser")
first_text = soup.get_text()
first_text
```

```
Out[14]:
'politicsNews Jones certified U.S. Senate winner despite Moore challenge (Reuters)
- Alabama officials on Thursday certified Democrat Doug Jones the winner of the sta
te's U.S. Senate race, after a state judge denied a challenge by Republican Roy Moo
re, whose campaign was derailed by accusations of sexual misconduct with teenage gi
```


rls. Jones won the vacant seat by about 22,000 votes, or 1.6 percentage points, election officials said. That made him the first Democrat in a quarter of a century to win a Senate seat in Alabama. The seat was previously held by Republican Jeff Sessions, who was tapped by U.S. President Donald Trump as attorney general. A state canvassing board composed of Alabama Secretary of State John Merrill, Governor Kay Ivey and Attorney General Steve Marshall certified the election results. Seating Jones will narrow the Republican majority in the Senate to 51 of 100 seats. In a statement, Jones called his victory "a new chapter" and pledged to work with both parties. Moore declined to concede defeat even after Trump urged him to do so. He stood by claims of a fraudulent election in a statement released after the certification and said he had no regrets, media outlets reported. An Alabama judge denied Moore's request to block certification of the results of the Dec. 12 election in a decision shortly before the canvassing board met. Moore's challenge alleged there had been potential voter fraud that denied him a chance of victory. His filing on Wednesday in the Montgomery Circuit Court sought to halt the meeting scheduled to ratify Jones' win on Thursday. Moore could ask for a recount, in addition to possible other court challenges, Merrill said in an interview with Fox News Channel. He would have to complete paperwork "within a timed period" and show he has the money for a challenge, Merrill said. "We've not been notified yet of their intention to do that," Merrill said. Regarding the claim of voter fraud, Merrill told CNN that more than 100 cases had been reported. "We've adjudicated more than 60 of those. We will continue to do that," he said. Republican lawmakers in Washington had distanced themselves from Moore and called for him to drop out of the race after several women accused him of sexual assault or misconduct dating back to when they were teenagers and he was in his early 30s. Moore has denied wrongdoing and Reuters has not been able to independently verify the allegations. '

Removal of Punctuation Marks and Special Characters

Let's now remove everything except uppercase / lowercase letters using Regular Expressions.

In [15]:

```
first_text = re.sub('\[[^]]*\]', ' ', first_text)
first_text = re.sub('[^a-zA-Z]', ' ', first_text) # replaces non-alphabets with spaces
first_text = first_text.lower() # Converting from uppercase to lowercase
first_text
```

Out[15]:

```
'politicsnews jones certified u s senate winner despite moore challenge reuters
alabama officials on thursday certified democrat doug jones the winner of the sta
te s u s senate race after a state judge denied a challenge by republican roy moo
re whose campaign was derailed by accusations of sexual misconduct with teenage gi
rls jones won the vacant seat by about votes or percentage points ele
ction officials said that made him the first democrat in a quarter of a century to
win a senate seat in alabama the seat was previously held by republican jeff ses
sions who was tapped by u s president donald trump as attorney general a state c
anvassing board composed of alabama secretary of state john merrill governor kay i
vey and attorney general steve marshall certified the election results seating jon
es will narrow the republican majority in the senate to of seats in a state
ment jones called his victory a new chapter and pledged to work with both partie
s moore declined to concede defeat even after trump urged him to do so he stood b
y claims of a fraudulent election in a statement released after the certification a
nd said he had no regrets media outlets reported an alabama judge denied moore s
request to block certification of the results of the dec election in a decision
shortly before the canvassing board met moore s challenge alleged there had been
potential voter fraud that denied him a chance of victory his filing on wednesday
in the montgomery circuit court sought to halt the meeting scheduled to ratify jone
```

s win on thursday moore could ask for a recount in addition to possible other court challenges merrill said in an interview with fox news channel he would have to complete paperwork within a timed period and show he has the money for a challenge merrill said we've not been notified yet of their intention to do that merrill said regarding the claim of voter fraud merrill told cnn that more than cases had been reported we've adjudicated more than of those we will continue to do that he said republican lawmakers in washington had distanced themselves from moore and called for him to drop out of the race after several women accused him of sexual assault or misconduct dating back to when they were teenagers and he was in his early s moore has denied wrongdoing and reuters has not been able to independently verify the allegations '

Removal of Stopwords

Let's remove stopwords like is,a,the... Which do not offer much insight.

```
In [16]:
nltk.download("stopwords")
from nltk.corpus import stopwords

# we can use tokenizer instead of split
first_text = nltk.word_tokenize(first_text)
[nltk_data] Downloading package stopwords to /usr/share/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

In [17]:
first_text = [ word for word in first_text if not word in set(stopwords.words("english"))]
```

Lemmatization

Lemmatization to bring back multiple forms of same word to their common root like 'coming', 'comes' into 'come'.

```
In [18]:
lemma = nltk.WordNetLemmatizer()
first_text = [ lemma.lemmatize(word) for word in first_text]
```

```
first_text = " ".join(first_text)
first_text
```

Out[18]:

'politicsnews jones certified u senate winner despite moore challenge reuters alabama official thursday certified democrat doug jones winner state u senate race state judge denied challenge republican roy moore whose campaign derailed accusation sexual misconduct teenage girl jones vacant seat vote percentage point election official said made first democrat quarter century win senate seat alabama seat previously held republican jeff session tapped u president donald trump attorney general state canvassing board composed alabama secretary state john merrill governor kay ivey attorney general steve marshall certified election result seating jones narrow republican majority senate seat statement jones called victory new chapter pledged work party moore declined concede defeat even trump urged stood claim fraudulent election statement released certification said regret medium outlet reported alabama judge denied moore request block certification result dec election decision shortly canvassing board met moore challenge alleged potential voter fraud denied chance victory filing wednesday montgomery circuit court sought halt meeting scheduled ratify jones win thursday moore could ask recount addition possible court challenge merrill

said interview fox news channel would complete paperwork within timed period show money challenge merrill said notified yet intention merrill said regarding claim voter fraud merrill told cnn case reported adjudicated continue said republican lawmaker washington distanced moore called drop race several woman accused sexual assault misconduct dating back teenager early moore denied wrongdoing reuters able independently verify allegation'

Perform it for all the examples

We performed the steps for a single example. Now let's perform it for all the examples in the data.

```
In [19]:
#Removal of HTML Contents
def remove_html(text):
    soup = BeautifulSoup(text, "html.parser")
    return soup.get_text()

#Removal of Punctuation Marks
def remove_punctuations(text):
    return re.sub('\[[^\]]*\]', '', text)

# Removal of Special Characters
def remove_characters(text):
    return re.sub("[^a-zA-Z]", " ", text)

#Removal of stopwords
def remove_stopwords_and_lemmatization(text):
    final_text = []
    text = text.lower()
    text = nltk.word_tokenize(text)

    for word in text:
        if word not in set(stopwords.words('english')):
            lemma = nltk.WordNetLemmatizer()
            word = lemma.lemmatize(word)
            final_text.append(word)
    return " ".join(final_text)

#Total function
def cleaning(text):
    text = remove_html(text)
    text = remove_punctuations(text)
    text = remove_characters(text)
    text = remove_stopwords_and_lemmatization(text)
    return text

#Apply function on text column
data['text'] = data['text'].apply(cleaning)

In [20]:
data.head()

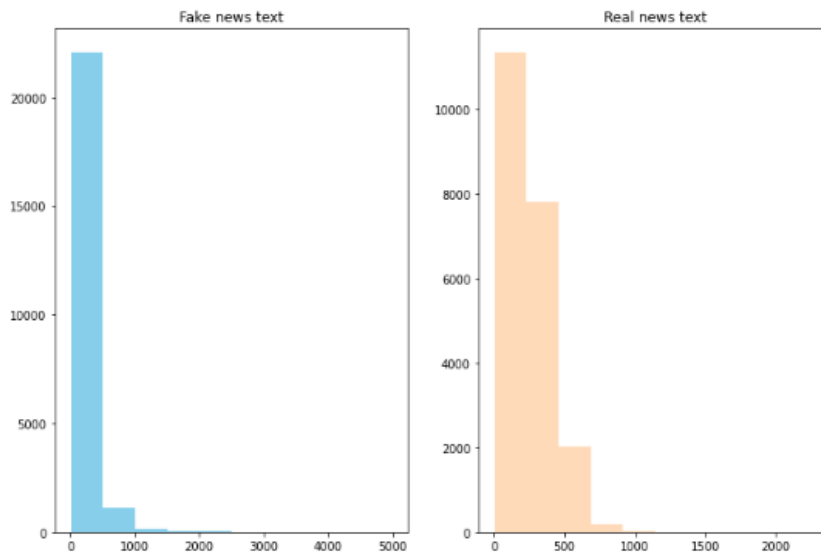
Out[20]:
```

	text	target
--	------	--------

Number of words in each text

In [23]:

```
fig,(ax1,ax2)=plt.subplots(1,2,figsize=(12,8))
text_len=data[data['target']==0]['text'].str.split().map(lambda x: len(x))
ax1.hist(text_len,color='SkyBlue')
ax1.set_title('Fake news text')
text_len=data[data['target']==1]['text'].str.split().map(lambda x: len(x))
ax2.hist(text_len,color='PeachPuff')
ax2.set_title('Real news text')
fig.suptitle('Words in texts')
plt.show()
```



```
texts = ' '.join(data['text'])
```

In [25]:

```
string = texts.split(" ")
```

In [26]:

```
def draw_n_gram(string,i):
    n_gram = (pd.Series(nltk.ngrams(string, i)).value_counts())[:15]
    n_gram_df=pd.DataFrame(n_gram)
    n_gram_df = n_gram_df.reset_index()
    n_gram_df = n_gram_df.rename(columns={"index": "word", 0: "count"})
    print(n_gram_df.head())
    plt.figure(figsize = (16,9))
    return sns.barplot(x='count',y='word', data=n_gram_df)
```

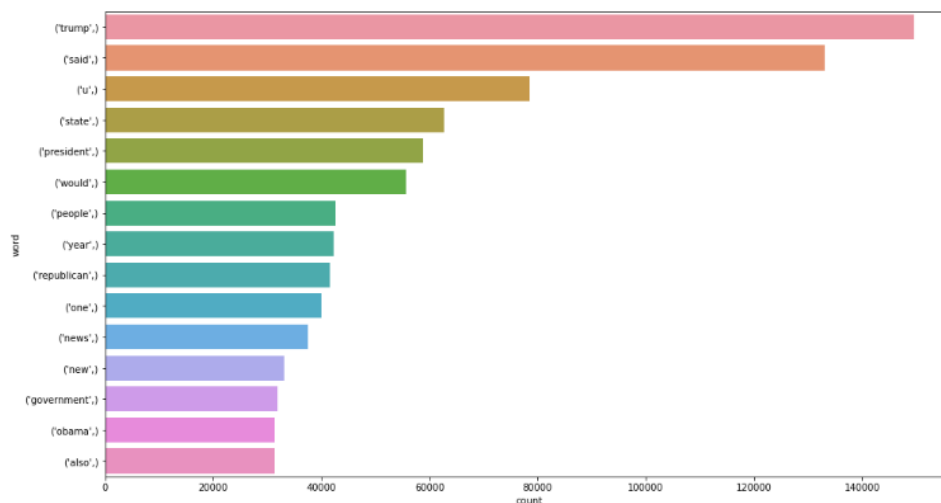
Unigram Analysis

In [27]:

```
draw_n_gram(string,1)
      word  count
0      (trump,) 149603
1      (said,) 133030
2      (u,) 78516
3      (state,) 62726
4      (president,) 58790
```

Out[27]:

```
<AxesSubplot:xlabel='count', ylabel='word'>
```



rain Test Split

In [30]:

```
X_train, X_test, y_train, y_test = train_test_split(data['text'], data['target'], random_state=0)
```

Tokenizing

- Tokenizing Text -> Representing each word by a number
- Mapping of original word to number is preserved in word_index property of tokenizer

Lets keep all news to 300, add padding to news with less than 300 words and truncating long ones

In [31]:

```
max_features = 10000
maxlen = 300
```

In [32]:

```
tokenizer = text.Tokenizer(num_words=max_features)
tokenizer.fit_on_texts(X_train)
tokenized_train = tokenizer.texts_to_sequences(X_train)
X_train = sequence.pad_sequences(tokenized_train, maxlen=maxlen)
```

```
In [33]:
tokenized_test = tokenizer.texts_to_sequences(X_test)
X_test = sequence.pad_sequences(tokenized_test, maxlen=maxlen)
```

Training LSTM Model

```
In [34]:
batch_size = 256
epochs = 10
embed_size = 100

In [35]:
model = Sequential()
#Non-trainable embedding layer
model.add(Embedding(max_features, output_dim=embed_size, input_length=maxlen, trainable=False))
#LSTM
model.add(LSTM(units=128, return_sequences=True, recurrent_dropout=0.25, dropout=0.25))
model.add(LSTM(units=64, recurrent_dropout=0.1, dropout=0.1))
model.add(Dense(units=32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile(optimizer=keras.optimizers.Adam(lr=0.01), loss='binary_crossentropy', metrics=['accuracy'])

In [36]:
model.summary()
Model: "sequential"
```

Layer (type)	Output Shape	Param #
=====		
embedding (Embedding)	(None, 300, 100)	1000000

lstm (LSTM)	(None, 300, 128)	117248

lstm_1 (LSTM)	(None, 64)	49408

dense (Dense)	(None, 32)	2080

dense_1 (Dense)	(None, 1)	33
=====		
Total params: 1,168,769		
Trainable params: 168,769		
Non-trainable params: 1,000,000		

```
In [37]:
history = model.fit(X_train, y_train, validation_split=0.3, epochs=10, batch_size=batch_size, shuffle=True, verbose=1)
Epoch 1/10
93/93 [=====] - 268s 3s/step - loss: 0.5514 - accuracy: 0.7044 - val_loss: 1.2749 - val_accuracy: 0.5668
Epoch 2/10
93/93 [=====] - 261s 3s/step - loss: 0.3611 - accuracy: 0.8452 - val_loss: 0.2542 - val_accuracy: 0.8987
Epoch 3/10
```

```

93/93 [=====] - 263s 3s/step - loss: 0.2870 - accuracy: 0.
8763 - val_loss: 0.2555 - val_accuracy: 0.8998
Epoch 4/10
93/93 [=====] - 264s 3s/step - loss: 0.2686 - accuracy: 0.
8857 - val_loss: 0.2131 - val_accuracy: 0.9171
Epoch 5/10
93/93 [=====] - 264s 3s/step - loss: 0.2209 - accuracy: 0.
9162 - val_loss: 0.1326 - val_accuracy: 0.9435
Epoch 6/10
93/93 [=====] - 263s 3s/step - loss: 0.1733 - accuracy: 0.
9389 - val_loss: 0.1308 - val_accuracy: 0.9392
Epoch 7/10
93/93 [=====] - 267s 3s/step - loss: 0.0712 - accuracy: 0.
9695 - val_loss: 0.0389 - val_accuracy: 0.9860
Epoch 8/10
93/93 [=====] - 269s 3s/step - loss: 0.0414 - accuracy: 0.
9843 - val_loss: 0.0402 - val_accuracy: 0.9838
Epoch 9/10
93/93 [=====] - 276s 3s/step - loss: 0.0418 - accuracy: 0.
9842 - val_loss: 0.0400 - val_accuracy: 0.9875
Epoch 10/10
93/93 [=====] - 270s 3s/step - loss: 0.0317 - accuracy: 0.
9886 - val_loss: 0.0454 - val_accuracy: 0.9828

```

Analysis After Training

```

In [38]:
print("Accuracy of the model on Training Data is - " , model.evaluate(X_train,y_train)[
1]*100 , "%")
print("Accuracy of the model on Testing Data is - " , model.evaluate(X_test,y_test)[1]*
100 , "%")
1053/1053 [=====] - 101s 96ms/step - loss: 0.0393 - accuracy: 0.9843
Accuracy of the model on Training Data is - 98.42603802680969 %
351/351 [=====] - 34s 97ms/step - loss: 0.0397 - accuracy: 0.9840
Accuracy of the model on Testing Data is - 98.39643836021423 %

```

```

In [39]:
plt.figure()
plt.plot(history.history["accuracy"], label = "Train")
plt.plot(history.history["val_accuracy"], label = "Test")
plt.title("Accuracy")
plt.ylabel("Acc")
plt.xlabel("epochs")
plt.legend()
plt.show()

```

```

In [40]:
plt.figure()
plt.plot(history.history["loss"], label = "Train")
plt.plot(history.history["val_loss"], label = "Test")
plt.title("Loss")
plt.ylabel("Acc")

```



```
plt.xlabel("epochs")
plt.legend()
plt.show()
```

```
In [41]:
```

```
pred = model.predict_classes(X_test)
print(classification_report(y_test, pred, target_names = ['Fake', 'Real']))
```

	precision	recall	f1-score	support
Fake	1.00	0.97	0.98	5858
Real	0.97	1.00	0.98	5367
accuracy			0.98	11225
macro avg	0.98	0.98	0.98	11225
weighted avg	0.98	0.98	0.98	11225

CONCLUSION:

NLP-based fake news detection models can be trained on datasets of labeled news articles, and then used to predict whether new news articles are real or fake. These models can be deployed in a variety of ways, such as in social media platforms or in news aggregator websites.

NLP-based fake news detection models have achieved promising results in recent years. For example, a study by researchers at Stanford University found that a deep learning model was able to detect fake news with an accuracy of over 90%. However, it is important to note that these models are still under development, and they can be fooled by sophisticated fake news articles.