A Project Report
on

Fake News Classification using Machine Learning

Submitted in partial fulfilment of the requirement
for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

Under The Supervision of
Name of Supervisor: Ms. Pragya Agarwal
Designation

Submitted By

S. No	Name	Admission Number
1	Aryan Singh	21SCSE1180138
2	Akarshan Upadhyay	21SCSE1280030
3	Ayush.	21SCSE1011056
4	Ankit Kumar	21SCSE1280033

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND
ENGINEERING GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
June ,2023

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"Fake News Classification using Machine Learning"** in partial fulfilment of the requirements for the award of the **B. Tech** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of **Ms. Pragya Agarwal** Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Aryan Singh, 21SCSE1180138

Akarshan Upadhyay, 21SCSE1280030

Ayush., 21SCSE1011056

Ankit Kumar, 21SCSE1280033

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Ms. Pragya Agarwal

Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of **Aryan Singh(21SCSE1180138), Akarshan Upadhyay(21SCSE1280030), Ayush.,(21SCSE1011056), Ankit Kumar(21SCSE1280033)** has been held on **FAKE NEWS CLASSFICATION USING MACHINE LEARNING** and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY**.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Program Chair **Signature of Dean**

Date: June ,2023

Place: Greater Noida

Abstract

Most of the smart phone users prefer to read the news via social media over internet. The news websites are publishing the news and provide the source of authentication. The question is how to authenticate the news and articles which are circulated among social media like WhatsApp groups, Facebook Pages, Twitter and other micro blogs & social networking sites. It is harmful for the society to believe on the rumors and pretend to be a news. The need of an hour is to stop the rumors especially in the developing countries like India, and focus on the correct, authenticated news articles. This paper demonstrates a model and the methodology for fake news detection. With the help of Machine learning and natural language processing, it is tried to aggregate the news and later determine whether the news is real or fake using Support Vector Machine (TF-IDF). The results of the proposed model is compared with existing models. The proposed model is working well and defining the correctness of results upto 92.0% of accuracy.

Table of Contents

Chapter: 1

Introduction

1.1) Overview

The advent of the World Wide Web and the rapid adoption of social media platforms (such as Facebook and Twitter) paved the way for information dissemination that has never been witnessed in the human history before. Besides other use cases, news outlets benefitted from the widespread use of social media platforms by providing updated news in near real time to its subscribers. The news media evolved from newspapers, tabloids, and magazines to a digital form such as online news platforms, blogs, social media feeds, and other digital media formats. It became easier for consumers to acquire the latest news at their fingertips. Facebook referrals account for 70% of traffic to news websites. These social media platforms in their current state are extremely powerful and useful for their ability to allow users to discuss and share ideas and debate over issues such as democracy, education, and health. However, such platforms are also used with a negative perspective by certain entities commonly for monetary gain and in other cases for creating biased opinions, manipulating mindsets, and spreading satire or absurdity. The phenomenon is commonly known as fake news.

1.2) Tools and Technology

1.2.1) Natural Language Processing:

NLP stands for Natural Language Processing, which is a part of Computer Science, Human language, and Artificial Intelligence**.** It is the technology that is used by machines to understand, analyse, manipulate, and interpret human's languages. It helps developers to organize knowledge for performing tasks such as translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction**,** and topic segmentation**.** The main reason for utilizing Natural Language Processing is to consider one or more specializations of system or an algorithm. The Natural Language Processing (NLP) rating of an algorithmic system enables the combination of speech understanding and speech generation. In addition, it could be utilized to detect actions with various languages. suggested a new ideal system for extraction actions from languages of English, Italian and Dutch speeches through utilizing various pipelines of various languages such as Emotion Analyzer and Detection, Named Entity Recognition (NER), Parts of Speech (POS) Taggers, Chunking, and Semantic Role Labeling made NLP good Subject of the search.

1.2.2) Machine Learning:

In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions. But can a machine also learn from experiences or past data like a human does? So here comes the role of **Machine Learning**.

Machine Learning is said as a subset of **artificial intelligence** that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own.

With the help of sample historical data, which is known as **training data**, machine learning algorithms build a **mathematical model** that helps in making predictions or decisions without being explicitly programmed. Machine learning brings computer science and statistics together for creating predictive models. Machine learning constructs or uses the algorithms that learn from historical data. The more we will provide the information, the higher will be the performance.

A machine has the ability to learn if it can improve its performance by gaining more data.

1.2.3) Fake News Classification Algorithms:

Detecting the fake news is one of the most difficult tasks for a human being. The fake news can easily be detected through the use of machine learning. There are different machine learning classifiers that can help in detecting the news is true or false. Nowadays, the dataset can easily be collected to train these classifiers. Different researchers used machine learning classifiers for checking the authenticity of news.

There are classifiers of machine learning that are used for detecting fake news. Some of these popular classifiers are given below that are used for this purpose.

Support Vector Machine:

This algorithm is mostly used for classification. This is a supervised machine learning algorithm that learns from the labeled data set. Researchers used various classifiers of machine learning and the support vector machine have given them the best results in detecting the fake news.

Naïve Bayes:

Naïve Bayes is also used for the classification tasks. This can be used to check whether the news is authentic or fake. Researchers used this classifier of machine learning to detect the false news. Logistic Regression: This classifier is used when the value to be predicted is categorical. For example, it can predict or give the result in true or false. Researchers have used this classifier to detect the news whether it is true or fake.

Random Forests:

In this classifier, there are different random forests that give a value and a value with more votes is the actual result of this classifier. In this, researchers have used different machine learning classifiers to detect the fake news. One of these classifiers is the random forest.

Recurrent Neural Network:

This classifier is also helpful for detecting the fake news. Researchers have used the recurrent neural network to classify the news as true or false. Neural Network: There are different algorithms of machine learning that are used to help in classification problems. One of these algorithms is the neural network. Researchers in (Kaliyar et al., 2020) have used the neural network to detect the fake news.

K-Nearest Neighbor:

This is a supervised algorithm of machine learning that is used for solving the classification problems. This stores the data about all the cases to classify the new case on the base of similarity. Researchers have used this classifier to detect fake news on social media.

Decision Tree:

This supervised algorithm of machine learning can help to detect the fake news. It breaks down the dataset into different smaller subsets. Researchers in (Kotteti et al., 2018) have used different machine learning classifiers and one of them is the decision tree. They have used these classifiers to detect the fake news.

Chapter: 2

Literature Survey

- In 2018 three students of Vivekananda Education Society's Institute of Technology, Mumbai published their research paper on fake news detection. They wrote in their research paper, social media age has started in 20th century. Eventually the web usage is increasing, the posts are increasing, the number of articles are increasing. They used various techniques and tool to detect fake news like NLP techniques, machine learning, and artificial intelligence.
- Facebook and WhatsApp are also working on fake news detection as they wrote in an article. They have been working for almost one year, and it is currently under the alpha phase.
- Nguyen Vo student of Ho Chi Minh City University of Technology (HCMUT) Cambodia did his research on fake news detection and implemented in 2017. He used Bi-directional GRU with Attention mechanism in his project fake news detection; Yang et al. originally proposed this mechanism. He also used some Deep learning algorithms and tried to implement other deep learning models such that AutoEncoders, GAN, CNN.

- Samir Bajaj of Stanford University published a research paper on fake news detection. He detects fake news with the help of NLP perspective and implements some other deep learning algorithm. He took an authentic data set from Signal Media News dataset.
- Rubin (2016) proposed a satire detection model with Support Vector Machine (SVM) based algorithm across 4 domains, such as science, business, soft news and civics. To verify the sources of news articles, authors have discussed various legitimate and satirical news websites. In that paper, five features were together chosen to predict the best predicting feature combination with 90% precision and 84% recall to identify satirical news which can help to minimize deception impact of satire.
- First, on a conceptual level, a distinction has been made between ’three types of fake news’ (Rubin) 2015: serious fabrications (i.e. news items about false and non-existing events or information such as celebrity gossip), hoaxes (i.e., providing false information via, for example, social media with the intention to be picked up by traditional news websites) and satire (i.e., humorous news items that mimic genuine news but contain irony and absurdity). Here, we focus on the first category, serious fabrication, in the two domains of general news (in six different

categories), as well as on celebrity gossip.

- Another attempts to differentiate satire from real news yielded promising results. The authors built a corpus of satire news (from The Onion and The Beaverton) and real news (The Toronto Star and The New York Times) in four domains (civics, science, business, soft news), resulting in a total of 240 news articles. The best classification performances were achieved with feature sets representing absurdity, punctuation, and grammar.
- Recently, a stylometric (i.e., writing-style) approach has been proposed for the identification of fake and genuine news articles (Potthast, 2017). The investigation used the Buzzfeed dataset2 of mainstream and hyperpartisan news articles of which the veracity was manually annotated. Stylometric features were, among others, character and stop word n-grams, readability indices, as well as features such as external links and the average number of words per paragraph. As a comparison, a topic-based feature set of a non-domain specific bag-of-words approach was used. The dataset used by (Potthast, 2017) consisted of 1,627 news articles that were obtainable from the original Buzzfeed dataset, including 299 fake news articles. Although the stylometric approach was promising for the classification of hyperpartisan versus mainstream

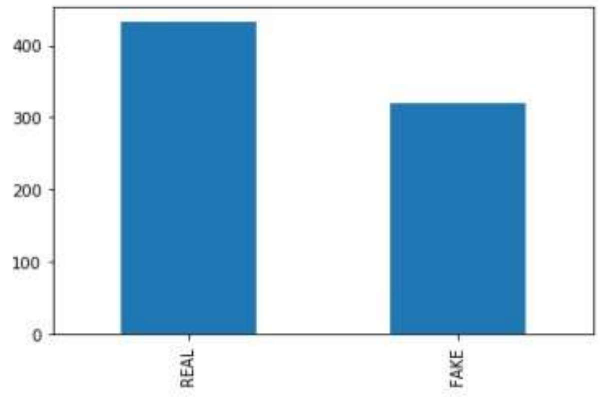
articles (accuracy: 0.75, compared to 0.71 for the topic-based feature set), both approaches were not able to differentiate fake from real news (accuracy: 0.55 and 0.52 for stylometric and topic-based feature sets, respectively).

Chapter: 3

Methodology

3.1) Data collection:

News data sets were collected manually because of the accounts and pages can post the news in different languages, and they were collected from sources that have relatively large amounts of followers and the most common news sources on Facebook because they can easily mislead or misinform the massive society like Afaan Oromo British Broadcasting Corporation (BBC), Afaan Oromo Voice of America (VOA), Afaan Oromo Fana Broadcasting Corporation (FBC), Oromia news network (ONN), Oromia Media Network (OMN), Gumaa Guddaa, Beekan Gulumma Iranaa facebook, Jawar Mohamed, Oromia Broadcasting Network (OBN) and etc. The dataset consisted of fake news from fake accounts that Facebook has been working to stamp out. News datasets were collected and properly labeled as real and fake news on similar topics to best show clear distinctions. The amount of dataset prepared for this study is relatively larger than the news dataset prepared for the Indonesian language and the Russian language but smaller when compared to the dataset for English so it requires further development. A total of 752 news datasets was used, out of these 320 were fake news articles and 432 were truthful articles as described in. Once collected, prepared and preprocessed in a required format a number of experiments were conducted for comparing model accuracy for the proper choice in developing system prototype.



3.2) Preprocessing:

Data preprocessing is done to convert the raw data into a required format. In this research, the datasets are collected from different resources that have different formats and attributes manually. Hence, the data can be duplicated and they may contain some attributes which are not useful. So, it is a must to convert the data into the required format with the required attributes that are used to train the model. To-do so different duplications and unnecessary tokens like punctuation, URL, and HTML should be removed.

3.2.1) Stemming:

A stemming algorithm refers to a step-by-step instruction that change terms to their common base form by removing off its derivational and inflectional suffixes

. Afterwords in the datasets are sleeted, normalized and tokenized, it is time to change the words into their stem. As stated Stemming is a derivation of words into their base stem, and reducing the number of classes of words in the data. While doing this the relevant document for retrieved is widen. For example, the words “hidhamtoota”, “hidhaa” and “hidhaman” “hidhan” with English corresponding meaning prisoners, prison, prisoned and they prison somebody else respectively and these all will be reduced to the word “hidh-” the stemmer used for this research is developed by the researcher.

3.3) Feature extraction:

The most important part of detecting if a given news is fake or not is to convert the news article into a news vector that contains the important features which are used to determine the nature of the news. This is one of the issues of text classification is irrelevant and repeatedly occurred features affect the accuracy and performance of the classifiers. Thus, it is best to perform feature reduction to reduce the text feature size and avoid large feature space dimension. There are several ways of feature extraction [12]. Different approaches to the same dataset are compared to determine which method gives the best accuracy. Three features extraction methods are applied in this research for comparing, namely: - Hashvectorizer, countvectorizer, and TF-IDFvectorizer. These methods are described one by one.

3.3.1) Term Frequency (TF):

Term Frequency is a method of representing the counts of words appearing in the news to calculate the similarity between the news. Described that each news that has an equal number of terms is represented by an equal length vector that contains the wordscounts. Then, each word vectors are normalized by adding one on each element. Then the result is converted into the probability of the words existing in the news. Consider, if the term is in a certain news one it will be represented as one, and if it is not in the document, it will be set to zero. Thus, each news is composed of several words. So TF represents words in the word vectors with a countvectorizer which indicates how many times the terms are there in the news. Therefore, in this research, a python module CountVectorizer from scikit-learn is used. It is used to display a table of terms that exist in the news, and its availability in the class. CountVectorizer capture the vocabulary from the news and selection of the words count features. Then, it displays a matrix with how many times a term exists for representing the news. These vectorizers were used in combination with n-grams and multinomial naïve Bayes classifiers and some preprocessing tasks.

3.3.2) Term Frequency-Inverse Document Frequency (TF-IDF):

Term Frequency-Inverse Document Frequency is also a method used to represent text in a format that can be easily processed by the machine learning algorithms. It is a numerical Afaan Oromo Text Content-Based Fake News Detection using Multinomial Naive Bayes statistic that shows how important a word is to news in a news dataset. The importance of a word is proportional to the number of times the word appears in the news (fake and real) but inversely proportional to the count of the existence of the word in the news dataset (fake or real). A problem with a term frequency approach is that the words with higher frequency becomes dominant. These words may not provide much information for the model. And due to this problem, domain-specific words that do not have a larger score may be discarded or ignored. In this Research, a computer learns, how to read and understand the differences between real news and fake news using Natural Language Processing (NLP). This is done by using TF-IDFvectorizer, countVectorizer, and hashVectorizer. TF-IDF is used to determine word importance in a given article in the entire news dataset. The frequency of the words is rescaled by considering how frequently the words occur in all the news dataset. Due to this, the scores for frequent words are also frequent among all the documents are reduced. This way of scoring is known as Term Frequency – Inverse Document Frequency. A word weight is directly proportional to how many times the words occur in the news, but, it is inversely proportional to a number of terms in the corpus. Let N represent a news corpus, let n represent a news $n \in$; a piece of news is defined as a set of terms w. Let $mw(n)$ represents how many times term w appears in news n. so, the size of news n is

$$|n| = \sum_{w \in m} m_w(n)$$

Described the TF, as term t with respect to news document n as

$$TF(W)_m = \frac{m_n(t)}{|n|}$$

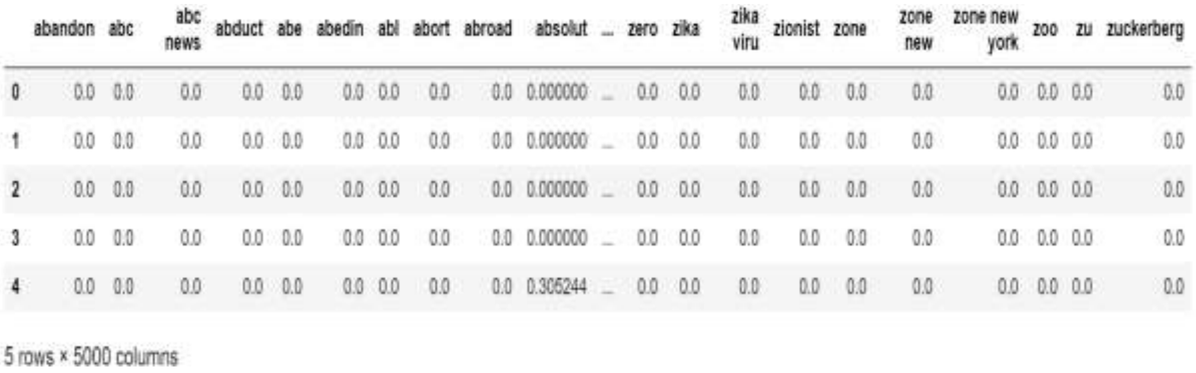
Also, present The IDF for a term t with respect to news corpus N, denoted $IDF(W)N$ is the logarithm of the total count of news in the news corpus divided by the number of news where this particular term appears and computed as follows:

$$IDF(W)_N = 1 + \log(\frac{|N|}{|\{d:N|t \in d\}|})$$

IDF is used for reducing the dominance of irrelevant words. For example, words such as “jiru” and “true” often appear in Afaan Oromo, if only TF is used, the above terms will dominate the frequency count. However, when IDF is used it scales down the impact of these words. Term Frequency –Inverted Document Frequency of the term T with respect to news n and corpus D is calculated as follows:

$$TF - IDF(T)_{nN} = TF(T)_n \times IDF(T)_N$$

The values of all text data into numeric data after applying TF-IDF vectorizer:



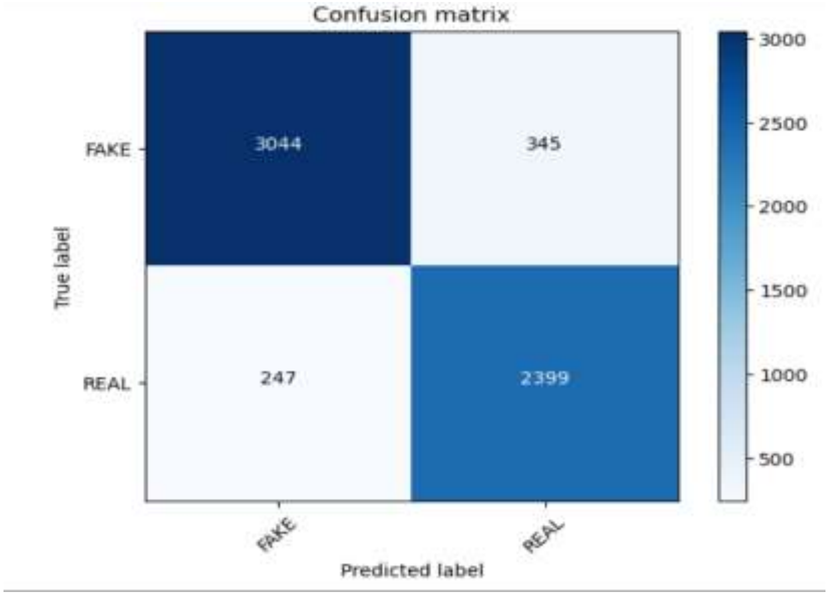
	abandon	abc	abc news	abduct	abe	abedin	abl	abort	abroad	absolut	...	zero	zika	zika viru	zionist	zone	zone new	zone new york	zoo	zu	zuckerberg
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.305244	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 5000 columns

3.3.3) MultinomialNB Algorithm:

The Multinomial Naive Bayes is one of the variants of the Naive Bayes algorithm in machine learning. It is very useful to use on a dataset that is distributed multinomially. This algorithm is especially preferred in classification tasks based on natural language processing. Spam detection is one of the applications where this algorithm can be used. If you have never used this algorithm for the machine learning problems based on classification before, this article is for you. In this article, I will take you through an introduction to the Multinomial Naive Bayes algorithm in machine learning and its implementation using Python.

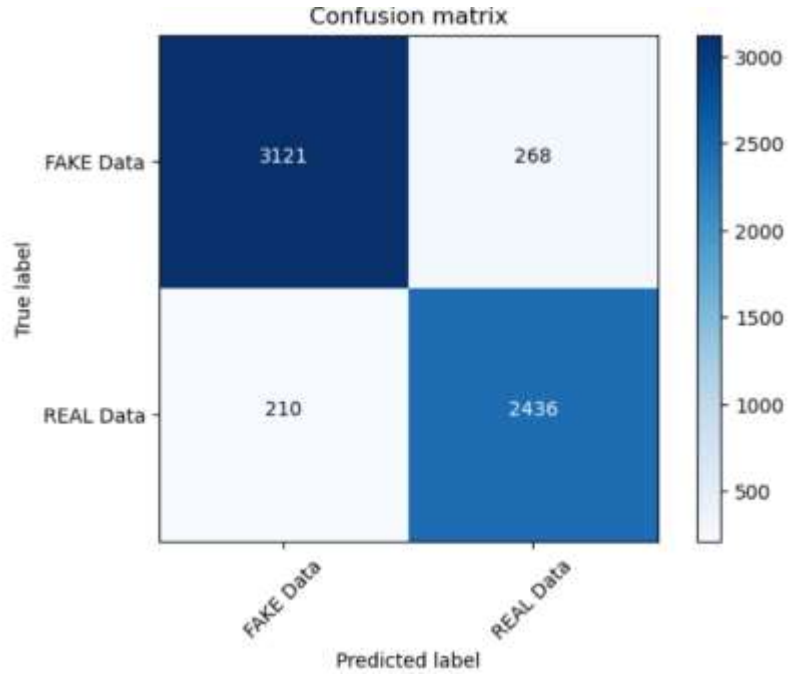
The confusion matrix by the algorithm used in our project is:



3.3.4) Passive aggressive classifier algorithm:

Passive-Aggressive algorithms are generally used for large-scale learning. It is one of the **online-learning algorithms**. In online machine learning algorithms, the input data comes in sequential order and the machine learning model is updated sequentially, as opposed to conventional batch learning, where the entire training dataset is used at once. This is very useful in situations where there is a huge amount of data, and it is computationally infeasible to train the entire dataset because of the sheer size of the data. A very good example of this would be to detect fake bulletin on a social media website like Twitter, WhatsApp where new data is being added every second. To dynamically read the data from Twitter continuously, the data would be huge, and using an online-learning algorithm would be ideal.

The confusion matrix by the algorithm used in our project is:



3.3.5) Multinomial Classifier with Hyperparameter:

The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. However, in practice, fractional counts such as TF-IDF may also work.

Chapter-4

Implementation

Fake News Classifier

Dataset: https://www.kaggle.com/c/fake-news/data#

In [1]:

```
import pandas as pd
```

In [2]:

```
df=pd.read_csv('train.csv')
```

In [3]:

```
df.head()
```

Out[3]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

In [4]:

```
## Get the Independent Features
X=df.drop('label',axis=1)
```

In [5]:

```
X.head()
```

Out[5]:

	id	title	author	text
0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...

	id	title	author	text
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...

In [6]:

```
## Get the Dependent features
y=df['label']
```

In [7]:

```
y.head()
```

Out[7]:

```
0    1
1    0
2    1
3    1
4    1
Name: label, dtype: int64
```

In [8]:

```
df.shape
```

Out[8]:

```
(20800, 5)
```

In [9]:

```
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer,
HashingVectorizer
```

In [10]:

```
df=df.dropna()
```

In [11]:

```
df.head(10)
```

Out[11]:

	id	title	author	text	label
0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0

	id	title	author	text	label
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1
5	5	Jackie Mason: Hollywood Would Love Trump if He...	Daniel Nussbaum	In these trying times, Jackie Mason is the Voi...	0
7	7	Benoît Hamon Wins French Socialist Party’s Pre...	Alissa J. Rubin	PARIS — France chose an idealistic, traditi...	0
9	9	A Back-Channel Plan for Ukraine and Russia, Co...	Megan Twohey and Scott Shane	A week before Michael T. Flynn resigned as nat...	0
10	10	Obama’s Organizing for Action Partners with So...	Aaron Klein	Organizing for Action, the activist group that...	0
11	11	BBC Comedy Sketch "Real Housewives of ISIS" Ca...	Chris Tomlinson	The BBC produced spoof on the “Real Housewives...	0

In [12]:

```
messages=df.copy()
```

In [13]:

```
messages.reset_index(inplace=True)
```

In [14]:

```
messages.head(10)
```

Out[14]:

	index	id	title	author	text	label
0	0	0	House Dem Aide: We Didn’t Even See Comey’s Let...	Darrell Lucus	House Dem Aide: We Didn’t Even See Comey’s Let...	1
1	1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0

	index	id	title	author	text	label
2	2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1
5	5	5	Jackie Mason: Hollywood Would Love Trump if He...	Daniel Nussbaum	In these trying times, Jackie Mason is the Voi...	0
6	7	7	Benoît Hamon Wins French Socialist Party’s Pre...	Alissa J. Rubin	PARIS — France chose an idealistic, traditi...	0
7	9	9	A Back-Channel Plan for Ukraine and Russia, Co...	Megan Twohey and Scott Shane	A week before Michael T. Flynn resigned as nat...	0
8	10	10	Obama’s Organizing for Action Partners with So...	Aaron Klein	Organizing for Action, the activist group that...	0
9	11	11	BBC Comedy Sketch "Real Housewives of ISIS" Ca...	Chris Tomlinson	The BBC produced spoof on the “Real Housewives...	0

In [15]:

```
messages['title'][6]
```

Out[15]:

```
'Benoît Hamon Wins French Socialist Party’s Presidential Nomination - The New
York Times'
```

In [16]:

```
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
import re
ps = PorterStemmer()
corpus = []
for i in range(0, len(messages)):
    review = re.sub('[^a-zA-Z]', ' ', messages['title'][i])
    review = review.lower()
    review = review.split()

    review = [ps.stem(word) for word in review if not word in
stopwords.words('english')]
    review = ' '.join(review)
    corpus.append(review)
```

In [17]:

```
corpus[3]
```

Out[17]:

```
'civilian kill singl us airstrik identifi'
```

In [18]:

```
## Applying Countvectorizer
# Creating the Bag of Words model
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features=5000,ngram_range=(1,3))
X = cv.fit_transform(corpus).toarray()
```

In [19]:

```
X.shape
```

Out[19]:

```
(18285, 5000)
```

In [20]:

```
y=messages['label']
```

In [21]:

```
## Divide the dataset into Train and Test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
random_state=0)
```

In [ ]:

In [22]:

```
cv.get_params()
```

Out[22]:

```
{'analyzer': 'word',
 'binary': False,
 'decode_error': 'strict',
 'dtype': numpy.int64,
 'encoding': 'utf-8',
 'input': 'content',
 'lowercase': True,
 'max_df': 1.0,
 'max_features': 5000,
 'min_df': 1,
 'ngram_range': (1, 3),
 'preprocessor': None,
 'stop_words': None,
 'strip_accents': None,
 'token_pattern': '(?u)\\b\\w\\w+\\b',
 'tokenizer': None,
 'vocabulary': None}
```

In [23]:

```
count_df = pd.DataFrame(X_train, columns=cv.get_feature_names_out())
```

In [24]:

```
count_df.head()
```

Out[24]:

	abandon	abc	abcnews	abduct	abe	abedin	abl	abort	abroad	absolut	...	zero	zika	zikaviru	zionist	zone	zonenew	zonenewyork	zoo	zu	zuckerberg
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	0	0	0	0	0

5 rows × 5000 columns

In [25]:

```
import matplotlib.pyplot as plt
```

In [26]:

```
def plot_confusion_matrix(cm, classes,
                          normalize=False,
                          title='Confusion matrix',
                          cmap=plt.cm.Blues):

    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    if normalize:
        cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
        print("Normalized confusion matrix")
    else:
        print('Confusion matrix, without normalization')

    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, cm[i, j],
                 horizontalalignment="center",
                 color="white" if cm[i, j] > thresh else "black")

    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')
```

MultinomialNB Algorithm

In [27]:

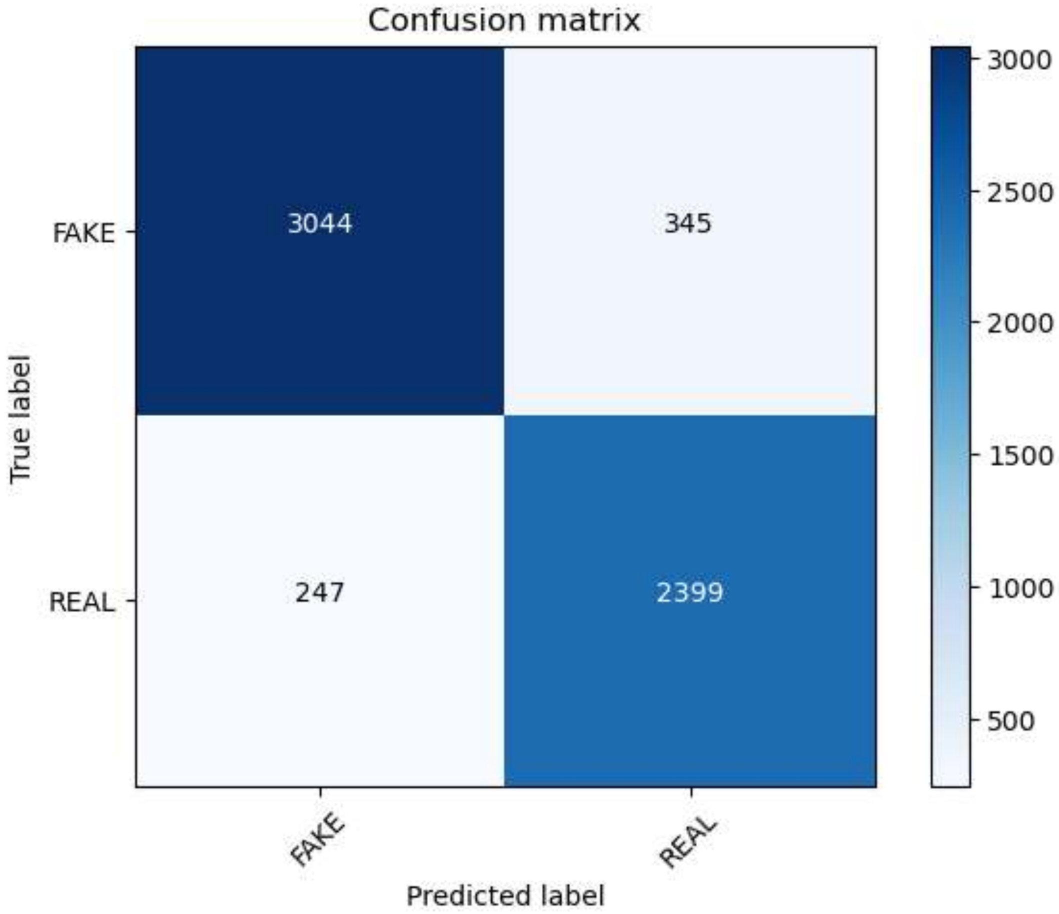
```
from sklearn.naive_bayes import MultinomialNB
classifier=MultinomialNB()
```

In [28]:

```
from sklearn import metrics
import numpy as np
import itertools
```

In [29]:

```
classifier.fit(X_train, y_train)
pred = classifier.predict(X_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy:   %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred)
plot_confusion_matrix(cm, classes=['FAKE', 'REAL'])
accuracy:   0.902
Confusion matrix, without normalization
```



In [30]:

```
classifier.fit(X_train, y_train)
pred = classifier.predict(X_test)
score = metrics.accuracy_score(y_test, pred)
score
```

Out[30]:

```
0.9019055509527755
```

In [31]:

```
y_train.shape
```

Out[31]:

```
(12250,)
```

Passive Aggressive Classifier Algorithm

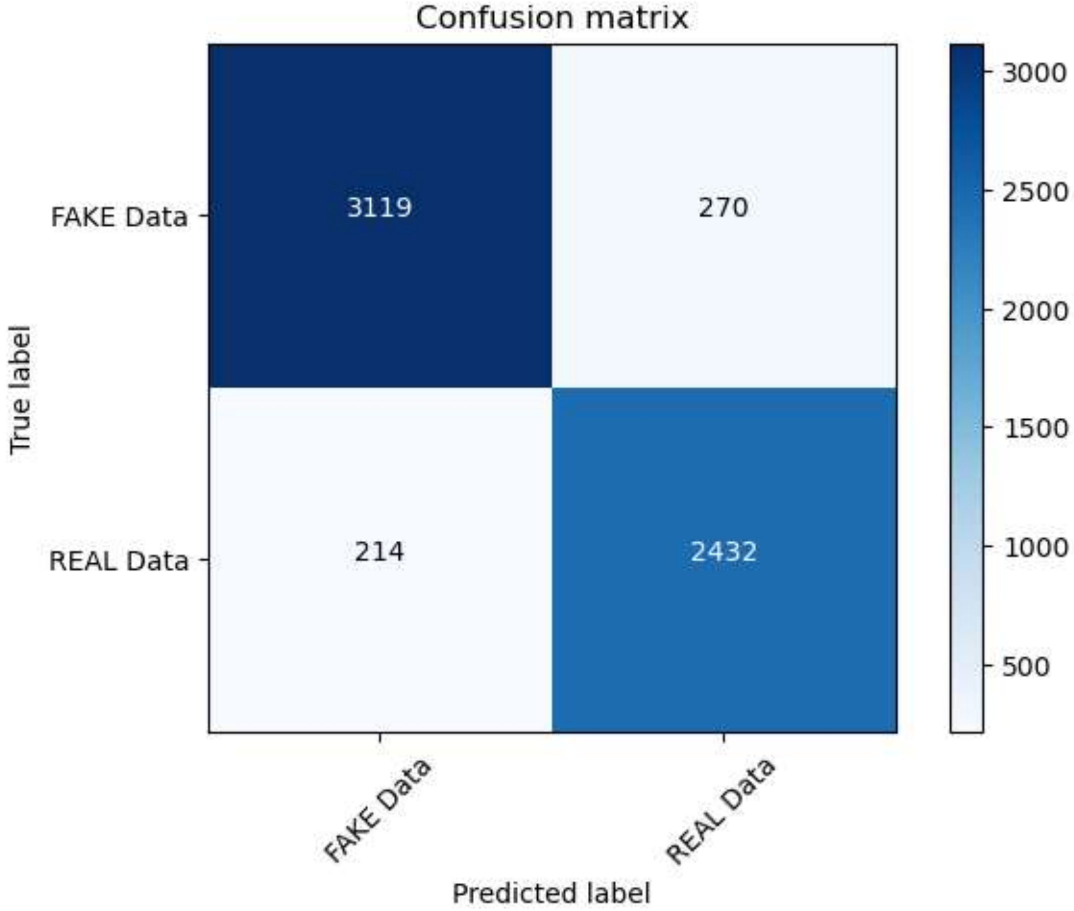
In [32]:

```
from sklearn.linear_model import PassiveAggressiveClassifier
linear_clf = PassiveAggressiveClassifier(max_iter=50)
```

In [33]:

```
linear_clf.fit(X_train, y_train)
pred = linear_clf.predict(X_test)
score = metrics.accuracy_score(y_test, pred)
print("accuracy:   %0.3f" % score)
cm = metrics.confusion_matrix(y_test, pred)
plot_confusion_matrix(cm, classes=['FAKE Data', 'REAL Data'])
```

```
accuracy:   0.920
Confusion matrix, without normalization
```



Multinomial Classifier with Hyperparameter

In [34]:

```
classifier=MultinomialNB(alpha=0.1)
```

In [35]:

```
previous_score=0
for alpha in np.arange(0,1,0.1):
    sub_classifier=MultinomialNB(alpha=alpha)
    sub_classifier.fit(X_train,y_train)
    y_pred=sub_classifier.predict(X_test)
    score = metrics.accuracy_score(y_test, y_pred)
    if score>previous_score:
        classifier=sub_classifier
    print("Alpha: {}, Score : {}".format(alpha,score))
```

```
C:\Users\ultra\anaconda3\anaconda\lib\site-packages\sklearn\naive_bayes.py:629
: FutureWarning: The default value for `force_alpha` will change to `True` in
1.4. To suppress this warning, manually set the value of `force_alpha`.
  warnings.warn(
C:\Users\ultra\anaconda3\anaconda\lib\site-packages\sklearn\naive_bayes.py:635
: UserWarning: alpha too small will result in numeric errors, setting alpha =
1.0e-10. Use `force_alpha=True` to keep alpha unchanged.
  warnings.warn(
Alpha: 0.0, Score : 0.8903065451532726
Alpha: 0.1, Score : 0.9020712510356255
Alpha: 0.2, Score : 0.9025683512841757
Alpha: 0.30000000000000004, Score : 0.9024026512013256
Alpha: 0.4, Score : 0.9017398508699255
Alpha: 0.5, Score : 0.9015741507870754
Alpha: 0.6000000000000001, Score : 0.9022369511184756
Alpha: 0.7000000000000001, Score : 0.9025683512841757
Alpha: 0.8, Score : 0.9015741507870754
Alpha: 0.9, Score : 0.9017398508699255
```

In [36]:

```
## Get Features names
feature_names = cv.get_feature_names_out()
```

In [37]:

```
classifier.feature_log_prob_[0]
```

Out[37]:

```
array([ -9.06056227,  -9.06056227, -10.28838629, ...,  -9.99212048,
       -11.45845755,  -9.42157562])
```

In [38]:

```
### Most real
sorted(zip(classifier.feature_log_prob_[0], feature_names), reverse=True)[:20]
```

Out[38]:

```
[(-2.9468577463990755, 'new'),
 (-2.994219848520549, 'time'),
 (-3.000566240637532, 'york'),
 (-3.0008020674474167, 'new york'),
 (-3.014815217142134, 'york time'),
 (-3.014815217142134, 'new york time'),
 (-3.9648310157438633, 'breitbart'),
 (-4.004573851696944, 'trump'),
 (-5.2756841712452855, 'donald'),
 (-5.282590276041697, 'donald trump'),
 (-5.755045510465673, 'say'),
 (-5.992274458590198, 'obama'),
 (-6.016039835625665, 'clinton'),
 (-6.106599412671392, 'presid'),
 (-6.122523316028115, 'state'),
 (-6.177512429323892, 'report'),
 (-6.188882648465076, 'attack'),
 (-6.253840544739848, 'hous'),
 (-6.259960514881633, 'brief'),
 (-6.316793989644799, 'hillari')]
```

In [39]:

```
### Most fake
sorted(zip(classifier.feature_log_prob_[0], feature_names))[:5000]
```

Out[39]:

```
[(-11.458457546147459, 'access pipelin protest'),
 (-11.458457546147459, 'acknowledg emf'),
 (-11.458457546147459, 'acknowledg emf damag'),
 (-11.458457546147459, 'acquit'),
 (-11.458457546147459, 'acr'),
 (-11.458457546147459, 'adhd'),
 (-11.458457546147459, 'airstrik kill'),
 (-11.458457546147459, 'al nusra'),
 (-11.458457546147459, 'america last'),
 (-11.458457546147459, 'america vote'),
 (-11.458457546147459, 'american concern'),
 (-11.458457546147459, 'american concern elect'),
 (-11.458457546147459, 'american peopl defeat'),
 (-11.458457546147459, 'aqsa'),
 (-11.458457546147459, 'arabian'),
 (-11.458457546147459, 'ariel noyola'),
 (-11.458457546147459, 'ariel noyola rodr'),
 (-11.458457546147459, 'ask question'),
 (-11.458457546147459, 'auf'),
 (-11.458457546147459, 'avail'),
 (-11.458457546147459, 'babi powder'),
 (-11.458457546147459, 'bad news'),
 (-11.458457546147459, 'badg'),
 (-11.458457546147459, 'ballot'),
 (-11.458457546147459, 'baltic'),
 (-11.458457546147459, 'banana'),
 (-11.458457546147459, 'banana republ'),
 (-11.458457546147459, 'barack obama delay'),
 (-11.458457546147459, 'beg'),
 (-11.458457546147459, 'behind trump'),
 (-11.458457546147459, 'bewar'),
 (-11.458457546147459, 'bias'),
 (-11.458457546147459, 'big pharma'),
 (-11.458457546147459, 'bishop'),
 (-11.458457546147459, 'black agenda'),
 (-11.458457546147459, 'bombshel'),
 (-11.458457546147459, 'bonus'),
 (-11.458457546147459, 'bottom'),
 (-11.458457546147459, 'break fbi'),
 (-11.458457546147459, 'break trump'),
 (-11.458457546147459, 'bribe'),
 (-11.458457546147459, 'buildup'),
 (-11.458457546147459, 'cafe open'),
 (-11.458457546147459, 'cafe open thread'),
 (-11.458457546147459, 'camp polic'),
 (-11.458457546147459, 'campaign chair'),
 (-11.458457546147459, 'campaign tri'),
 (-11.458457546147459, 'campaign tri hack'),
 (-11.458457546147459, 'cancer cell'),
 (-11.458457546147459, 'cannabi'),
 (-11.458457546147459, 'cartoon'),
 (-11.458457546147459, 'case testimoni'),
 (-11.458457546147459, 'case testimoni pennsylvania'),
 (-11.458457546147459, 'caught tape'),
 (-11.458457546147459, 'cdc'),
 (-11.458457546147459, 'chaffetz'),
 (-11.458457546147459, 'charg obstruct'),
 (-11.458457546147459, 'charg obstruct justic'),
 (-11.458457546147459, 'chart day'),
 (-11.458457546147459, 'christ'),
 (-11.458457546147459, 'civil unrest'),
 (-11.458457546147459, 'civil unrest survey'),
 (-11.458457546147459, 'clinton camp'),
 (-11.458457546147459, 'clinton campaign chair'),
 (-11.458457546147459, 'clinton crime'),
 (-11.458457546147459, 'clinton crime famili'),
 (-11.458457546147459, 'clinton elect'),
 (-11.458457546147459, 'clinton inc'),
 (-11.458457546147459, 'clinton presid'),
 (-11.458457546147459, 'clinton propos'),
 (-11.458457546147459, 'clinton propos rig'),
 (-11.458457546147459, 'clinton vote'),
 (-11.458457546147459, 'clown'),
 (-11.458457546147459, 'coconut'),
 (-11.458457546147459, 'codesod'),
 (-11.458457546147459, 'collect evolut'),
 (-11.458457546147459, 'color revolut'),
 (-11.458457546147459, 'comment hillari'),
 (-11.458457546147459, 'comment hillari clinton'),
 (-11.458457546147459, 'commiss one'),
 (-11.458457546147459, 'commiss one want'),
 (-11.458457546147459, 'concern elect'),
 (-11.458457546147459, 'concern elect violenc'),
 (-11.458457546147459, 'conscious'),
 (-11.458457546147459, 'conserv daili'),
 (-11.458457546147459, 'conserv daili post'),
 (-11.458457546147459, 'contrarian'),
 (-11.458457546147459, 'contrarian read'),
 (-11.458457546147459, 'cooler'),
 (-11.458457546147459, 'coordin'),
 (-11.458457546147459, 'costum'),
 (-11.458457546147459, 'could go prison'),
 (-11.458457546147459, 'craig robert'),
 (-11.458457546147459, 'craze'),
 (-11.458457546147459, 'creamer'),
 (-11.458457546147459, 'creepi'),
 (-11.458457546147459, 'crook hillari'),
 (-11.458457546147459, 'crop'),
 (-11.458457546147459, 'crucial'),
 (-11.458457546147459, 'cult'),
 (-11.458457546147459, 'daesh'),
 (-11.458457546147459, 'daili contrarian'),
 (-11.458457546147459, 'daili contrarian read'),
 (-11.458457546147459, 'daili post'),
 (-11.458457546147459, 'damn'),
 (-11.458457546147459, 'dapl'),
 (-11.458457546147459, 'dapl protest'),
 (-11.458457546147459, 'day elect'),
 (-11.458457546147459, 'de la'),
 (-11.458457546147459, 'de lo'),
 (-11.458457546147459, 'dea'),
 (-11.458457546147459, 'debbi menon'),
 (-11.458457546147459, 'decor'),
 (-11.458457546147459, 'defeat oligarchi'),
 (-11.458457546147459, 'defeat oligarchi rule'),
 (-11.458457546147459, 'del'),
 (-11.458457546147459, 'delay suspend'),
 (-11.458457546147459, 'delay suspend elect'),
 (-11.458457546147459, 'demonetis'),
 (-11.458457546147459, 'den'),
 (-11.458457546147459, 'der'),
 (-11.458457546147459, 'design fail'),
 (-11.458457546147459, 'destroy trump'),
 (-11.458457546147459, 'devast'),
 (-11.458457546147459, 'diabet'),
 (-11.458457546147459, 'dinucci'),
 (-11.458457546147459, 'director comey'),
 (-11.458457546147459, 'disabl'),
 (-11.458457546147459, 'disgust'),
 (-11.458457546147459, 'disord'),
 (-11.458457546147459, 'dr david'),
 (-11.458457546147459, 'dr david duke'),
 (-11.458457546147459, 'dr duke'),
 (-11.458457546147459, 'dr eowyn'),
 (-11.458457546147459, 'duck'),
 (-11.458457546147459, 'eastern outlook'),
 (-11.458457546147459, 'ein'),
 (-11.458457546147459, 'elect fraud'),
 (-11.458457546147459, 'elect hillari'),
 (-11.458457546147459, 'elect hillari forc'),
 (-11.458457546147459, 'elect night'),
 (-11.458457546147459, 'elect paul'),
 (-11.458457546147459, 'elect paul craig'),
 (-11.458457546147459, 'elect rig'),
 (-11.458457546147459, 'elect video'),
 (-11.458457546147459, 'elect violenc'),
 (-11.458457546147459, 'electr'),
 (-11.458457546147459, 'electron vote'),
 (-11.458457546147459, 'electron vote machin'),
 (-11.458457546147459, 'email clinton'),
 (-11.458457546147459, 'email found'),
 (-11.458457546147459, 'email reveal'),
 (-11.458457546147459, 'emf'),
 (-11.458457546147459, 'emf damag'),
 (-11.458457546147459, 'en el'),
 (-11.458457546147459, 'end hillari'),
 (-11.458457546147459, 'end hillari clinton'),
 (-11.458457546147459, 'entertain'),
 (-11.458457546147459, 'eowyn'),
 (-11.458457546147459, 'eras'),
 (-11.458457546147459, 'euro'),
 (-11.458457546147459, 'everywher'),
 (-11.458457546147459, 'expos hillari'),
 (-11.458457546147459, 'fame star'),
 (-11.458457546147459, 'farm grow'),
 (-11.458457546147459, 'fascin'),
 (-11.458457546147459, 'fbi agent'),
 (-11.458457546147459, 'fbi clinton'),
 (-11.458457546147459, 'fbi director comey'),
 (-11.458457546147459, 'fbi email'),
 (-11.458457546147459, 'fbi email investig'),
 (-11.458457546147459, 'fbi investig'),
 (-11.458457546147459, 'fbi reopen'),
 (-11.458457546147459, 'fbi reopen hillari'),
 (-11.458457546147459, 'fbi reopen investig'),
 (-11.458457546147459, 'fda'),
 (-11.458457546147459, 'feast cafe'),
 (-11.458457546147459, 'feast cafe open'),
 (-11.458457546147459, 'find american concern'),
 (-11.458457546147459, 'fli zone'),
 (-11.458457546147459, 'flu'),
 (-11.458457546147459, 'fluorid'),
 (-11.458457546147459, 'forc new fbi'),
 (-11.458457546147459, 'former congressman'),
 (-11.458457546147459, 'freedom rider'),
 (-11.458457546147459, 'fuck'),
 (-11.458457546147459, 'furiou'),
 (-11.458457546147459, 'get readi civil'),
 (-11.458457546147459, 'gmo'),
 (-11.458457546147459, 'go prison year'),
 (-11.458457546147459, 'gold silver'),
 (-11.458457546147459, 'grab musket'),
 (-11.458457546147459, 'gruber'),
 (-11.458457546147459, 'guardian liberti'),
 (-11.458457546147459, 'guardian liberti voic'),
 (-11.458457546147459, 'guez'),
 (-11.458457546147459, 'hack wikileak'),
 (-11.458457546147459, 'health benefit'),
 (-11.458457546147459, 'herb'),
 (-11.458457546147459, 'hillari campaign'),
 (-11.458457546147459, 'hillari clinton charg'),
 (-11.458457546147459, 'hillari clinton propos'),
 (-11.458457546147459, 'hillari clinton vote'),
 (-11.458457546147459, 'hillari elect'),
 (-11.458457546147459, 'hillari forc'),
 (-11.458457546147459, 'hillari forc new'),
 (-11.458457546147459, 'hillari investig'),
 (-11.458457546147459, 'hillari support'),
 (-11.458457546147459, 'hillari win'),
 (-11.458457546147459, 'hodg'),
 (-11.458457546147459, 'homeless woman'),
 (-11.458457546147459, 'horrifi'),
 (-11.458457546147459, 'human right council'),
 (-11.458457546147459, 'iceland'),
 (-11.458457546147459, 'illeg alien muslim'),
 (-11.458457546147459, 'illuminati'),
 (-11.458457546147459, 'im'),
 (-11.458457546147459, 'imperi'),
 (-11.458457546147459, 'implod'),
 (-11.458457546147459, 'incred'),
 (-11.458457546147459, 'investig hillari'),
 (-11.458457546147459, 'investig hillari clinton'),
 (-11.458457546147459, 'iraqi armi'),
 (-11.458457546147459, 'iraqi troop'),
 (-11.458457546147459, 'ist'),
 (-11.458457546147459, 'iv'),
 (-11.458457546147459, 'jason chaffetz'),
 (-11.458457546147459, 'johnson johnson'),
 (-11.458457546147459, 'jungl'),
 (-11.458457546147459, 'kiev'),
 (-11.458457546147459, 'kill cancer'),
 (-11.458457546147459, 'kuznetsov'),
 (-11.458457546147459, 'leader acquit'),
 (-11.458457546147459, 'leak audio'),
 (-11.458457546147459, 'leak email'),
 (-11.458457546147459, 'let go'),
 (-11.458457546147459, 'let go new'),
 (-11.458457546147459, 'lib'),
 (-11.458457546147459, 'liberti voic'),
 (-11.458457546147459, 'liberti writer'),
 (-11.458457546147459, 'liberti writer news'),
 (-11.458457546147459, 'live blog'),
 (-11.458457546147459, 'live tv'),
 (-11.458457546147459, 'lose grab'),
 (-11.458457546147459, 'lose grab musket'),
 (-11.458457546147459, 'malheur'),
 (-11.458457546147459, 'malheur wildlif'),
 (-11.458457546147459, 'malheur wildlif refug'),
 (-11.458457546147459, 'manlio'),
 (-11.458457546147459, 'manlio dinucci'),
 (-11.458457546147459, 'materi'),
 (-11.458457546147459, 'matrix'),
 (-11.458457546147459, 'medit'),
 (-11.458457546147459, 'menon'),
 (-11.458457546147459, 'meter case'),
 (-11.458457546147459, 'meter case testimoni'),
 (-11.458457546147459, 'meyssan'),
 (-11.458457546147459, 'militar polic'),
 (-11.458457546147459, 'monetari'),
 (-11.458457546147459, 'moveabl'),
 (-11.458457546147459, 'moveabl feast'),
 (-11.458457546147459, 'moveabl feast cafe'),
 (-11.458457546147459, 'musket'),
 (-11.458457546147459, 'muslim invad'),
 (-11.458457546147459, 'muslim migrant'),
 (-11.458457546147459, 'muslim women'),
 (-11.458457546147459, 'neocon'),
 (-11.458457546147459, 'new clinton'),
 (-11.458457546147459, 'new clinton email'),
 (-11.458457546147459, 'new eastern'),
 (-11.458457546147459, 'new eastern outlook'),
 (-11.458457546147459, 'new fbi email'),
 (-11.458457546147459, 'new moon'),
 (-11.458457546147459, 'new report'),
 (-11.458457546147459, 'new world'),
 (-11.458457546147459, 'new world order'),
 (-11.458457546147459, 'newstick'),
 (-11.458457546147459, 'nov'),
 (-11.458457546147459, 'novemb daili'),
 (-11.458457546147459, 'novemb daili contrarian'),
 (-11.458457546147459, 'nuke'),
 (-11.458457546147459, 'nusra'),
 (-11.458457546147459, 'obama clinton'),
 (-11.458457546147459, 'obamacar design'),
 (-11.458457546147459, 'oct'),
 (-11.458457546147459, 'octob surpris'),
 (-11.458457546147459, 'oligarch'),
 (-11.458457546147459, 'oligarchi'),
 (-11.458457546147459, 'oligarchi rule'),
 (-11.458457546147459, 'one want'),
 (-11.458457546147459, 'one want acknowledg'),
 (-11.458457546147459, 'open thread'),
 (-11.458457546147459, 'opinion conserv'),
 (-11.458457546147459, 'oregon standoff'),
 (-11.458457546147459, 'os'),
 (-11.458457546147459, 'overnight'),
 (-11.458457546147459, 'palestin elect'),
 (-11.458457546147459, 'paper ballot'),
 (-11.458457546147459, 'par'),
 (-11.458457546147459, 'para'),
 (-11.458457546147459, 'passag'),
 (-11.458457546147459, 'patholog'),
 (-11.458457546147459, 'patriot act'),
 (-11.458457546147459, 'paul craig'),
 (-11.458457546147459, 'paul craig robert'),
 (-11.458457546147459, 'paul ryan rel'),
 (-11.458457546147459, 'pay play'),
 (-11.458457546147459, 'pennsylvania public'),
 (-11.458457546147459, 'pennsylvania public util'),
 (-11.458457546147459, 'peopl defeat'),
 (-11.458457546147459, 'peopl defeat oligarchi'),
 (-11.458457546147459, 'percentfedup'),
 (-11.458457546147459, 'percentfedup com'),
 (-11.458457546147459, 'peter thiel'),
 (-11.458457546147459, 'pharma'),
 (-11.458457546147459, 'physicist'),
 (-11.458457546147459, 'pilger'),
 (-11.458457546147459, 'pimp'),
 (-11.458457546147459, 'piss'),
 (-11.458457546147459, 'pm water'),
 (-11.458457546147459, 'pm water cooler'),
 (-11.458457546147459, 'podesta email'),
 (-11.458457546147459, 'podestaemail'),
 (-11.458457546147459, 'poison'),
 (-11.458457546147459, 'por'),
 (-11.458457546147459, 'powder'),
 (-11.458457546147459, 'predict trump'),
 (-11.458457546147459, 'presstitut'),
 (-11.458457546147459, 'primer'),
 (-11.458457546147459, 'project verita'),
 (-11.458457546147459, 'project verita video'),
 (-11.458457546147459, 'propos rig'),
 (-11.458457546147459, 'protector'),
 (-11.458457546147459, 'public util'),
 (-11.458457546147459, 'public util commiss'),
 (-11.458457546147459, 'puppet'),
 (-11.458457546147459, 'que'),
 (-11.458457546147459, 'ransom'),
 (-11.458457546147459, 'rather'),
 (-11.458457546147459, 'readi civil'),
 (-11.458457546147459, 'readi civil unrest'),
 (-11.458457546147459, 'rebuild'),
 (-11.458457546147459, 'regim chang'),
 (-11.458457546147459, 'remedi'),
 (-11.458457546147459, 'reopen hillari'),
 (-11.458457546147459, 'report novemb'),
 (-11.458457546147459, 'reveal clinton'),
 (-11.458457546147459, 'reveal hillari'),
 (-11.458457546147459, 'rider'),
 (-11.458457546147459, 'rig elect'),
 (-11.458457546147459, 'rig palestin'),
 (-11.458457546147459, 'rig palestin elect'),
 (-11.458457546147459, 'rodr'),
 (-11.458457546147459, 'rodr guez'),
 (-11.458457546147459, 'ron paul'),
 (-11.458457546147459, 'russian border'),
 (-11.458457546147459, 'russophobia'),
 (-11.458457546147459, 'ryan rel'),
 (-11.458457546147459, 'saker'),
 (-11.458457546147459, 'saudi arabian'),
 (-11.458457546147459, 'se'),
 (-11.458457546147459, 'selfi'),
 (-11.458457546147459, 'shadow govern'),
 (-11.458457546147459, 'shock video'),
 (-11.458457546147459, 'shocker'),
 (-11.458457546147459, 'show clinton'),
 (-11.458457546147459, 'sich'),
 (-11.458457546147459, 'sie'),
 (-11.458457546147459, 'signific'),
 (-11.458457546147459, 'sinc cold'),
 (-11.458457546147459, 'sinc cold war'),
 (-11.458457546147459, 'sioux'),
 (-11.458457546147459, 'smart meter'),
 (-11.458457546147459, 'smart meter case'),
 (-11.458457546147459, 'steal elect'),
 (-11.458457546147459, 'steal trump'),
 (-11.458457546147459, 'surfac'),
 (-11.458457546147459, 'survey find american'),
 (-11.458457546147459, 'suspend elect'),
 (-11.458457546147459, 'suspend elect hillari'),
 (-11.458457546147459, 'switch trump'),
 (-11.458457546147459, 'switch vote'),
 (-11.458457546147459, 'syrian war'),
 (-11.458457546147459, 'syrian war report'),
 (-11.458457546147459, 'testimoni pennsylvania'),
 (-11.458457546147459, 'testimoni pennsylvania public'),
 (-11.458457546147459, 'thiel'),
 (-11.458457546147459, 'thierri'),
 (-11.458457546147459, 'thierri meyssan'),
 (-11.458457546147459, 'thousand wild'),
 (-11.458457546147459, 'thread'),
 (-11.458457546147459, 'thug'),
 (-11.458457546147459, 'timelin'),
 (-11.458457546147459, 'tonight'),
 (-11.458457546147459, 'tri hack'),
 (-11.458457546147459, 'tri hack wikileak'),
 (-11.458457546147459, 'trophi'),
 (-11.458457546147459, 'trump lose grab'),
 (-11.458457546147459, 'trump need'),
 (-11.458457546147459, 'trump star'),
 (-11.458457546147459, 'trump truthfe'),
 (-11.458457546147459, 'trump video'),
 (-11.458457546147459, 'trump vote'),
 (-11.458457546147459, 'truthfe'),
 (-11.458457546147459, 'tu'),
 (-11.458457546147459, 'ufo'),
 (-11.458457546147459, 'una'),
 (-11.458457546147459, 'unesco'),
 (-11.458457546147459, 'unrest survey'),
 (-11.458457546147459, 'unrest survey find'),
 (-11.458457546147459, 'us airstrik'),
 (-11.458457546147459, 'us elect'),
 (-11.458457546147459, 'us intellig'),
 (-11.458457546147459, 'us militari'),
 (-11.458457546147459, 'us presid'),
 (-11.458457546147459, 'usapoliticsnow'),
 (-11.458457546147459, 'util commiss'),
 (-11.458457546147459, 'util commiss one'),
 (-11.458457546147459, 'verita'),
 (-11.458457546147459, 'verita video'),
 (-11.458457546147459, 'vertic'),
 (-11.458457546147459, 'vertic farm'),
 (-11.458457546147459, 'vertic farm grow'),
 (-11.458457546147459, 'video expos'),
 (-11.458457546147459, 'video trump'),
 (-11.458457546147459, 'vineyard saker'),
 (-11.458457546147459, 'vote clinton'),
 (-11.458457546147459, 'vote warmong'),
 (-11.458457546147459, 'vote warmong hillari'),
 (-11.458457546147459, 'vote world'),
 (-11.458457546147459, 'vote world war'),
 (-11.458457546147459, 'voter suppress'),
 (-11.458457546147459, 'walk fame'),
 (-11.458457546147459, 'want acknowledg'),
 (-11.458457546147459, 'want acknowledg emf'),
 (-11.458457546147459, 'want know'),
 (-11.458457546147459, 'war iii'),
 (-11.458457546147459, 'war report'),
 (-11.458457546147459, 'war report novemb'),
 (-11.458457546147459, 'war syria'),
 (-11.458457546147459, 'warmong'),
 (-11.458457546147459, 'warmong hillari'),
 (-11.458457546147459, 'warmong hillari clinton'),
 (-11.458457546147459, 'watch hillari'),
 (-11.458457546147459, 'water cooler'),
 (-11.458457546147459, 'water protector'),
 (-11.458457546147459, 'weed'),
 (-11.458457546147459, 'wikileak hillari'),
 (-11.458457546147459, 'win elect'),
 (-11.458457546147459, 'wire transfer'),
 (-11.458457546147459, 'wolf'),
 (-11.458457546147459, 'world first'),
 (-11.458457546147459, 'world order'),
 (-11.458457546147459, 'world war iii'),
 (-11.458457546147459, 'would look'),
 (-11.458457546147459, 'wreck'),
 (-11.458457546147459, 'writer news'),
 (-11.458457546147459, 'ww'),
 (-11.458457546147459, 'wwii'),
 (-11.458457546147459, 'yo'),
 (-11.458457546147459, 'zionist'),
 (-11.458457546147459, 'zu'),
 (-10.711243144317237, 'abstain'),
 (-10.711243144317237, 'access pipelin'),
 (-10.711243144317237, 'accus trump'),
 (-10.711243144317237, 'african american'),
 (-10.711243144317237, 'akbar'),
 (-10.711243144317237, 'alcohol'),
 (-10.711243144317237, 'alert'),
 (-10.711243144317237, 'allahu'),
 (-10.711243144317237, 'allahu akbar'),
 (-10.711243144317237, 'amaz'),
 (-10.711243144317237, 'ambush'),
 (-10.711243144317237, 'american peopl'),
 (-10.711243144317237, 'amnesti'),
 (-10.711243144317237, 'ap'),
 (-10.711243144317237, 'architect'),
 (-10.711243144317237, 'arctic'),
 (-10.711243144317237, 'ariel'),
 (-10.711243144317237, 'assembl'),
 (-10.711243144317237, 'audio hillari'),
 (-10.711243144317237, 'audio hillari clinton'),
 (-10.711243144317237, 'avert'),
 (-10.711243144317237, 'awesom'),
 (-10.711243144317237, 'balanc'),
 (-10.711243144317237, 'birthday'),
 (-10.711243144317237, 'black voter'),
 (-10.711243144317237, 'blackout'),
 (-10.711243144317237, 'blm'),
 (-10.711243144317237, 'blog'),
 (-10.711243144317237, 'boast'),
 (-10.711243144317237, 'boost'),
 (-10.711243144317237, 'brick'),
 (-10.711243144317237, 'brink'),
 (-10.711243144317237, 'bro'),
 (-10.711243144317237, 'buffalo'),
 (-10.711243144317237, 'bug'),
 (-10.711243144317237, 'bundi'),
 (-10.711243144317237, 'bundi brother'),
 (-10.711243144317237, 'calai'),
 (-10.711243144317237, 'call end'),
 (-10.711243144317237, 'call hillari'),
 (-10.711243144317237, 'cam'),
 (-10.711243144317237, 'campaign promis'),
 (-10.711243144317237, 'cat'),
 (-10.711243144317237, 'cave'),
 (-10.711243144317237, 'central bank'),
 (-10.711243144317237, 'charter'),
 (-10.711243144317237, 'clinton email investig'),
 (-10.711243144317237, 'clinton support'),
 (-10.711243144317237, 'cohen'),
 (-10.711243144317237, 'colon'),
 (-10.711243144317237, 'com'),
 (-10.711243144317237, 'comey letter'),
 (-10.711243144317237, 'conced'),
 (-10.711243144317237, 'conduct'),
 (-10.711243144317237, 'congratul'),
 (-10.711243144317237, 'cop killer'),
 (-10.711243144317237, 'creation'),
 (-10.711243144317237, 'crew'),
 (-10.711243144317237, 'crime famili'),
 (-10.711243144317237, 'crook'),
 (-10.711243144317237, 'dakota access'),
 (-10.711243144317237, 'dakota access pipelin'),
 (-10.711243144317237, 'david duke'),
 (-10.711243144317237, 'decept'),
 (-10.711243144317237, 'dementia'),
 (-10.711243144317237, 'demon'),
 (-10.711243144317237, 'demonstr'),
 (-10.711243144317237, 'descend'),
 (-10.711243144317237, 'despair'),
 (-10.711243144317237, 'deterior'),
 (-10.711243144317237, 'deutsch'),
 (-10.711243144317237, 'devot'),
 (-10.711243144317237, 'discoveri'),
 (-10.711243144317237, 'document reveal'),
 (-10.711243144317237, 'donna brazil'),
 (-10.711243144317237, 'eastern'),
 (-10.711243144317237, 'economist'),
 (-10.711243144317237, 'either'),
 (-10.711243144317237, 'elderli'),
 (-10.711243144317237, 'electron'),
 (-10.711243144317237, 'email case'),
 (-10.711243144317237, 'email investig'),
 (-10.711243144317237, 'email scandal'),
 (-10.711243144317237, 'endors donald'),
 (-10.711243144317237, 'endors donald trump'),
 (-10.711243144317237, 'es'),
 (-10.711243144317237, 'et'),
 (-10.711243144317237, 'ethnic'),
 (-10.711243144317237, 'extinct'),
 (-10.711243144317237, 'extraordinari'),
 (-10.711243144317237, 'extraterrestri'),
 (-10.711243144317237, 'feast'),
 (-10.711243144317237, 'federalist'),
 (-10.711243144317237, 'ferguson'),
 (-10.711243144317237, 'find american'),
 (-10.711243144317237, 'fl'),
 (-10.711243144317237, 'fleet'),
 (-10.711243144317237, 'flower'),
 (-10.711243144317237, 'forc new'),
 (-10.711243144317237, 'forev'),
 (-10.711243144317237, 'found dead'),
 (-10.711243144317237, 'front line'),
 (-10.711243144317237, 'garner'),
 (-10.711243144317237, 'get readi'),
 (-10.711243144317237, 'glitch'),
 (-10.711243144317237, 'go full'),
 (-10.711243144317237, 'go new'),
 (-10.711243144317237, 'go new york'),
 (-10.711243144317237, 'go prison'),
 (-10.711243144317237, 'googl facebook'),
 (-10.711243144317237, 'gop senat'),
 (-10.711243144317237, 'govt'),
 (-10.711243144317237, 'gowdi'),
 (-10.711243144317237, 'guess'),
 (-10.711243144317237, 'ha'),
 (-10.711243144317237, 'halloween'),
 (-10.711243144317237, 'hardest'),
 (-10.711243144317237, 'harri reid'),
 (-10.711243144317237, 'harsh'),
 (-10.711243144317237, 'headquart'),
 (-10.711243144317237, 'heal'),
 (-10.711243144317237, 'hilari'),
 (-10.711243144317237, 'hillari clinton campaign'),
 (-10.711243144317237, 'hillari email'),
 (-10.711243144317237, 'holi'),
 (-10.711243144317237, 'hollywood star'),
 (-10.711243144317237, 'honest'),
 (-10.711243144317237, 'horizon'),
 (-10.711243144317237, 'hq'),
 (-10.711243144317237, 'hunter'),
 (-10.711243144317237, 'hybrid'),
 (-10.711243144317237, 'ident polit'),
 (-10.711243144317237, 'iii'),
 (-10.711243144317237, 'incit violenc'),
 (-10.711243144317237, 'infiltr'),
 (-10.711243144317237, 'info'),
 (-10.711243144317237, 'insan'),
 (-10.711243144317237, 'interf'),
 (-10.711243144317237, 'invad'),
 (-10.711243144317237, 'invent'),
 (-10.711243144317237, 'investig clinton'),
 (-10.711243144317237, 'investig end'),
 (-10.711243144317237, 'invis'),
 (-10.711243144317237, 'justic depart'),
 (-10.711243144317237, 'keef'),
 (-10.711243144317237, 'kill civilian'),
 (-10.711243144317237, 'la vega'),
 (-10.711243144317237, 'lab'),
 (-10.711243144317237, 'landslid'),
 (-10.711243144317237, 'larger'),
 (-10.711243144317237, 'launder'),
 (-10.711243144317237, 'leonard'),
 (-10.711243144317237, 'libertarian'),
 (-10.711243144317237, 'liquid'),
 (-10.711243144317237, 'liter'),
 (-10.711243144317237, 'makeup'),
 (-10.711243144317237, 'manipul'),
 (-10.711243144317237, 'mask'),
 (-10.711243144317237, 'meddl'),
 (-10.711243144317237, 'medic marijuana'),
 (-10.711243144317237, 'militar'),
 (-10.711243144317237, 'milk'),
 (-10.711243144317237, 'million dollar'),
 (-10.711243144317237, 'mirror'),
 (-10.711243144317237, 'mitt'),
 (-10.711243144317237, 'mitt romney'),
 (-10.711243144317237, 'monopoli'),
 (-10.711243144317237, 'monsanto'),
 (-10.711243144317237, 'monster'),
 (-10.711243144317237, 'motiv'),
 (-10.711243144317237, 'msm'),
 (-10.711243144317237, 'necessari'),
 (-10.711243144317237, 'need know'),
 (-10.711243144317237, 'new email'),
 (-10.711243144317237, 'new fbi'),
 (-10.711243144317237, 'new poll'),
 (-10.711243144317237, 'new video'),
 (-10.711243144317237, 'notic'),
 (-10.711243144317237, 'novemb'),
 (-10.711243144317237, 'nowher'),
 (-10.711243144317237, 'obamacar architect'),
 (-10.711243144317237, 'observ'),
 (-10.711243144317237, 'obstruct justic'),
 (-10.711243144317237, 'obviou'),
 (-10.711243144317237, 'occupi'),
 (-10.711243144317237, 'oliv'),
 (-10.711243144317237, 'open letter'),
 (-10.711243144317237, 'outcom'),
 (-10.711243144317237, 'outlet'),
 (-10.711243144317237, 'pa'),
 (-10.711243144317237, 'palestin'),
 (-10.711243144317237, 'partial'),
 (-10.711243144317237, 'payrol'),
 (-10.711243144317237, 'peak'),
 (-10.711243144317237, 'pedophil'),
 (-10.711243144317237, 'penalti'),
 (-10.711243144317237, 'plastic'),
 (-10.711243144317237, 'plu'),
 (-10.711243144317237, 'poll number'),
 (-10.711243144317237, 'poll station'),
 (-10.711243144317237, 'popular vote'),
 (-10.711243144317237, 'possess'),
 (-10.711243144317237, 'pr'),
 (-10.711243144317237, 'prof'),
 (-10.711243144317237, 'proof'),
 (-10.711243144317237, 'proven'),
 (-10.711243144317237, 'pundit'),
 (-10.711243144317237, 'pure'),
 (-10.711243144317237, 'pussi'),
 (-10.711243144317237, 'raqqa'),
 (-10.711243144317237, 'raw'),
 (-10.711243144317237, 'readi go'),
 (-10.711243144317237, 'real reason'),
 (-10.711243144317237, 'reason vote'),
 (-10.711243144317237, 'recent'),
 (-10.711243144317237, 'reduc'),
 (-10.711243144317237, 'regist'),
 (-10.711243144317237, 'religion'),
 (-10.711243144317237, 'reopen investig'),
 (-10.711243144317237, 'repeatedli'),
 (-10.711243144317237, 'republ'),
 (-10.711243144317237, 'reset'),
 (-10.711243144317237, 'resourc'),
 (-10.711243144317237, 'right council'),
 (-10.711243144317237, 'ritual'),
 (-10.711243144317237, 'rock protest'),
 (-10.711243144317237, 'rocki'),
 (-10.711243144317237, 'rt'),
 (-10.711243144317237, 'safe space'),
 (-10.711243144317237, 'salt'),
 (-10.711243144317237, 'sand'),
 (-10.711243144317237, 'satan'),
 (-10.711243144317237, 'satellit'),
 (-10.711243144317237, 'say hillari'),
 (-10.711243144317237, 'scorpio'),
 (-10.711243144317237, 'scream'),
 (-10.711243144317237, 'script'),
 (-10.711243144317237, 'secreci'),
 (-10.711243144317237, 'server'),
 (-10.711243144317237, 'sex scandal'),
 (-10.711243144317237, 'shelter'),
 (-10.711243144317237, 'sieg'),
 (-10.711243144317237, 'silent coup'),
 (-10.711243144317237, 'slush'),
 (-10.711243144317237, 'smash'),
 (-10.711243144317237, 'smoke gun'),
 (-10.711243144317237, 'spill'),
 (-10.711243144317237, 'stand rock protest'),
 (-10.711243144317237, 'state depart'),
 (-10.711243144317237, 'stock market'),
 (-10.711243144317237, 'stole'),
 (-10.711243144317237, 'su'),
 (-10.711243144317237, 'suck'),
 (-10.711243144317237, 'suppos'),
 (-10.711243144317237, 'suppress'),
 (-10.711243144317237, 'surrog'),
 (-10.711243144317237, 'survey find'),
 (-10.711243144317237, 'sustain'),
 (-10.711243144317237, 'swing state'),
 (-10.711243144317237, 'tale'),
 (-10.711243144317237, 'teach'),
 (-10.711243144317237, 'terribl'),
 (-10.711243144317237, 'thin'),
 (-10.711243144317237, 'thing learn'),
 (-10.711243144317237, 'thing need'),
 (-10.711243144317237, 'thing need know'),
 (-10.711243144317237, 'think tank'),
 (-10.711243144317237, 'tire'),
 (-10.711243144317237, 'tobi'),
 (-10.711243144317237, 'top democrat'),
 (-10.711243144317237, 'torch'),
 (-10.711243144317237, 'tragic'),
 (-10.711243144317237, 'transform'),
 (-10.711243144317237, 'tri steal'),
 (-10.711243144317237, 'tribun'),
 (-10.711243144317237, 'trump fan'),
 (-10.711243144317237, 'trump favor'),
 (-10.711243144317237, 'trump hollywood'),
 (-10.711243144317237, 'trump lose'),
 (-10.711243144317237, 'trump warn'),
 (-10.711243144317237, 'ukrainian'),
 (-10.711243144317237, 'ultim'),
 (-10.711243144317237, 'unknown'),
 (-10.711243144317237, 'unrest'),
 (-10.711243144317237, 'unusu'),
 (-10.711243144317237, 'up'),
 (-10.711243144317237, 'upon'),
 (-10.711243144317237, 'upsid'),
 (-10.711243144317237, 'use privat'),
 (-10.711243144317237, 'util'),
 (-10.711243144317237, 'vaccin'),
 (-10.711243144317237, 'vega'),
 (-10.711243144317237, 'versu'),
 (-10.711243144317237, 'vineyard'),
 (-10.711243144317237, 'viral'),
 (-10.711243144317237, 'vitamin'),
 (-10.711243144317237, 'von'),
 (-10.711243144317237, 'vote donald'),
 (-10.711243144317237, 'vote donald trump'),
 (-10.711243144317237, 'vote hillari'),
 (-10.711243144317237, 'vote machin'),
 (-10.711243144317237, 'war russia'),
 (-10.711243144317237, 'washington dc'),
 (-10.711243144317237, 'washington state'),
 (-10.711243144317237, 'wast'),
 (-10.711243144317237, 'wealthi'),
 (-10.711243144317237, 'weather'),
 (-10.711243144317237, 'weiner laptop'),
 (-10.711243144317237, 'weird'),
 (-10.711243144317237, 'whistleblow'),
 (-10.711243144317237, 'wikileak email'),
 (-10.711243144317237, 'wikileak releas'),
 (-10.711243144317237, 'wildlif refug'),
 (-10.711243144317237, 'win presid'),
 (-10.711243144317237, 'world seri'),
 (-10.711243144317237, 'wow'),
 (-10.711243144317237, 'year sinc'),
 (-10.711243144317237, 'yet anoth'),
 (-10.288386293497204, 'abc news'),
 (-10.288386293497204, 'abl'),
 (-10.288386293497204, 'absurd'),
 (-10.288386293497204, 'accid'),
 (-10.288386293497204, 'achiev'),
 (-10.288386293497204, 'acknowledg'),
 (-10.288386293497204, 'acquitt'),
 (-10.288386293497204, 'affect'),
 (-10.288386293497204, 'affili'),
 (-10.288386293497204, 'ai'),
 (-10.288386293497204, 'air forc'),
 (-10.288386293497204, 'aircraft'),
 (-10.288386293497204, 'al gore'),
 (-10.288386293497204, 'alex jone'),
 (-10.288386293497204, 'ancient'),
 (-10.288386293497204, 'appar'),
 (-10.288386293497204, 'appli'),
 (-10.288386293497204, 'asid'),
 (-10.288386293497204, 'averag'),
 (-10.288386293497204, 'aviv'),
 (-10.288386293497204, 'awaken'),
 (-10.288386293497204, 'ballist'),
 (-10.288386293497204, 'ballist missil'),
 (-10.288386293497204, 'bang'),
 (-10.288386293497204, 'bare'),
 (-10.288386293497204, 'battleground state'),
 (-10.288386293497204, 'begun'),
 (-10.288386293497204, 'bibl'),
 (-10.288386293497204, 'boil'),
 (-10.288386293497204, 'bound'),
 (-10.288386293497204, 'brave'),
 (-10.288386293497204, 'break new'),
 (-10.288386293497204, 'brilliant'),
 (-10.288386293497204, 'brought'),
 (-10.288386293497204, 'bull'),
 (-10.288386293497204, 'cage'),
 (-10.288386293497204, 'capabl'),
 (-10.288386293497204, 'carbon'),
 (-10.288386293497204, 'catastroph'),
 (-10.288386293497204, 'celeb'),
 (-10.288386293497204, 'censorship'),
 (-10.288386293497204, 'chariti'),
 (-10.288386293497204, 'chart'),
 (-10.288386293497204, 'child rape'),
 (-10.288386293497204, 'cleveland'),
 (-10.288386293497204, 'clinton foundat'),
 (-10.288386293497204, 'clinton investig'),
 (-10.288386293497204, 'clinton win'),
 (-10.288386293497204, 'cloth'),
 (-10.288386293497204, 'cold war'),
 (-10.288386293497204, 'collud'),
 (-10.288386293497204, 'combin'),
 (-10.288386293497204, 'come back'),
 (-10.288386293497204, 'command'),
 (-10.288386293497204, 'common core'),
 (-10.288386293497204, 'common ground'),
 (-10.288386293497204, 'complain'),
 (-10.288386293497204, 'concuss'),
 (-10.288386293497204, 'confess'),
 (-10.288386293497204, 'contest'),
 (-10.288386293497204, 'contractor'),
 (-10.288386293497204, 'contribut'),
 (-10.288386293497204, 'could go'),
 (-10.288386293497204, 'crack'),
 (-10.288386293497204, 'creativ'),
 (-10.288386293497204, 'crossroad'),
 (-10.288386293497204, 'cultiv'),
 (-10.288386293497204, 'cunningham'),
 (-10.288386293497204, 'current'),
 (-10.288386293497204, 'cycl'),
 (-10.288386293497204, 'da'),
 (-10.288386293497204, 'definit'),
 (-10.288386293497204, 'depress'),
 (-10.288386293497204, 'determin'),
 (-10.288386293497204, 'directli'),
 (-10.288386293497204, 'dirti'),
 (-10.288386293497204, 'diseas'),
 (-10.288386293497204, 'divis'),
 (-10.288386293497204, 'dna'),
 (-10.288386293497204, 'donald trump elect'),
 (-10.288386293497204, 'dow'),
 (-10.288386293497204, 'drill'),
 (-10.288386293497204, 'drink water'),
 (-10.288386293497204, 'drought'),
 (-10.288386293497204, 'duke'),
 (-10.288386293497204, 'dumb'),
 (-10.288386293497204, 'earli vote'),
 (-10.288386293497204, 'elect day'),
 (-10.288386293497204, 'elect donald'),
 (-10.288386293497204, 'elect donald trump'),
 (-10.288386293497204, 'elect result'),
 (-10.288386293497204, 'elect trump new'),
 (-10.288386293497204, 'email show'),
 (-10.288386293497204, 'employ'),
 (-10.288386293497204, 'en'),
 (-10.288386293497204, 'endors hillari'),
 (-10.288386293497204, 'endors hillari clinton'),
 (-10.288386293497204, 'enrag'),
 (-10.288386293497204, 'er'),
 (-10.288386293497204, 'erupt'),
 (-10.288386293497204, 'evan mcmullin'),
 (-10.288386293497204, 'evangel'),
 (-10.288386293497204, 'evict'),
 (-10.288386293497204, 'examin'),
 (-10.288386293497204, 'excit'),
 (-10.288386293497204, 'exec'),
 (-10.288386293497204, 'fals flag'),
 (-10.288386293497204, 'fbi director jame'),
 (-10.288386293497204, 'fema'),
 (-10.288386293497204, 'fewer'),
 (-10.288386293497204, 'filmmak'),
 (-10.288386293497204, 'flash'),
 (-10.288386293497204, 'flip'),
 (-10.288386293497204, 'fold'),
 (-10.288386293497204, 'fool'),
 (-10.288386293497204, 'footag'),
 (-10.288386293497204, 'forest'),
 (-10.288386293497204, 'four year'),
 (-10.288386293497204, 'frack'),
 (-10.288386293497204, 'freak'),
 (-10.288386293497204, 'ft'),
 (-10.288386293497204, 'fukushima'),
 (-10.288386293497204, 'futur new'),
 (-10.288386293497204, 'futur new york'),
 (-10.288386293497204, 'garden'),
 (-10.288386293497204, 'gaza'),
 (-10.288386293497204, 'genocid'),
 (-10.288386293497204, 'ghost'),
 (-10.288386293497204, 'giuliani'),
 (-10.288386293497204, 'glass'),
 (-10.288386293497204, 'gm'),
 (-10.288386293497204, 'go back'),
 (-10.288386293497204, 'gore'),
 (-10.288386293497204, 'greater'),
 (-10.288386293497204, 'grid'),
 (-10.288386293497204, 'guardian'),
 (-10.288386293497204, 'gym'),
 (-10.288386293497204, 'gymnast'),
 (-10.288386293497204, 'hat'),
 (-10.288386293497204, 'hater'),
 (-10.288386293497204, 'haunt'),
 (-10.288386293497204, 'heard'),
 (-10.288386293497204, 'hike'),
 (-10.288386293497204, 'hip'),
 (-10.288386293497204, 'home invas'),
 (-10.288386293497204, 'homeless man'),
 (-10.288386293497204, 'hors'),
 (-10.288386293497204, 'howard'),
 (-10.288386293497204, 'huffington'),
 (-10.288386293497204, 'humanitarian'),
 (-10.288386293497204, 'hurrican matthew'),
 (-10.288386293497204, 'idaho'),
 (-10.288386293497204, 'identifi'),
 (-10.288386293497204, 'imposs'),
 (-10.288386293497204, 'inc'),
 (-10.288386293497204, 'incid'),
 (-10.288386293497204, 'infant'),
 (-10.288386293497204, 'influenti'),
 (-10.288386293497204, 'inject'),
 (-10.288386293497204, 'instal'),
 (-10.288386293497204, 'intend'),
 (-10.288386293497204, 'intensifi'),
 (-10.288386293497204, 'intimid'),
 (-10.288386293497204, 'isra settlement'),
 (-10.288386293497204, 'janet'),
 (-10.288386293497204, 'jesu'),
 (-10.288386293497204, 'judg order'),
 (-10.288386293497204, 'kurd'),
 (-10.288386293497204, 'lash'),
 (-10.288386293497204, 'laugh'),
 (-10.288386293497204, 'lead new'),
 (-10.288386293497204, 'legitim'),
 (-10.288386293497204, 'lewd'),
 (-10.288386293497204, 'live stream'),
 (-10.288386293497204, 'loan'),
 (-10.288386293497204, 'loretta'),
 (-10.288386293497204, 'loretta lynch'),
 (-10.288386293497204, 'lover'),
 (-10.288386293497204, 'loyalti'),
 (-10.288386293497204, 'lynch'),
 (-10.288386293497204, 'mail'),
 (-10.288386293497204, 'marco rubio'),
 (-10.288386293497204, 'match'),
 (-10.288386293497204, 'mayhem'),
 (-10.288386293497204, 'mcilroy'),
 (-10.288386293497204, 'mcmullin'),
 (-10.288386293497204, 'mental health'),
 (-10.288386293497204, 'mention'),
 (-10.288386293497204, 'meter'),
 (-10.288386293497204, 'mit'),
 (-10.288386293497204, 'mitchel'),
 (-10.288386293497204, 'monument'),
 (-10.288386293497204, 'mouth'),
 (-10.288386293497204, 'mp'),
 (-10.288386293497204, 'must see'),
 (-10.288386293497204, 'myth'),
 (-10.288386293497204, 'nail'),
 (-10.288386293497204, 'nanci pelosi'),
 (-10.288386293497204, 'natur'),
 (-10.288386293497204, 'neg'),
 (-10.288386293497204, 'nevada'),
 (-10.288386293497204, 'new evid'),
 (-10.288386293497204, 'new realiti'),
 (-10.288386293497204, 'newt'),
 (-10.288386293497204, 'next year'),
 (-10.288386293497204, 'nightmar'),
 (-10.288386293497204, 'norway'),
 (-10.288386293497204, 'novel'),
 (-10.288386293497204, 'nra'),
 (-10.288386293497204, 'nuclear war'),
 (-10.288386293497204, 'nurs'),
 (-10.288386293497204, 'nut'),
 (-10.288386293497204, 'nypd'),
 (-10.288386293497204, 'oath'),
 (-10.288386293497204, 'obama doj'),
 (-10.288386293497204, 'obstruct'),
 (-10.288386293497204, 'octob'),
 (-10.288386293497204, 'oldest'),
 (-10.288386293497204, 'one step'),
 (-10.288386293497204, 'opec'),
 ...]
```

Chapter: 5

CONCLUSIONS

5.1 SUMMARY

Based on the above the result are very promising. From the results, methods state that term frequency is able to predict fake news. To the best of my knowledge, it is first step toward using machine classification for identification of fake new in Afaan Oromo. The best performing models are a combination of multinomial naïve Bayes classifier with term frequency, feature extraction and unigram which is 96% classification accuracy. The model was evaluated using thresholds of 0.7, which may not be the most reliable for models where probability scoring is not well calibrated. TF performs better, but frequent words but not important words affect the result. These challenges bound the analysis and prevent us from broader generalizability. So, it was planned to address these issues and abut slang words in a future work.

5.2 FUTURE SCOPE

In the future, a web-based GUI can be created for the proposed fake news detection system to classify the news as fake or real on real-time social media platforms such as Facebook, Instagram, Twitter, WhatsApp, etc. Also, the annotated dataset in the sequence of images (with textual content written on them) will be collected and maintained from Facebook and Reddit platforms. The annotated dataset is often used for detecting fake images within the future as no such dataset is out there at the present. The proposed system has the potential to provide an impulse to various emerging applications such as controlling the spread of fake news during elections, terrorism, natural calamities, crimes for the betterment of society. In the future, the efficiency and accuracy of the prototype can be enhanced to a certain level, and also enhance the user interface of the proposed model.