|  |  |  |
| --- | --- | --- |
| **Tech Saksham**  Final Project Report  **DEEP DIVE AI** |  |  |

**“Fake News Detection”**

**“MIT college of Arts & Science for Women”**

|  |  |
| --- | --- |
| **ROLL NO** | **NAME** |
| CB20S 202964 | G.Ragavi |
| CB20S 202955 | M.Kaviya |
| CB20S 202959 | D.Kowsalya |
| CB20S 202976 | S.Susmitha |

|  |  |
| --- | --- |
|  |  |
|  | Trainer Name |
|  | Master Trainer |

**Mrs.Mayank Shrivastava**

**ABSTRACT**

* 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.
* With the current usage of social media platforms, consumers are creating and sharing more information than ever before, some of which are misleading with no relevance to reality.
* Automated classification of a text article as misinformation or disinformation is a challenging task. Even an expert in a particular domain has to explore multiple aspects before giving a verdict on the truthfulness of an article.
* In this work, we propose to use machine learning ensemble approach for automated classification of news articles.
* Our study explores different textual properties that can be used to distinguish fake contents from real.
* By using those properties, we train a combination of different machine learning algorithms using various ensemble methods and evaluate their performance on 4 real world datasets.
* Experimental evaluation confirms the superior performance of our proposed ensemble learner approach in comparison to individual learn

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Table of Contents** | **Page No.** |
| 1 | Chapter 1: Introduction | 3 |
| 2 | Chapter 2: Services and Tools Required | 7 |
| 3 | Chapter 3: Project Architecture | 12 |
| 4 | Chapter 4: Architecture Blocks Detail Working | 13 |
| 5 | Conclusion | 16 |
| 6 | References | 17 |
| 7 | Code | 18 |

**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview**

Fake news detection is a subtask of text classification [1] and is often defined as the task of classifying news as real or fake. The term 'fake news' refers to the false or misleading information that appears as real news. It aims to deceive or mislead people.

* 1. **Feature**

Fake news detection is a subtask of text classification [1] and is often defined as the task of classifying news as real or fake. The term 'fake news' refers to the false or misleading information that appears as real news. It aims to deceive or mislead people.

**1.3 Advantages**

Apart from the pandemic, hundreds of incidents where people or political groups used fabricated information or altered data to misinform people. Obviously, for self-interests. It’s time to know can fake news be of any benefit; if yes, then who?

### 1. Advertisers take the Advantages of Fake News

[Digital News Media](https://www.myayan.com/advantages-and-disadvantages-of-digital-news-media) is the cheapest and easiest way to get in touch with the users and deliver information. So, that’s how it goes; a fake new website display draws the reader’s attention and makes money whenever users click an ad.

The websites partner with ad providers, giving them commissions based on how many viewers you manage to get through fake news. The biggest reason why readers love to read fake news is it’s compelling and arouses curiosity among users.

### 2. Influencers also take benefits of Fake News

Influencers with millions of fan following get money from their sponsors Social media influencers need daily content to draw their followers’ attention. Especially those who don‘t have special talent but republish the content or misrepresent the reality to grab public attention are the ones who **benefit from fake news**.

TikTok Video App is a better example of what many users do by scripting the real scenario and fooling the viewers. Many others spread false information through YouTube to influence the public and show they are the experts.

### 3. Political Warfare

The majority of the conversation these days revolves around its role in political campaigns and influencing civilians views’ on political parties. Today, fundamentalist violent extremist organizations have utilized social media to gather support for their groups by posting videos and articles of their actions online for the public to see.

Incidents like riots or videos that incite communal disharmony are trends that easily get millions of views within minutes. The rest is done by the political leaders who look for such information to blame the opposition. Most political parties take the **advantages of fake news** during elections.

### 4. Fun and Entertainment

Many individuals use fake news in the form of satire or parody to have fun and get attention from their friends. Everyone loves to see funny stuff online, and most of the content you’ll find on the internet is either fabricated or twisted to make it more entertaining for viewers.

The actor uses doctored images to manipulate users and let them believe the content is original and genuine. People sharing links of money-making apps encouraging other group members to follow the same is another example of manipulating

**1.4 Scope**

There are four perspectives given the literature for automatic detection of fake news: the unfactual knowledge it conveys, its style of writing or content-based, its propagation patterns or social, and the credibility of its source.

media rather than ancient fourth estate. social media has conjointly been accustomed unfold

pretend news, that has sturdy negative impacts on individual users and broader society. We have

a tendency to explore the pretend news drawback by reviewing existing literature in two phases

characterization and detection. Within the characterization part, we have a tendency to introduced

the essential ideas and principles of faux news in each ancient media and social media. Within the

detection part, we have tendency to reviewed existing pretend news detection approaches from a

knowledge mining perspective, together with feature extraction and model construction. We

have tendency to conjointly more mentioned the datasets, analysis metrics, and promising future

directions in pretend news detection analysis and expand the sphere to alternative applications

**1.5 Future Work**

This advanced python project of detecting fake news deals with fake and real news. Using sklearn, we build a TfidfVectorizer on our dataset. Then, we initialize a PassiveAggressive Classifier and fit the model. In the end, the accuracy score and the confusion matrix tell us how well our model fares.

**CHAPTER 2**

**SERVICES AND TOOLS REQUIRED**

**2.1 Services Used**

Fake News Detection based on Machine learning . The Service we provide the provision of tools as a service which can be applied to allow:

-Multimedia data gathering and data lake model creation,

-Developing tools based on machine learning and data analytics towards detecting the content of fake or real news.

-detection of source in fake news activities.

And providing a score of potential fake news(including the reason for such indication),and classification of news to allow journalists and digitalmedia companies to detect fake news

**2.1.1 Liberty Profile**

We collect and analyze user profile features from different aspects, i.e., implicit and explicit. Implicit features are not directly available but are inferred from user meta information or online behaviors, such as historical tweets. Explicit features are obtained directly from meta-data returned by querying social media site APIs. The implicit features include: age, personality, location, profile image, political bias. Due to space limitation, we ignore the description of age and personality. For explicit features, we have similar observations for explicit profile features as in , so we omit the discussion due to the space limitation. All features analysis are included here.

**Location**: Research has shown an inseparable relationship between user profiles and geo-locations. However, the location fields are usually very sparse. Thus, we exploit user-posted content to predict the user’s location . The idea is to identify “location indicative words” (LIW), which can encode an association with a particular location. The implementation of a pre-trained LIW model is integrated into an open source tool named pigeo , which is utilized here to predict the geolocations of users in U (f) and U (r) . The predicted results of pigeo are at the city-level and also include (latitude, longitude) pairs and we observe that:

(1) there are overall more users located in the US than other places, which is because most of the real/fake news items in our particular datasets are published and related to US politics and entertainments;

(2) the location distribution is different for fake and real news on both datasets, and the red and blue dots demonstrate the degree of differences. For example, there are general more real news share in east region of US in our datasets.

**Profile Image**: Profile images are important visual components of users on social media. Various studies have demonstrated the correlation between the choice of profile images with user personalities, behaviors, and activities. We classify the object types in profile images. With the recent development of deep learning in the computer vision domain, convolutional neural networks (CNN) have shown good performance for detecting objects in images. We chose the pre-trained VGG16 model [8] as it is the widely-used CNN architecture. We see that: the distributions of profile image classes are different for users in U (f) and U (r) on both datasets. For example, there are specific image types, such as “wig” and “mask” dominating the image categories for users spreading fake news, and “website” and “envelope” dominating the image categories for users spreading real news, on both datasets consistently. Political Bias: Political bias plays an important role in shaping users’ profiles and affecting their news consumption choices. Sociological studies on journalism demonstrate the correlation between partisan bias and news content authenticity (i.e., fake or real news) . Reports have shown people’s political affiliation is correlated with their attributes and behaviors . Thus, we adopt method in to measure user political bias scores by exploiting users’ interests

**2.2 Tools and Softwares used**

**Python:** Python is often used as a support language for software developers, for build control and management, testing, and in many other ways. SCons for build control. Buildbot and Apache Gump for automated continuous compilation and testing. Roundup or Trac for bug tracking and project management.

**Anaconda Navigator:** Anaconda Navigator is included in the Anaconda distribution, and allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages, install them in an environment, run the packages and update them.

**Jupiter Notebook:** Anaconda Navigator is a GUI tool that is included in the Anaconda distribution and makes it easy to configure, install, and launch tools such as Jupyter Notebook. ◆ A Conda Python environment is an isolated environment. It allows you to install packages without modifying your system's Python installation.

**TfidVectorizer:** The TfidfVectorizer turns a set of raw documents into a TF-IDF feature matrix. Python implementation of Us with and Word2Vec word embeddings.

**fit\_transform:** It is used to train data in order to scale it and learn the scaling parameters.

**LogisticRegression:** Based on a collection of independent variables, logistic regression assesses the likelihood of an event occurring, such as voting or not voting. Because the outcome is a probability, the dependent variable is limited to values between 0 and 1.

**DecisionTreeClassifier:** The DecisionTreeClassifier class may conduct multi-class classification on a dataset. If numerous classes have the same and highest probability, the classifier will forecast the class with the lowest index among those classes.

**Re**: The functions in this module allow you to determine whether a given text fits a given regular expression, known as a regular expression.

**String:** You can use the Python library NLTK, or Natural Language Toolkit, for NLP. A large portion of the data you might be examining is unstructured and contains text humans can read. Preprocessing that data is necessary before you can programmatically evaluate it.

**train\_test\_split():**[Machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning) algorithms applicable to prediction-based algorithms and applications are evaluated using the train-test split. We can compare the output of our own machine-learning model to that of other machines using this quick and simple process.

**INSTALLING THE LIBRARIES**

**Pandas:** Working with "relational" or "labeled" data can be simple and intuitive thanks to the Python module pandas, which offers quick, adaptable, and expressive data structures.

!pip install pandas

**Numpy:** The Python package NumPy is used to manipulate arrays.Additionally, it has matrices, Fourier transform, and functions for working in the area of linear algebra.

!pip install numpy

**Seaborn:** A package called [Seaborn](https://www.simplilearn.com/tutorials/python-tutorial/python-seaborn) uses Matplotlib as its foundation to plot graphs. In order to see random distributions, it will be used.

!pip install seaborn

**Matplotlib:** For the Python programming language & its NumPy numerical mathematics add-on, Matplotlib is a graphing library. It offers an object-oriented [API](https://www.simplilearn.com/building-blocks-of-api-development-guide-pdf) for integrating charts into programs utilizing all-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

!pip install matplotlib

**Sklearn:** It includes a variety of classification, regression, and clustering methods, such as support vector machines, random forests, gradient boosting, k-means, and DBSCAN, and is built to work with Python's NumPy and SciPy scientific and numerical libraries.

!pip3 install sklearn

**2.2.1 NodeJS**

Node. js is an open-source, cross-platform JavaScript runtime environment and library for running web applications outside the client's browser . Building Fake News Detection using Angular 6 in the frontend,Node JS in the backend to build API using Express JS and Python Scikit Learn machine learning package for detecting FAKE NEWS.

**2.2.2 HTML**

Detecting Fake News with BERT and LSTMs.Developing a web app with Flask and HTML for the project.

**2.2.3 Cloud foundry**

**Cloud Computing :**

In parallel with the increase in data, there has also been an increase in the sources of data,

which has led to a phenomenon called big data. With the emergence of big data, this concept

has proved its value in many areas, such as economics, and it is likely to grow even more in

the future. Increased data sharing in social networks has led to the foundation of cloud

computing as a result of the increase in the number of users, the formation of data lakes and

thus their storage and presentation at any time or place increasing data sharing in social

networks, the increase in the number of users, the formation of data lakes, and thus their

storage and presentation in places and times ready for use have led to the foundations of cloud

computing. . Cloud computing has emerged as software and applications running on an

infrastructure where data is shared over a network and is usually related to big data. Cloud

services are offered by CSPs such as Amazon Web services (AWS), Microsoft Azure, IBM

cloud, etc., and are generally reviewed in three categories .

* Infrastructure as a service (IaaS): It represents the infrastructure needs that can be

considered as the basis of cloud computing, servers and other physical storage environments

provided and managed over the Internet. IaaS is often used in data storage, backup and

recovery scenarios to prevent sudden data losses, in big data analysis, in high performance

computing, web applications, testing and development activities [20].

* Platform as a service (PaaS): Software, database management systems, business

development tools, business intelligence AI, ML tools are found in this platform as services.

In this section, more business intelligence and analysis processes are included [21,22].

* Software as a service (SaaS): The use of this service in detecting fake news is

generally for users to use and share various office tools and databases over the internet. This

service is used for various business applications such as email, customer relationship

management (CRM), human resources management, financial management, database

management, and enterprise resource planning .

**TYPES OF CLOUD:**

* Public Cloud: Quick access to information technology (IT) resources is provided in the public cloud. In this category of services, users do not need to purchase infrastructure, basic needs belong to and are managed by service providers.

In public cloud services, users do not need to purchase hardware, software, or supporting infrastructure, as such essential elements are owned and managed by providers.

In this type of service, innovative SaaS business applications are offered for various applications from source and process management to data analysis in detecting fake news. It is often used for short -term storage services and is suitable for cloud-based application development and distribution environments .

* Private Cloud: The private cloud is more suitable for personal use or more private use

in the corporate context and is purpose oriented. In this way, efficient use of resources and

productivity is maximized. The limited number of users is important in terms of implementing

more effective data analysis and achieving the correct result. The most basic feature of this

service is the flexibility and ability to plan and analyze for the purposes. Since the user knows

the need, he/she can take the initiative to reach the goal and security is optimized .

* Hybrid Cloud: The basis of the hybrid cloud is the private cloud, but in the form of

this service, there is a purposeful integration and use of public cloud services. Resources in the

hybrid cloud type are generally not considered separately from the public cloud. As the use of

cloud computing increases professionally, managing private cloud and public cloud workloads

has resulted in the use of the hybrid cloud. The hybrid cloud basically allows the analyzed data

and critical applications to be stored in a traditional data center environment or in a private

cloud. For example, the environment where previously analyzed and non-critical data is stored

is the public cloud environment, while the storage environment for critical and confidential

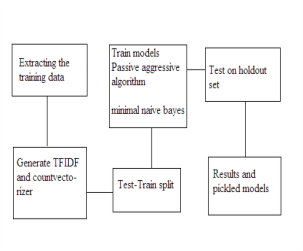
information should be a private cloud. The hybrid cloud serves as an effective and useful

cloud service in terms of this purpose .

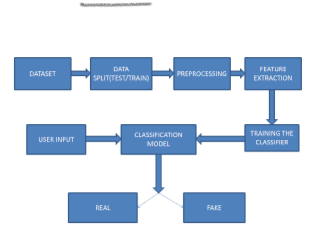
**CHAPTER 3**

**FAKE NEWS DETECTION ARCHITECTURE**

**3.1 Architecture**

****

**SYSTEM ARCHITECTURE:**

****

**CHAPTER 4**

**ARCHITECTURE BLOCKS DETAIL WORKING**

**4.1 Blocks**

**Block Diagram:**

The block diagram for the specified project is depicted in the graphic above. The initial stage, as indicated in the block diagram, is to extract the training data and check for any null values. The count vectorizer and tfidf vectorizers are used to extract features from the data sets in the second stage. As a result, the models will be made. Classifiers such as the nave bayes classifier and the passive aggressive classifier will be used to verify these models. On this foundation, the confusion matrix will be built. The next stage will be to verify the correctness of each model's output. The last stage is to determine the news fake or real.

**System Architecture**:

**Working Of The Back End**

Step :1: Importing the required libraries.

Step 2: Reading the data from the dataset.

Step 3: Pre-processing the data.

Step 4: Splitting the data as train set and test set.

Step 5: Extracting the features from the text using various Vectorizers.

a) Using bag of words or Count vectorizer for bith training and test data set.

b) Using tf-idf vectorizer for both training and test data set.

Step 6: Checking whether both the dataframes are equal or not.

Step 7: Performing classification using classifiers

a) Using naïve bayes classifier for both the models created with count vectorizer and tfidf vectorizer. Also creating a confusion matrix showing the details.

b) Using passive aggressive classifier for both the models created with count vectorizer and tf-idf vectorizer. Also creating a confusion matrix showing the details

Step 8: Checking the accuracy provided by each model and selecting the best model Step 9: Saving the best model created using python library.

Step 10: Also creating a function using which we can enter any news and check whether it is real or fake.

**Working Of The Front End**:

In the front end there will be a user interface model created using html and CSS. User can enter any news in the space provided and click on the predict button provided below it. You will get the output as real if the news is true and as fake if the news is false.

**Angular JS:**

AngularJS is a discontinued free and open-source JavaScript-based web framework for developing single-page applications. It was maintained mainly by Google and a community of individuals and corporations.

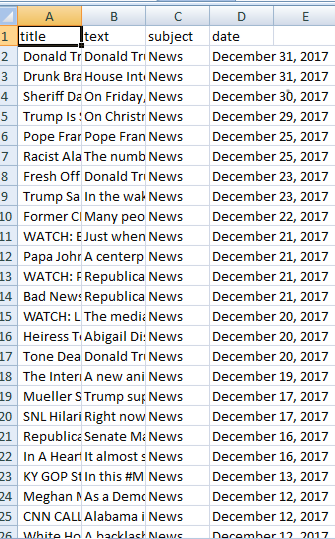
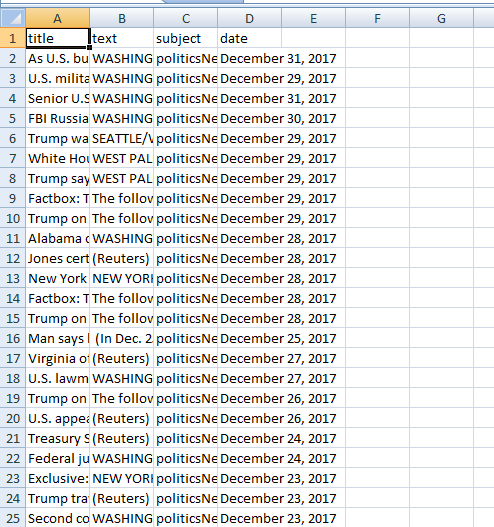
**HTML:**

Detecting Fake News with BERT and LSTMs.Developing a web app with Flask and HTML for the project.

**Data base:**

Data Base is an the Back end Storage of the project dataset.That we want using the dataset is fake and true.That is created on the Excel sheet in Microsoft office

**Fake news Dataset: True News Dataset:**

 ****

**CONCLUSION**

With the increasing quality of social media, additional individuals consume news from social

media rather than ancient fourth estate. social media has conjointly been accustomed unfold

pretend news, that has sturdy negative impacts on individual users and broader society. We have

a tendency to explore the pretend news drawback by reviewing existing literature in two phases

characterization and detection. Within the characterization part, we have a tendency to introduced

the essential ideas and principles of faux news in each ancient media and social media. Within the

detection part, we have tendency to reviewed existing pretend news detection approaches from a

knowledge mining perspective, together with feature extraction and model construction. We

have tendency to conjointly more mentioned the datasets, analysis metrics, and promising future

directions in pretend news detection analysis and expand the sphere to alternative applications

With the increasing quality of social media, additional individuals consume news from social

media rather than ancient fourth estate. social media has conjointly been accustomed unfold

pretend news, that has sturdy negative impacts on individual users and broader society. We have

a tendency to explore the pretend news drawback by reviewing existing literature in two phases

characterization and detection. Within the characterization part, we have a tendency to introduced

the essential ideas and principles of faux news in each ancient media and social media. Within the

detection part, we have tendency to reviewed existing pretend news detection approaches from a

knowledge mining perspective, together with feature extraction and model construction. We

have tendency to conjointly more mentioned the datasets, analysis metrics, and promising future

directions in pretend news detection analysis and expand the sphere to alternative applications

With the increasing quality of social media, additional individuals consume news from social

media rather than ancient fourth estate. social media has conjointly been accustomed unfold

pretend news, that has sturdy negative impacts on individual users and broader society. We have

a tendency to explore the pretend news drawback by reviewing existing literature in two phases

characterization and detection. Within the characterization part, we have a tendency to introduced

the essential ideas and principles of faux news in each ancient media and social media. Within the

detection part, we have tendency to reviewed existing pretend news detection approaches from a

knowledge mining perspective, together with feature extraction and model construction. We

have tendency to conjointly more mentioned the datasets, analysis metrics, and promising future

directions in pretend news detection analysis and expand the sphere to alternative applications

* **The rise of fake news has become a global problem that even major tech companies like Facebook and Google are struggling to solve.**
* **It can be difficult to determine whether a text is factual without additional context and human judgement .**
* **NLP models can detect the topic and the facts of a news article, then classify the same as trustworthy articles and compare the given article topic with the trustworthy articles.**

media rather than ancient fourth estate. social media has conjointly been accustomed unfold

pretend news, that has sturdy negative impacts on individual users and broader society. We have

a tendency to explore the pretend news drawback by reviewing existing literature in two phases

characterization and detection. Within the characterization part, we have a tendency to introduced

the essential ideas and principles of faux news in each ancient media and social media. Within the

detection part, we have tendency to reviewed existing pretend news detection approaches from a

knowledge mining perspective, together with feature extraction and model construction. We

have tendency to conjointly more mentioned the datasets, analysis metrics, and promising future

directions in pretend news detection analysis and expand the sphere to alternative applications

**REFERENCES**

1. <http://www.anaconda.com/>
2. http://en. Wikipedia.org/wiki/Anaconda (python \_distribution)
3. <http://numpy.org/>
4. <http://www.oreily.com/data/the-new-artificial-inteligence-market.csp>
5. <http://seaborn.pydata.org>/
6. <http://www.kaggle.com/>
7. <http://pandas.pydata.org/>
8. <http://matplotlib.org/>
9. <http://scikit-learn.org/>
10. Link for data set;-https://www.simlilearn.com/learn-mac…
11. You tube channel:- <https://bit>.zVT4WtH

**CODE**

# Imporing Libraries

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import accuracy\_score

from sklearn.metrics import classification\_report

import re

import string

#import Datasets

df\_fake = pd.read\_csv("Fake[1].csv")

df\_true = pd.read\_csv("True[2].csv")

df\_fake.head()

OUTPUT

| **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 |

df\_true.head(5)

**OUTPUT**

|  | **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 |
|  |  |  |  |  |

**# Inserting a column "class" as target feature**

**df\_fake["class"] = 0**

**df\_true["class"] = 1**

**df\_fake.shape, df\_true.shape**

**OUTPUT**

((23481, 5), (21417, 5))

**# Removing last 10 rows for manual testing**

**df\_fake\_manual\_testing = df\_fake.tail(10)**

**for i in range(23480,23470,-1):**

**df\_fake.drop([i], axis = 0, inplace = True)**

**df\_true\_manual\_testing = df\_true.tail(10)**

**for i in range(21416,21406,-1):**

**df\_true.drop([i], axis = 0, inplace = True)**

**df\_fake.shape, df\_true.shape**

**OUTPUT**

((23471, 5), (21407, 5))

**df\_fake\_manual\_testing["class"] = 0**

**df\_true\_manual\_testing["class"] = 1**

C:\Users\ELCOT\AppData\Local\Temp\ipykernel\_452\860779283.py:1: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_fake\_manual\_testing["class"] = 0

C:\Users\ELCOT\AppData\Local\Temp\ipykernel\_452\860779283.py:2: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.

Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy>

df\_true\_manual\_testing["class"] = 1

**df\_fake\_manual\_testing.head(10)**

**OUTPUT**

|  | **title** | **text** | **subject** | **date** | **Class** |
| --- | --- | --- | --- | --- | --- |
| **23471** | Seven Iranians freed in the prisoner swap have... | 21st Century Wire says This week, the historic... | Middle-east | January 20, 2016 | 0 |
| **23472** | #Hashtag Hell & The Fake Left | By Dady Chery and Gilbert MercierAll writers ... | Middle-east | January 19, 2016 | 0 |
| **23473** | Astroturfing: Journalist Reveals Brainwashing ... | Vic Bishop Waking TimesOur reality is carefull... | Middle-east | January 19, 2016 | 0 |
| **23474** | The New American Century: An Era of Fraud | Paul Craig RobertsIn the last years of the 20t... | Middle-east | January 19, 2016 | 0 |
| **23475** | Hillary Clinton: ‘Israel First’ (and no peace ... | Robert Fantina CounterpunchAlthough the United... | Middle-east | January 18, 2016 | 0 |
| **23476** | McPain: John McCain Furious That Iran Treated ... | 21st Century Wire says As 21WIRE reported earl... | Middle-east | January 16, 2016 | 0 |
| **23477** | JUSTICE? Yahoo Settles E-mail Privacy Class-ac... | 21st Century Wire says It s a familiar theme. ... | Middle-east | January 16, 2016 | 0 |
| **23478** | Sunnistan: US and Allied ‘Safe Zone’ Plan to T... | Patrick Henningsen 21st Century WireRemember ... | Middle-east | January 15, 2016 | 0 |
| **23479** | How to Blow $700 Million: Al Jazeera America F... | 21st Century Wire says Al Jazeera America will... | Middle-east | January 14, 2016 | 0 |
| **23480** | 10 U.S. Navy Sailors Held by Iranian Military ... | 21st Century Wire says As 21WIRE predicted in ... | Middle-e |  |  |

**df\_true\_manual\_testing.head(10)**

**OUTPUT**

|  | **title** | **text** | **subject** | **date** | **Class** |
| --- | --- | --- | --- | --- | --- |
| **21407** | Mata Pires, owner of embattled Brazil builder ... | SAO PAULO (Reuters) - Cesar Mata Pires, the ow... | worldnews | August 22, 2017 | 1 |
| **21408** | U.S., North Korea clash at U.N. forum over nuc... | GENEVA (Reuters) - North Korea and the United ... | worldnews | August 22, 2017 | 1 |
| **21409** | U.S., North Korea clash at U.N. arms forum on ... | GENEVA (Reuters) - North Korea and the United ... | worldnews | August 22, 2017 | 1 |
| **21410** | Headless torso could belong to submarine journ... | COPENHAGEN (Reuters) - Danish police said on T... | worldnews | August 22, 2017 | 1 |
| **21411** | North Korea shipments to Syria chemical arms a... | UNITED NATIONS (Reuters) - Two North Korean sh... | worldnews | August 21, 2017 | 1 |
| **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 |

**df\_manual\_testing=pd.concat([df\_fake\_manual\_testing,df\_true\_manual\_testing], axis = 0)**

**df\_manual\_testing.to\_csv("manual\_testing.csv")**

**# Merging True and Fake Dataframes**

**df\_merge = pd.concat([df\_fake, df\_true], axis =0)**

**df\_merge.head(10)**

**OUTPUT**

|  | **title** | **text** | **subject** | **date** | **Class** |
| --- | --- | --- | --- | --- | --- |
| **0** | Donald Trump Sends Out Embarrassing New Year’... | Donald Trump just couldn t wish all Americans ... | News | December 31, 2017 | 0 |
| **1** | Drunk Bragging Trump Staffer Started Russian ... | House Intelligence Committee Chairman Devin Nu... | News | December 31, 2017 | 0 |
| **2** | Sheriff David Clarke Becomes An Internet Joke... | On Friday, it was revealed that former Milwauk... | News | December 30, 2017 | 0 |
| **3** | Trump Is So Obsessed He Even Has Obama’s Name... | On Christmas day, Donald Trump announced that ... | News | December 29, 2017 | 0 |
| **4** | Pope Francis Just Called Out Donald Trump Dur... | Pope Francis used his annual Christmas Day mes... | News | December 25, 2017 | 0 |
| **5** | Racist Alabama Cops Brutalize Black Boy While... | The number of cases of cops brutalizing and ki... | News | December 25, 2017 | 0 |
| **6** | Fresh Off The Golf Course, Trump Lashes Out A... | Donald Trump spent a good portion of his day a... | News | December 23, 2017 | 0 |
| **7** | Trump Said Some INSANELY Racist Stuff Inside ... | In the wake of yet another court decision that... | News | December 23, 2017 | 0 |
| **8** | Former CIA Director Slams Trump Over UN Bully... | Many people have raised the alarm regarding th... | News | December 22, 2017 | 0 |
| **9** | WATCH: Brand-New Pro-Trump Ad Features So Muc... | Just when you might have thought we d get a br... | News | December 21, 2017 | 0 |

**df\_merge.columns**

**OUTPUT**

Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')

**# Removing columns which are not required**

**df = df\_merge.drop(["title", "subject", "date"], axis = 1)**

**df.isnull().sum()**

**OUTPUT**

text 0

class 0

dtype: int64

**# Random Shuffling the dataframe**

**df = df.sample(frac = 1)**

**df.head()**

**OUTPUT**

|  | **text** | **class** |
| --- | --- | --- |
| **1388** | Apparently, Fox News hosts do not want their v... | 0 |
| **11724** | BERLIN (Reuters) - Largely unperturbed by Ange... | 1 |
| **16803** | ZURICH (Reuters) - Austria s likely next chanc... | 1 |
| **3243** | Barring something completely unforeseen, Donal... | 0 |
| **12841** | People have a right to understand who it is t... | 0 |

**df.reset\_index(inplace = True)**

**df.drop(["index"], axis = 1, inplace = True)**

**df.columns**

**OUTPUT**

Index(['text', 'class'], dtype='object')

**df.head()**

**OUTPUT**

|  | **text** | **class** |
| --- | --- | --- |
| **0** | Apparently, Fox News hosts do not want their v... | 0 |
| **1** | BERLIN (Reuters) - Largely unperturbed by Ange... | 1 |
| **2** | ZURICH (Reuters) - Austria s likely next chanc... | 1 |
| **3** | Barring something completely unforeseen, Donal... | 0 |
| **4** | People have a right to understand who it is t... | 0 |

**# Creating a function to process the texts**

**def wordopt(text):**

**text = text.lower()**

**text = re.sub('\[.\*?\]', '',text)**

**text = re.sub("\\W"," ",text)**

**text = re.sub('http?://\S+|www\.\S+', '',text)**

**text = re.sub('<.\*?>+', '', text)**

**text = re.sub('[%s]' % re.escape(string.punctuation), '', text)**

**text = re.sub('\n', '', text)**

**text = re.sub('\w\*\d\w\*', '', text)**

**return text**

**df["text"] = df["text"].apply(wordopt)**

**# Defining dependent and independent variables**

**x = df["text"]**

**y = df["class"]**

**# Splitting Training and Testing**

**x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size=0.25)**

**# Convert text to vectors**

**from sklearn.feature\_extraction.text import TfidfVectorizer**

**vectorization = TfidfVectorizer()**

**xv\_train = vectorization.fit\_transform(x\_train)**

**xv\_test = vectorization.transform(x\_test)**

**# Logistic Regression**

**from sklearn.linear\_model import LogisticRegression**

**LR = LogisticRegression()**

**LR.fit(xv\_train,y\_train)**

**OUTPUT**

LogisticRegression()

**pred\_lr=LR.predict(xv\_test)**

**LR.score(xv\_test, y\_test)**

**OUTPUT**

0.9867201426024955

**print(classification\_report(y\_test, pred\_lr))**

**OUTPUT**

precision recall f1-score support

0 0.99 0.99 0.99 5864

1 0.98 0.99 0.99 5356

accuracy 0.99 11220

macro avg 0.99 0.99 0.99 11220

weighted avg 0.99 0.99 0.99 11220

**# Decision Tree Classification**

**from sklearn.tree import DecisionTreeClassifier**

**DT = DecisionTreeClassifier()**

**DT.fit(xv\_train, y\_train)**

**OUTPUT**

DecisionTreeClassifier()

**pred\_dt = DT.predict(xv\_test)**

**DT.score(xv\_test, y\_test)**

**OUTPUT**

0.995632798573975

**print(classification\_report(y\_test, pred\_dt))**

**OUTPUT**

precision recall f1-score support

0 0.99 1.00 1.00 5864

1 1.00 0.99 1.00 5356

accuracy 1.00 11220

macro avg 1.00 1.00 1.00 11220

weighted avg 1.00 1.00 1.00 11220

**# Gradient Boosting Classifier**

**from sklearn.ensemble import GradientBoostingClassifier**

**GBC = GradientBoostingClassifier(random\_state=0)**

**GBC.fit(xv\_train, y\_train)**

**OUTPUT**

GradientBoostingClassifier(random\_state=0)

**pred\_gbc = GBC.predict(xv\_test)**

**GBC.score(xv\_test, y\_test)**

**OUTPUT**

0.9953654188948307

**print(classification\_report(y\_test, pred\_gbc))**

**OUTPUT**

precision recall f1-score support

0 1.00 0.99 1.00 5864

1 0.99 1.00 1.00 5356

accuracy 1.00 11220

macro avg 1.00 1.00 1.00 11220

weighted avg 1.00 1.00 1.00 11220

**# Random Forest Classifier**

**from sklearn.ensemble import RandomForestClassifier**

**RFC = RandomForestClassifier(random\_state=0)**

**RFC.fit(xv\_train, y\_train)**

**OUTPUT**

RandomForestClassifier(random\_state=0)

**pred\_rfc = RFC.predict(xv\_test)**

**RFC.score(xv\_test, y\_test)**

**OUTPUT**

0.9908199643493761

**print(classification\_report(y\_test, pred\_rfc))**

**OUTPUT**

precision recall f1-score support

0 0.99 0.99 0.99 5864

1 0.99 0.99 0.99 5356

accuracy 0.99 11220

macro avg 0.99 0.99 0.99 11220

weighted avg 0.99 0.99 0.99 11220

**# Model Testing**

**def output\_lable(n):**

**if n == 0:**

**return "Fake News"**

**elif n == 1:**

**return "Not A Fake News"**

**def manual\_testing(news):**

**testing\_news = {"text":[news]}**

**new\_def\_test = pd.DataFrame(testing\_news)**

**new\_def\_test["text"] = new\_def\_test["text"].apply(wordopt)**

**new\_x\_test = new\_def\_test["text"]**

**new\_xv\_test = vectorization.transform(new\_x\_test)**

**pred\_LR = LR.predict(new\_xv\_test)**

**pred\_DT = DT.predict(new\_xv\_test)**

**pred\_GBC = GBC.predict(new\_xv\_test)**

**pred\_RFC = RFC.predict(new\_xv\_test)**

**return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Prediction: {}".format(output\_lable(pred\_LR[0]), output\_lable(pred\_DT[0]),**

**output\_lable(pred\_GBC[0]),**

**output\_lable(pred\_RFC[0])))**

**news = str(input())**

**manual\_testing(news)**

OUTPUT

Donald Trump Sends Out Embarrassing New Year’s Eve Message; This is Disturbing,"Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his enemies, haters and the very dishonest fake news media. The former reality show star had just one job to do and he couldn t do it. As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year, President Angry Pants tweeted. 2018 will be a great year for America! As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year. 2018 will be a great year for America! Donald J. Trump (@realDonaldTrump) December 31, 2017Trump s tweet went down about as welll as you d expect.What kind of president sends a New Year s greeting like this despicable, petty, infantile gibberish? Only Trump! His lack of decency won t even allow him to rise above the gutter long enough to wish the American citizens a happy new year! Bishop Talbert Swan (@TalbertSwan) December 31, 2017no one likes you Calvin (@calvinstowell) December 31, 2017Your impeachment would make 2018 a great year for America, but I ll also accept regaining control of Congress. Miranda Yaver (@mirandayaver) December 31, 2017Do you hear yourself talk? When you have to include that many people that hate you you have to wonder? Why do the they all hate me? Alan Sandoval (@AlanSandoval13) December 31, 2017Who uses the word Haters in a New Years wish?? Marlene (@marlene399) December 31, 2017You can t just say happy new year? Koren pollitt (@Korencarpenter) December 31, 2017Here s Trump s New Year s Eve tweet from 2016.Happy New Year to all, including to my many enemies and those who have fought me and lost so badly they just don t know what to do. Love! Donald J. Trump (@realDonaldTrump) December 31, 2016This is nothing new for Trump. He s been doing this for years.Trump has directed messages to his enemies and haters for New Year s, Easter, Thanksgiving, and the anniversary of 9/11. pic.twitter.com/4FPAe2KypA Daniel Dale (@ddale8) December 31, 2017Trump s holiday tweets are clearly not presidential.How long did he work at Hallmark before becoming President? Steven Goodine (@SGoodine) December 31, 2017He s always been like this . . . the only difference is that in the last few years, his filter has been breaking down. Roy Schulze (@thbthttt) December 31, 2017Who, apart from a teenager uses the term haters? Wendy (@WendyWhistles) December 31, 2017he s a fucking 5 year old Who Knows (@rainyday80) December 31, 2017So, to all the people who voted for this a hole thinking he would change once he got into power, you were wrong! 70-year-old men don t change and now he s a year older.Photo by Andrew Burton/Getty Images.",News,"December 31, 2017"

LR Prediction: Fake News

DT Prediction: Fake News

GBC Prediction: Fake News

RFC Prediction: Fake News

news = str(input())

manual\_testing(news)

OUTPUT

U.S. military to accept transgender recruits on Monday: Pentagon,"WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. military starting on Monday as ordered by federal courts, the Pentagon said on Friday, after President Donald Trump’s administration decided not to appeal rulings that blocked his transgender ban. Two federal appeals courts, one in Washington and one in Virginia, last week rejected the administration’s request to put on hold orders by lower court judges requiring the military to begin accepting transgender recruits on Jan. 1. A Justice Department official said the administration will not challenge those rulings. “The Department of Defense has announced that it will be releasing an independent study of these issues in the coming weeks. So rather than litigate this interim appeal before that occurs, the administration has decided to wait for DOD’s study and will continue to defend the president’s lawful authority in District Court in the meantime,” the official said, speaking on condition of anonymity. In September, the Pentagon said it had created a panel of senior officials to study how to implement a directive by Trump to prohibit transgender individuals from serving. The Defense Department has until Feb. 21 to submit a plan to Trump. Lawyers representing currently-serving transgender service members and aspiring recruits said they had expected the administration to appeal the rulings to the conservative-majority Supreme Court, but were hoping that would not happen. Pentagon spokeswoman Heather Babb said in a statement: “As mandated by court order, the Department of Defense is prepared to begin accessing transgender applicants for military service Jan. 1. All applicants must meet all accession standards.” Jennifer Levi, a lawyer with gay, lesbian and transgender advocacy group GLAD, called the decision not to appeal “great news.” “I’m hoping it means the government has come to see that there is no way to justify a ban and that it’s not good for the military or our country,” Levi said. Both GLAD and the American Civil Liberties Union represent plaintiffs in the lawsuits filed against the administration. In a move that appealed to his hard-line conservative supporters, Trump announced in July that he would prohibit transgender people from serving in the military, reversing Democratic President Barack Obama’s policy of accepting them. Trump said on Twitter at the time that the military “cannot be burdened with the tremendous medical costs and disruption that transgender in the military would entail.” Four federal judges - in Baltimore, Washington, D.C., Seattle and Riverside, California - have issued rulings blocking Trump’s ban while legal challenges to the Republican president’s policy proceed. The judges said the ban would likely violate the right under the U.S. Constitution to equal protection under the law. The Pentagon on Dec. 8 issued guidelines to recruitment personnel in order to enlist transgender applicants by Jan. 1. The memo outlined medical requirements and specified how the applicants’ sex would be identified and even which undergarments they would wear. The Trump administration previously said in legal papers that the armed forces were not prepared to train thousands of personnel on the medical standards needed to process transgender applicants and might have to accept “some individuals who are not medically fit for service.” The Obama administration had set a deadline of July 1, 2017, to begin accepting transgender recruits. But Trump’s defense secretary, James Mattis, postponed that date to Jan. 1, 2018, which the president’s ban then put off indefinitely. Trump has taken other steps aimed at rolling back transgender rights. In October, his administration said a federal law banning gender-based workplace discrimination does not protect transgender employees, reversing another Obama-era position. In February, Trump rescinded guidance issued by the Obama administration saying that public schools should allow transgender students to use the restroom that corresponds to their gender identity. ",politicsNews,"December 29, 2017 "

LR Prediction: Not A Fake News

DT Prediction: Not A Fake News

GBC Prediction: Not A Fake News

RFC Prediction: Not A Fake News

**Please Provide Code through Git Hub Repo Link**