A

Project Report

On

FAKE NEWS DETECTION SYSYTEM

By:

JAI GROVER

DEPARTMENT OF COMPUTER APPLICATIONS

JSS Academy of Technical Education Noida

College Code – 091 C-20/1 Sector-62, Noida Uttar Pradesh

[May’2020]

A

Project Report

On

FAKE NEWS DETECTION SYSYTEM

In partial fulfillment of requirements for degree of

Master of Computer Applications

SUBMITTED BY:

Jai Grover

Under the Guidance of

Sir.Hanumanta Rao

DEPARTMENT OF COMPUTER APPLICATIONS

JSS Academy of Technical Education, Noida

College Code – 091 C-20/1 Sector-62, Noida Uttar Pradesh

[May'2020]

CERTIFICATE

This is to certify that the project entitled “FAKE NEWS DETECTION SYSTEM” has Been Carried out by the Jai Grover under my guidance in partial fulfillment of the degree of Master of Computer Applications of Uttar Pradesh Technical University Lucknow during the period Jan’2020-May’2020.

Name of the student:

Jai Grover

Date:

Place: Noida,U.P

Internal Guide:

Sir Hanumantha Rao

PAGE INDEX

1.INTRODUCTION

This section gives a scope description and overview of everything included in this Report. Also, the purpose for this document is described and definitions is provided.

1.1. Purpose: The topic of fake news detection on social media has recently attracted tremendous attention. The basic countermeasure of comparing websites against a list of labeled fake news sources is inflexible, and so a machine learning approach is desirable. Our project aims to detect fake news directly, based on the text content of news articles.

1.2. Scope: The scope of this project is very diverse, it ranges from various websites , news , medias to fake blogs, fake websites that deceive the users in one way or the other.

1.3. Definitions: Fake news is news articles that are intentionally and verifiably false and could mislead readers . There are two key features of this definition: *authenticity* and *intent*. First, fake news includes false information that can be verified as such. Second, fake news is created with dishonest intention to mislead consumers. Broader definitions of fake news focus on the either authenticity or intent of the news content.

2.THE OVERALL DESCRIPTION

Window Based Application which is work over the command prompt for input ,output , and browse the data. The user wants to determine about news is fake or real. So user browse the news and gives over on cmd as a input for finding the news is real or fake which comes as an output.

2.1. Product Functions: Product (Fake news Detection System) Function is to determine the news is fake or real. The user will be able to determine about news by using the WBA. The user will search the news coppy from the user and paste over on cmd as an input after press an enter will get an output ,and output will comes either one is fake or real.

2.2. User Characteristics:

There is one type of user which is interacts with the system : User(itself)

The users can use the Window Based Application to find a news. This means that the user have to be able to search for news is fake or real, choose a news from that browser and then copy from the browser and paste on cmd. In order for the users to get a relevant result .

2.3. Constraints:

The Internet connection is also a constraint for the application. Since the user fetches data from the browser Internet, it is crucial that there is an Internet connection for the user to search for the input.

3.SPECIFIC REQUIREMENTS

3.1. External Interfaces:

3.1.1. User Interfaces: Command Prompt

3.1.2. Hardware Interfaces: Laptop/Personal Computer , Internet

3.1.3. Software Interfaces: Jupyter(IDE) , Anaconda Prompt , Web Browser

3.2. Dataset Requirements: Dataset is required for giving training to the model , and filtering the dataset by using feature extraction , first is Tfdvectorizer , second is Countvectorizer . and Using Naïve Byes model for easily predict the news is fake or real and for better accuracy.

3.3. Software System Attributes:

The requirements in this section specify the required reliability, availability, security and maintainability of the software system.

3.4.1. Reliability:

1. The reliability that the system gives the right result on a search.

2. Measurements obtained from 1000 searches during testing.

3.4.2. Availability:

1. The availability of the system when it is used.

2. The average system availability (not considering network failing).

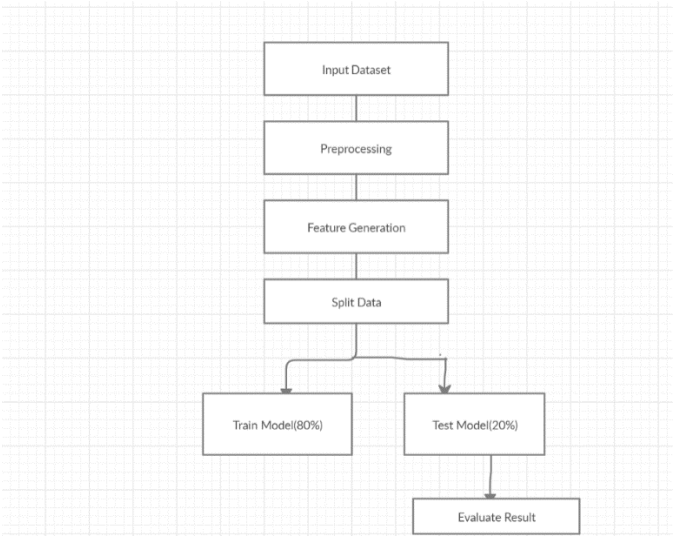
3.4.3. Maintainability:

1. The application should be easy to extend. The code should be written in a way that it favors implementation of new functions.

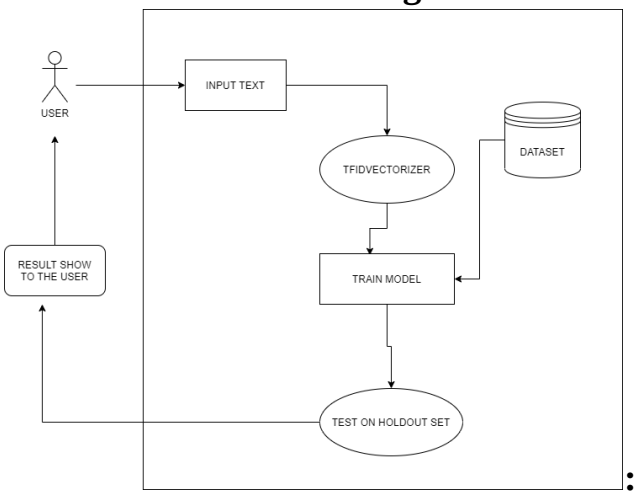
2.In order for future functions to be implemented easily to the application.

3.5. List Of Diagrams :

3.5.1. DFD:

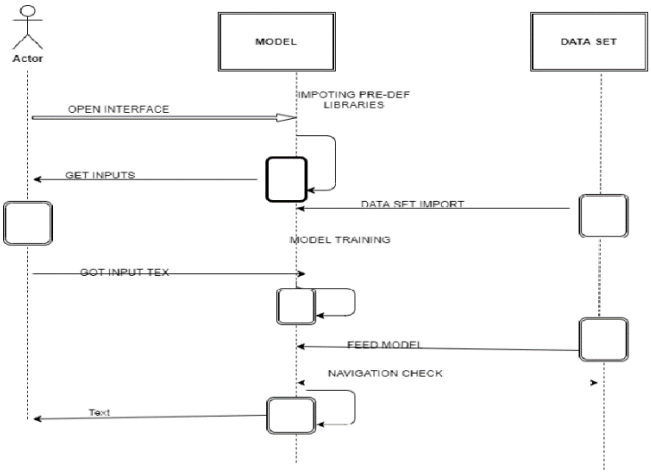


3.5.2. Use Case Diagram



:

3.5.3. Sequence Diagram:

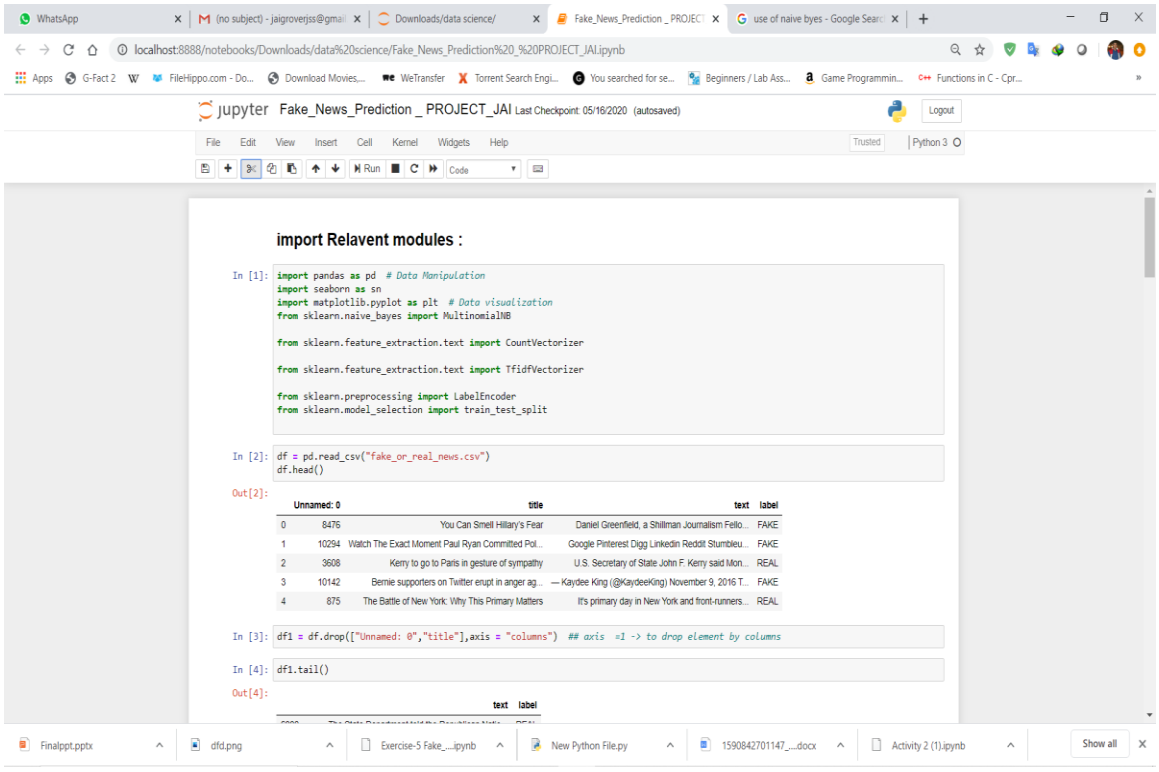


4. IMPLEMENTATION RESULTS

```
E:\new>python "New Python File.py"
Enter a News Headline : The highest circulating print newspapers in the United States include USA Today, The Wall Street
 Journal, The New York Times, and The Washington Post. Each newspaper's corresponding website has varying levels of acce
ss. Many newspapers rely on content from The Associated Press and Reuters. There is also a large number of local newspap
ers—more than 1,300 in the United States alone according to the Newspaper Association of America.
Results:  ['FAKE']

E:\new>
```

5. SOURCE CODE



jupyter Fake_News_Prediction _ PROJECT_JAI Last Checkpoint: 05/16/2020 (autosaved)

import Relavent modules :

```
In [1]: import pandas as pd  # Data Manipulation
        import seaborn as sn
        import matplotlib.pyplot as plt  # Data visualization
        from sklearn.naive_bayes import MultinomialNB

        from sklearn.feature_extraction.text import CountVectorizer

        from sklearn.feature_extraction.text import TfidfVectorizer

        from sklearn.preprocessing import LabelEncoder
        from sklearn.model_selection import train_test_split
```

```
In [2]: df = pd.read_csv("fake_or_real_news.csv")
        df.head()
```

Out[2]:

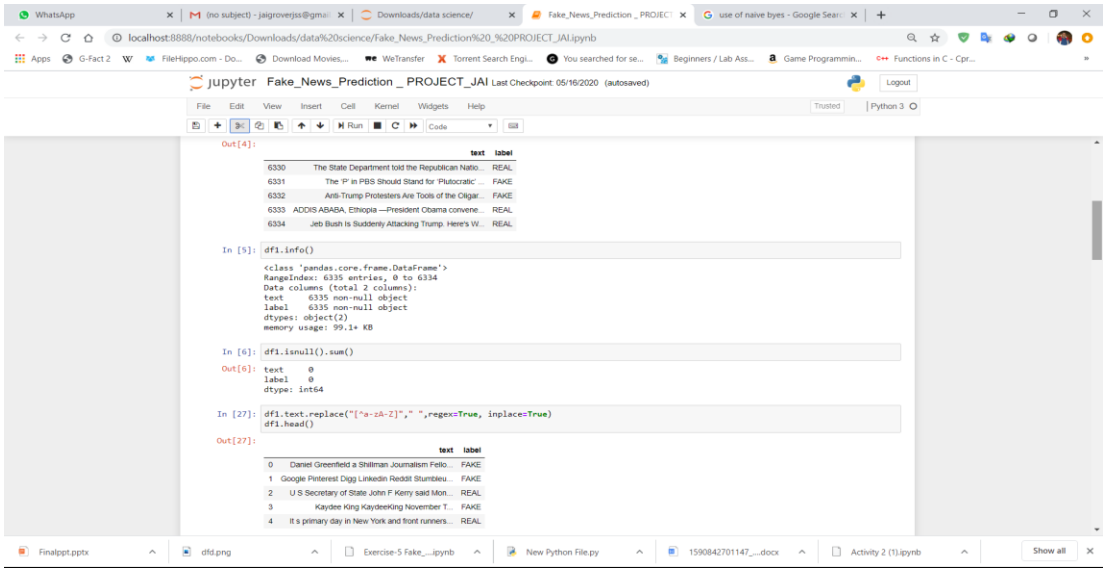
	Unnamed: 0	title	text	label
0	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	10294	Watch The Exact Moment Paul Ryan Committed Pol...	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	3608	Kerry to go to Paris in gesture of sympathy	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	10142	Bernie supporters on Twitter erupt in anger ag...	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	875	The Battle of New York: Why This Primary Matters	It's primary day in New York and front-runners...	REAL

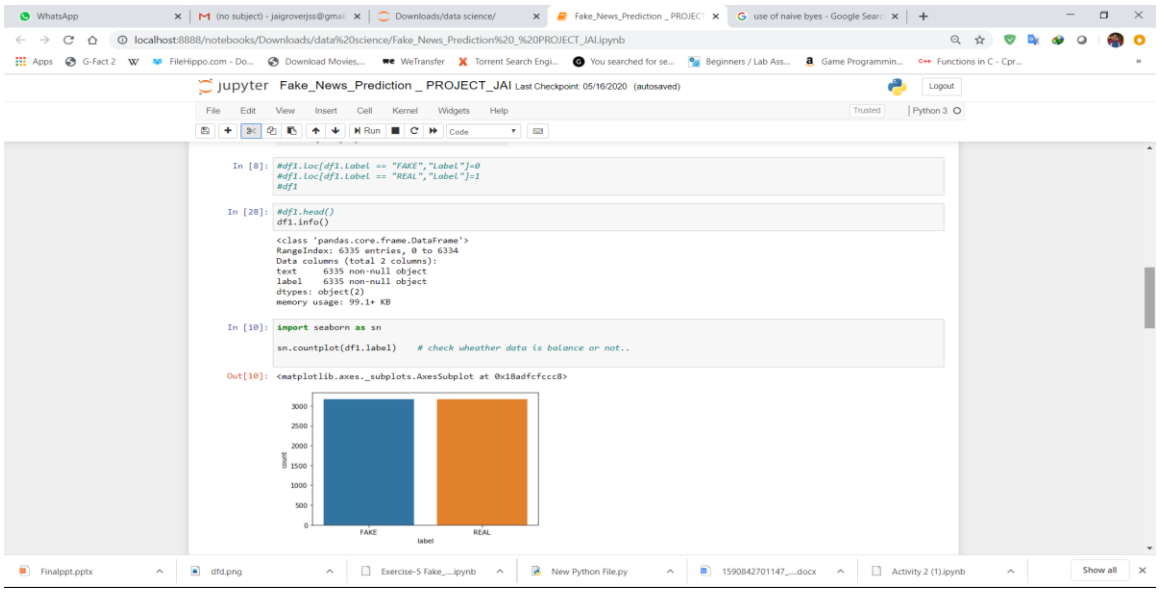
```
In [3]: df1 = df.drop(["Unnamed: 0","title"],axis = "columns")  ## axis  =1 -> to drop element by columns
```

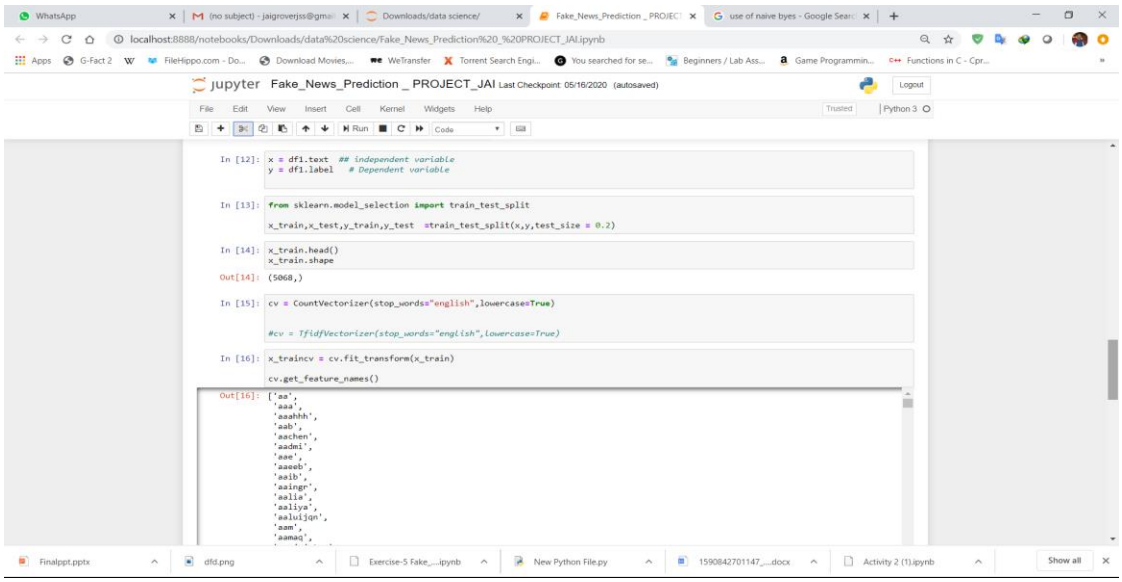
```
In [4]: df1.tail()
```

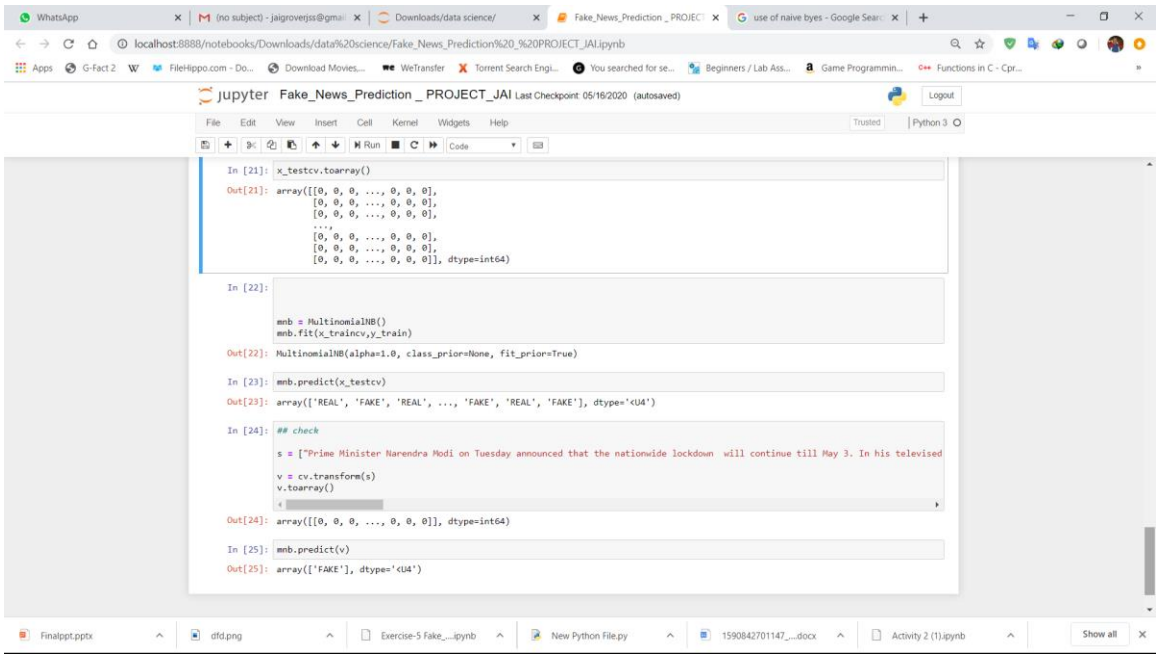
Out[4]:

	text	label


jupyter Fake_News_Prediction _ PROJECT_JAI Last Checkpoint: 05/16/2020 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
Out[4]:
text label
6330 The State Department told the Republican Natio... REAL
6331 The 'P' in PBS Should Stand for 'Plutocratic' ... FAKE
6332 Anti-Trump Protesters Are Tools of the Oligar... FAKE
6333 ADDIS ABABA, Ethiopia —President Obama convene... REAL
6334 Jeb Bush Is Suddenly Attacking Trump. Here's W... REAL
In [5]: df1.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6335 entries, 0 to 6334
Data columns (total 2 columns):
text 6335 non-null object
label 6335 non-null object
dtypes: object(2)
memory usage: 99.1+ KB
In [6]: df1.isnull().sum()
Out[6]: text 0
label 0
dtype: int64
In [27]: df1.text.replace("[^a-zA-Z]"," ",regex=True, inplace=True)
df1.head()
Out[27]:
text label
0 Daniel Greenfield a Shillman Journalism Fello... FAKE
1 Google Pinterest Digg Linkedin Reddit Stumbleu... FAKE
2 U S Secretary of State John F Kerry said Mon... REAL
3 Kaydee King KaydeeKing November T... FAKE
4 It s primary day in New York and front runners... REAL


jupyter Fake_News_Prediction _ PROJECT_JAI Last Checkpoint: 05/16/2020 (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted Python 3
In [8]: #df1.loc[df1.Label == "FAKE","Label"]=0
#df1.loc[df1.Label == "REAL","Label"]=1
#df1
In [28]: #df1.head()
df1.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6335 entries, 0 to 6334
Data columns (total 2 columns):
text 6335 non-null object
label 6335 non-null object
dtypes: object(2)
memory usage: 99.1+ KB
In [10]: import seaborn as sn
sn.countplot(df1.label) # check wheather data is balance or not..
Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x18adfcfccc8>
3000
2500
2000
1500
1000
500
0
count
FAKE
REAL
label





6.LIMITATIONS AND FUTURE ENHANCEMENT

Future Scope

There is always a scope for enhancements in any developed system, especially when our nature of the project is iterative which allows us to rethink on the method of development to adopt changes in the project. Below mentioned are some of the changes possible in

thefuture to increase the adaptability, and efficiency of the system

• Increase the dataset

• Increase the processing speed.

• Try to bring the domain as close as possible to the real world.

• Quality of dataset can be improved.

Limitations :

Fake News is not something that is new however, as technology evolves and advances over time, the detection of Fake News also becomes more challenging as social media continues to dominate our everyday lives and hence accelerating the speed of which Fake News travel .In a recent study published by the journal Science, it analysed millions of tweets sent between 2006 and 2017 and it was found that: “Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information.” It also concluded that “it took the truth about six times as long as falsehood to reach 1,500 people.” Also other than just the sheer speed of how fast fake news travel, it is also more challenging to detect it simply because of how attractive most fake news articles are titled as.

Fake news detection is still a challenge even to deep learning methods such as Convolutional Neural Network (CNN), Recurrent neural network (RNN) and etc. because the content of fake news is planned in a way it resembles the truth so as to deceive readers; and without cross referencing and fact checking, it is often difficult to determine veracity by text analysis alone.

7. REFERENCES

1.Marco L, E. Tacchini, S. Moret, G. Ballarin, “Automatic Online Fake News Detection Combining Content and

Social Signals,” [Online]. Available:

https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017.

2.H. Shaori, W. C. Wibowo, “Fake News Identification Chatacteristics Using Named Entity Recognition and Phrase

Detection,” 2018, 10th ICITEE, Universitas Indonesia.

1.https://patentimages.storage.googleapis.com/6c/34/81/c390e0d0b7a340/US8185448.pdf

2.https://patents.google.com/patent/CA2984904A1/en?oq=CA2984904A1

3.https://link.springer.com/chapter/10.1007/978-3-319-69155-8_9