

**FAKE NEWS DETECTION USING MACHINE LEARNING WITH NLP**

**BY**

**MD. MESBAH HOSSAIN**

**ID: 182-15-11461**

This Report Presented in Partial Fulfillment of the Requirements for the  
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Most. Hasna Hena**

Assistant Professor

Department of CSE

Daffodil International University

Co-Supervised By

**Mr. Narayan Ranjan Chakraborty**

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**MAY 2021**

## **APPROVAL**

This Project titled “**Fake News Detection Using Machine Learning with NLP**”, submitted by \***Md. Mesbah Hossain**\* to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on \***May 31, 2021**\*

## **BOARD OF EXAMINERS**

**Chairman**



---

**Dr. Touhid Bhuiyan**

**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University



---

**Nazmun Nessa Moon**

**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

*Aniruddha Rakshit*

---

**Aniruddha Rakshit**

**Senior Lecturer**

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

**Internal Examiner**

*Dr. Mohammad Shorif Uddin*

---

**Dr. Mohammad Shorif Uddin**

**Professor**

Department of Computer Science and Engineering

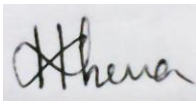
Jahangirnagar University

**External Examiner**

## DECLARATION

I hereby declare that, this project has been done by me under the supervision of Most. Hasna Hena, **Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**



---

**Most. Hasna Hena**  
Assistant Professor  
Department of CSE  
Daffodil International University

**Co-Supervised by:**



---

**Mr. Narayan Ranjan Chakraborty**  
Assistant Professor  
Department of CSE  
Daffodil International University

**Submitted by:**



---

**Md. Mesbah Hossain**  
ID: -182-15-11461  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project successfully.

I really grateful and wish my profound my indebtedness to **Most. Hasna Hena**, Assistant Professor, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of “*Machine Learning*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to **Professor Dr. Touhid Bhuiyan**, Head, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank my entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

## **ABSTRACT**

With the advancement of technology, fake news has become a common trend. It has the ability to cause harm to citizens, protests, and even a threat to democracy. It gives agitators the ability to spread fake news or false propaganda, triggering civil unrest. Since then, identifying fake news has become a critical concern, with every company attempting to develop a system for detecting and combating it. Fake news has had a negative impact on society in terms of politics and culture. It has had a negative impact on both online and offline social networking networks, as well as groups and conversations. In this study, I propose a model for detecting false news that combines NLP and machine learning techniques. To determine which model produces the best results, I compare various classifier models and extraction techniques. Introduction, motivation, rationality of the study, research questions, expected output, terminologies, related works, comparative analysis, challenges, research methodology, statistical analysis, applied mechanism, results and discussion are briefly written here. Impact on society, environment and sustainability are also explained here. Tables and Figures are included and listed. Then pertinent information is repeated in the summary section for convenience. At the end of the report an annotated reference list is included for ease in finding other useful guidance.

**LIST OF TABLES**

<b>TABLES</b>	<b>PAGE NO</b>
Table 1: Various Fake News Detection datasets	6
Table 2: Quantity of various types of news	8
Table 3: Accuracy scores of different models	14

**LIST OF FIGURES**

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 1: Inserting required libraries	9
Figure 2: Dropping unnecessary columns	9
Figure 3: Convert text into vectors using TFIDF	10
Figure 4: Logistic Regression	11
Figure 5: Decision Tree Classifier	11
Figure 6: Gradient Boosting Classifier	12
Figure 7: Random Forest Classifier	12
Figure 8: Detecting result of Fake News	14
Figure 9: Detecting result of Real News	15



## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Approval Page	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
List of Tables	v
List of Figures	vi
<b>CHAPTER</b>	
<b>CHAPTER 1: INTRODUCTION</b>	<b>1-3</b>
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Output	2
1.6 Report Layout	3
<b>CHAPTER 2: BACKGROUND</b>	<b>5-7</b>
2.1 Preliminaries/Terminologies	5
2.2 Related Works	5
2.3 Comparative Analysis and Summary	6
2.4 Scope of the Problem	7

2.5 Challenges	7
<b>CHAPTER 3: RESEARCH METHODOLOGY</b>	<b>8-13</b>
3.1 Research Subject and Instrumentation	8
3.2 Data Collection Procedure/Dataset Utilized	8
3.3 Statistical Analysis	8
3.4 Proposed Methodology/Applied Mechanism	9
3.5 Implementation Requirements	13
<b>CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION</b>	<b>14-15</b>
4.1 Experimental Setup	14
4.2 Experimental Results & Analysis	14
4.3 Discussion	15
<b>CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY</b>	<b>16-17</b>
5.1 Impact on Society	16
5.2 Impact on Environment	16
5.3 Ethical Aspect	16
5.4 Sustainability Plan	17
<b>CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH</b>	<b>18-19</b>
6.1 Summary of the Study	18
6.2 Conclusions	18
6.3 Implication for Further Study	18

<b>APPENDIX</b>	<b>20</b>
<b>REFERENCES</b>	<b>21</b>
<b>PLAGIARISM REPORT</b>	<b>23</b>

## **CHAPTER 1**

### **Introduction**

#### **1.1 Introduction**

Fake news and misinformation have become more prevalent in journalism, news reporting, social media, and other types of online information consumption as our society has become increasingly digitized. False information from these sources has led to a slew of issues, including the stoking of unfounded fears during medical outbreaks like Ebola. [1] The spread and repercussions of fake news are worsening as a result of the rise of prominent social media apps and other online outlets with insufficient fact-checking or third-party filtering, allowing anyone to broadcast fake news quickly and widely. While efforts to track and remove fake news, it continues to pose a serious threat to society. According to a Pew Research Center report [2], half of Americans consider fake news to be a serious issue, ranking it higher than violent crime. Furthermore, according to the survey, 68 percent of Americans believe fake news has a major effect on their faith in the government, while 54 percent believe it has a strong impact on their trust in one another. As a result, research into the identification of false news is critical for society.

The use of social media for news consumption has two sides. On the one hand, people search out and consume news from social media because of its low cost, easy access, and rapid distribution of information. On the other hand, it facilitates the widespread dissemination of "fake news," or low-quality news that contains deliberately false facts. The extensive spread of fake news has the potential for extremely negative impacts on individuals and society.

To build a fake news detection model, I'll need a lot of data. There are many published datasets in this field of research. Some dataset contains fewer samples; it is limited in its ability to contribute to the study of fake news. The dataset I propose to overcome these limitations faced by short datasets; from Kaggle [3] I collect two recent datasets for detecting fake news that includes more than 40 thousand samples.

## **1.2 Motivation**

Our dream a peaceful society where is no crime. There are various types of people in our society. Many of the crimes create by spreading rumor using fake news. Some people share and some people believe it. Major of the news are contains political news and government. We can able to remove this problem using fake news detection. People can able to check the news as real or fake. Then he/she can decide to believe it or not. That's why I am interested in the study of fake news detection.

## **1.3 Rationale of the Study**

There are many fake news detection projects. Researchers are very interested in this topic. But maximum of those projects failed to get maximum accuracy. Then it failed to give accurate results. That's why I use some different techniques in this study; I wish it gives better result.

## **1.4 Research Questions**

Is it possible to detect fake news?

Is proposed method an improvement over the current methods?

How feasible it is to implement proposed method?

How will people benefited from the method?

## **1.5 Expected Output**

- Able to detect fake or real news.
- Better improvement of the existing methods.
- People will get benefited.

## **1.6 Report Layout**

Approval Page

Declaration

Acknowledgement

Abstract

List of Tables

List of Figures

### **Chapter 1: Introduction**

- Introduction
- Motivation
- Rationale of the Study
- Research Questions
- Expected Output
- Report Layout

### **Chapter 2: Background**

- Preliminaries/Terminologies
- Related Works
- Comparative Analysis and Summary
- Scope of the Problem
- Challenges

### **Chapter 3: Research Methodology**

- Research Subject and Instrumentation
- Data Collection Procedure/Dataset Utilized
- Statistical Analysis
- Proposed Methodology/Applied Mechanism
- Implementation Requirements

### **Chapter 4: Experimental Results and Discussion**

- Experimental Setup
- Experimental Results & Analysis

- Discussion

### **Chapter 5: Impact on Society, Environment and Sustainability**

- Impact on Society
- Impact on Environment
- Ethical Aspects
- Sustainability Plan

### **Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research**

- Summary of the Study
- Conclusions
- Implication for Further Study

### **Appendix**

### **References**

## CHAPTER 2

### Background

#### 2.1 Preliminaries/Terminologies

Fake news has existed for almost as long as news has been widely circulated since the printing press was invented in 1439[4]. However, there is no consensus about what constitutes "fake news". The task of detecting fake news, which consists of deliberate misinformation or hoaxes spread through traditional news media or online social media, is known as fake news detection. Detecting fake news on social media presents a number of novel and difficult research challenges. Though fake news is not a new problem—nation and groups have used the news media to carry out propaganda and manipulate operations for centuries—the rise of web-generated news on social media has amplified the power of fake news, challenging conventional journalistic norms.

#### 2.2 Related Works

Since political language on TV interviews, Facebook updates, and Twitter posts is mainly short sentences, detecting fake news is more difficult than detecting misleading reviews. Fake news can have a wide-ranging negative impact, and it can also influence or even exploit major public events. For example, during the final three months of the 2016 US presidential election, fake news created to benefit one of the two candidates was believed by many people and shared over 37 million times on Facebook [5]. There has been a significant amount of research into the characteristics of fake news. For example, Jin et al. [6] used multimedia datasets to analyze news article images for fake news detection. They looked at a variety of visual and statistical picture features to see whether they could predict the veracity of different posts. They also proposed a method for detecting fake news that relies on a reputation dissemination network created by analyzing competing opinions extracted from tweets. Yang et al. [7] suggested a model for detecting fake news early using a multivariate time series to distinguish news dissemination paths.



A variety of datasets for fake news detection have been published in recent years. These are listed in Table 1, along with their specific characteristics.

TABLE 1: VARIOUS FAKE NEWS DETECTION DATASETS

Dataset	Size (# of samples)	Source	Data Category
LIAR	12,836	Politifact	Political
BUZZFEEDNEWS	2,282	Facebook	Political
BUZZFACE	2,263	Facebook	Political
Some-like-it-hoax	15,500	Facebook	Scientifix/conspiracy
PHEME	330	Twitter	Variety
Breaking	700	BS Detector	Political

A few patterns emerge when comparing fake news datasets. The majority of the datasets are small, which can make current machine learning models that require large amounts of training data ineffective. Furthermore, many datasets divide their data into a limited number of groups, such as false vs. real. More fine-grained labels can be found in datasets like LIAR (Wang, 2017) [8], and Fake News Corpus. Many datasets include data from particular fields, such as politics and celebrity gossip (Tacchini et al., 2017) [9]; (Abu Salem et al., 2019) [10]; (Santia and Williams, 2018) [11], while others have data from a number of categories (Zubiaga et al., 2016) [12]. Because of the small number of categories, these data samples may have minimal meaning and writing styles. From an NLP perspective, researchers have studied numerous aspects of credibility of online information. For example, [13] & [14] they used TF-IDF. Then some models and classifiers apply on it. [15] Uses Confusion Matrix.

### 2.3 Comparative Analysis and Summary

Above mentioned work have less efficiency. With the combination of TF-IDF Vectorizer and Logistic Regression it achieved an efficiency of approximately 85%. But we need more efficiency for better result.

I plan to work by analyzing and perform not only similar procedures, but also including some other procedures using different datasets from Kaggle [3] for improving efficiency;

it will help expand the spectrum and improving accuracy. I want to use maximum model & classifier.

## **2.4 Scope of the Problem**

There are many types of news in journalism and social media. This is very tough to classify that news as real or fake for readers. News publishes several times. There are huge amount of news data. This is very hard to work with all news data without making dataset. I have need dataset. There are some datasets those contains mixed news, some datasets are only for fake and only for real news.

## **2.5 Challenges**

Information authenticity has long been a problem for companies and culture, impacting both printed and digital media. On social networks, the influence and effects of information distributed are so rapid and amplified that skewed, misleading, or false information has a huge potential to have real-world consequences for millions of users within minutes. To remove this problem administrator need to develop an effective system. Ensure better efficiency, better accuracy and accurate result. Our challenge, better expected outcomes from the study.

## CHAPTER 3

### Research Methodology

#### 3.1 Research Subject and Instrumentation

Our research subject is Fake News Detection Using Machine Learning with NLP. I used here 2 datasets from kaggle [3]; one is fake news dataset and another one is real/true news dataset. I apply on the dataset TFIDF from NLP. Also use Jupyter Notebook for code implementations.

#### 3.2 Data Collection Procedure/Dataset Utilized

I collected dataset from kaggle[3]. I used here 2 datasets; fake and true.

#### 3.3 Statistical Analysis

Here I use 23,481 fake news data and 21,417 real news data. There are various types of news showing in Table 2:

TABLE 2: QUANTITY OF VARIOUS TYPES OF NEWS

Subject	Quantity
politicsNews	11,272
Worldnews	10,145
Governmnet News	1,570
Middle-east	778
News	9,050
US_News	783
Left-news	4,459
Politics	6,841
Total	44,898

Maximum used news is political news. Real and fake both; by spreading fake news there causes violence. So, it is more important for political news to detect fake news. Mainly in election season many fake news spread.

### 3.4 Proposed Methodology/Applied Mechanism

First of all I import some necessary libraries of Python and Machine Learning showing in figure 1.

#### Importing Required Libraries

```
In [2]: import pandas as pd
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import accuracy_score
        from sklearn.metrics import classification_report
        import re
        import string
```

Figure 1: Inserting required libraries.

Inserting Fake and Real Datasets. Then classify both different dataset. I removed last 10 rows from both dataset for manual testing. Again classify both manual testing dataset and merge both in single dataset and save it in a csv file. Merge main fake and real data frame to apply further methods. I need only 2 types of data for analysis, news data and class. So, now drop other columns from dataset showing in figure 2.

**"title", "subject" and "date" columns is not required for detecting the fake news,**

```
In [142]: df = df_merge.drop(["title", "subject", "date"], axis = 1)
          df.head()
```

Out[142]:

	text	class
0	Donald Trump just couldn't wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0

Figure 2 : Dropping unnecessary columns.

I set the 2 variable as x and y. Here, x is independent variable and y is dependent variable. Then splitting the dataset into training set and testing set. Checking null values

from the dataset. Then randomly shuffling the data frame and reset index. I use here a function Regular Expression to convert the text in lowercase; remove extra space, special character, url and links.

Text is converted to feature vectors, which can then be used as input to an estimator. Vocabulary is a dictionary that translates each token (word) into a feature index in the matrix, with a feature index for each specific token. The numbers (weights) in each vector reflect the TFIDF score features. Now I used TFIDF to convert text into vectors showing in figure 3.

## Convert text to vectors ¶

```
In [38]: from sklearn.feature_extraction.text import TfidfVectorizer
```

```
In [39]: vectorization = TfidfVectorizer()
          xv_train = vectorization.fit_transform(x_train)
          xv_test = vectorization.transform(x_test)
```

Figure 3: Convert text into vectors using TFIDF.

After completing vectorization I apply 4 classifier from 3 models; From linear model I use Logistic Regression, from decision tree I use Decision Tree Classifier, from ensemble model I use Gradient Boosting Classifier & Random Forest Classifier. Those are showing on figure 4, figure 5, figure 6 and figure 7 respectively.

## Logistic Regression

```
In [40]: from sklearn.linear_model import LogisticRegression
```

```
In [41]: LR = LogisticRegression()
```

```
In [42]: LR.fit(xv_train, y_train)
```

```
Out[42]: LogisticRegression()
```

```
In [64]: pred_lr = LR.predict(xv_test)
```

```
In [65]: pred_lr = LR.predict(xv_train)
```

```
In [66]: LR.score(xv_train, y_train)
```

```
Out[66]: 0.9905736360110968
```

```
In [67]: LR.score(xv_test, y_test)
```

```
Out[67]: 0.9869351230425056
```

Figure 4 : Logistic Regression.

## Decision Tree Classification

```
In [45]: from sklearn.tree import DecisionTreeClassifier
```

```
In [46]: DT = DecisionTreeClassifier()  
DT.fit(xv_train, y_train)
```

```
Out[46]: DecisionTreeClassifier()
```

```
In [68]: pred_dt = DT.predict(xv_train)
```

```
In [69]: pred_dt = DT.predict(xv_test)
```

```
In [75]: DT.score(xv_train, y_train)
```

```
Out[75]: 0.9999701697342124
```

```
In [76]: DT.score(xv_test, y_test)
```

```
Out[76]: 0.9943624161073825
```

Figure 5: Decision Tree Classifier.

## Gradient Boosting Classifier

```
In [85]: from sklearn.ensemble import GradientBoostingClassifier
```

```
In [84]: GBC = GradientBoostingClassifier()
GBC.fit(xv_train,y_train)
```

```
Out[84]: GradientBoostingClassifier()
```

```
In [89]: pred_gbc = GBC.predict(xv_train)
```

```
In [90]: pred_gbc = GBC.predict(xv_test)
```

```
In [91]: GBC.score(xv_train,y_train)
```

```
Out[91]: 0.9972854458133222
```

```
In [92]: GBC.score(xv_test,y_test)
```

```
Out[92]: 0.9948993288590604
```

Figure 6 : Gradient Boosting Classifier.

## Random Forest Classifier

```
In [94]: from sklearn.ensemble import RandomForestClassifier
```

```
In [95]: RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train,y_train)
```

```
Out[95]: RandomForestClassifier(random_state=0)
```

```
In [101]: pred_rfc = RFC.predict(xv_train)
```

```
In [102]: pred_rfc = RFC.predict(xv_test)
```

```
In [103]: RFC.score(xv_train,y_train)
```

```
Out[103]: 0.9999701697342124
```

```
In [104]: RFC.score(xv_test,y_test)
```

```
Out[104]: 0.9902460850111857
```

Figure 7: Random Forest Classifier.

### **3.5 Implementation Requirements**

To analyze required dataset I need to apply some methods. To get expected outcomes also need it. Not only apply, I should ensure proper utilization of the methods. Our data utilization process effects on the accuracy. I have got the better performance from the system. I use TFIDF from NLP to convert text data into vector. I use 4 different classifiers to ensure better accuracy and more accurate results.



## CHAPTER 4

### Experimental Results and Discussion

#### 4.1 Experimental Setup

I separate 20 data from the total dataset for manual testing. Whole mechanism I would apply to the main dataset.

#### 4.2 Experimental Results & Analysis

After running several simulations, following are the results of different combinations of models used in our study. Now I mention accuracy scores of different models on Table 3,

TABLE 3 : ACCURACY SCORES OF DIFFERENT MODELS

Models/Classifier	Accuracy Score
Leaner Model/Logistic Regression	0.986
Decision Tree/Decision Tree Classifier	0.994
Ensemble/Gradient Boosting Classifier	0.994
Ensemble/Random Forest Classifier	0.990

I also got result for fake and real news, showing on figure 8 and figure 9 respectively.

```
In [74]: news = str(input())
manual_testing(news)
```

Donald Trump just signed the GOP tax scam into law. Of course, that meant that he invited all of his craven, cruel GOP sycophants down from their perches on Capitol Hill to celebrate in the Rose Garden at the White House. Now, that part is bad enough celebrating tax cuts for a bunch of rich hedge fund managers and huge corporations at the expense of everyday Americans. Of course, Trump is beside himself with glee, as this represents his first major legislative win since he started squatting in the White House almost a year ago. Thanks to said glee, in true Trumpian style, he gave a free-wheeling address, and a most curious subject came up as Trump was thanking the goons from the Hill. Somehow, Trump veered away from tax cuts, and started talking about the Congressional baseball shooting that happened over the summer. In that shooting, Rep. Steve Scalise, who is also the House Majority Whip, was shot and almost lost his life. Thanks to this tragic and stunning act of political violence, Scalise had a long recovery; in fact he is still in physical therapy. But, of course, vain and looks-obsessed Trump decided that he would congratulate Scalise, not on his survival and on his miraculous recovery, but on the massive amount of weight Scalise lost while he was practically dying. And make no mistake Scalise is VERY lucky to be alive. According to doctors, when he arrived at the hospital, Scalise was actually, quote, in imminent risk of death. Here is the quote, via Twitter: How stunningly tone deaf does one have to be to say something like that? I never thought I'd say this about a Republican that I, by all reasonable accounts, absolutely loathe, but I feel sorry for him. I am sorry he got shot, and I am even sorer that he now has to stand there and listen to that orange buffoon talk about him like that. I am sure that Scalise is a much tougher man than Trump, though. I am equally sure that he also knows that Trump is an international embarrassment and a crazy man who never should have been allowed anywhere near the White House. Featured image via Alex Wong/Getty Images

LR Prediction: Fake News  
DT Prediction: Fake News  
GBC Prediction: Fake News  
RFC Prediction: Fake News

Figure 8 : Detecting result of Fake news.

```

news = str(input())
manual_testing(news)

(Reuters) - A lottery drawing to settle a tied Virginia legislative race that could shift the statehouse balance of power has been indefinitely postponed, state election officials said on Tuesday, after the Democratic candidate mounted a legal fight. The decision to put off the high-stakes lotto, originally scheduled for Wednesday, marks the latest twist in a dramatic election recount that at one point showed Democrat Shelly Simonds beating Republican incumbent David Yancey by a single vote. A victory by Simonds would shift Republicans' slim control of the 100-member House of Delegates to an even 50-50 split with the Democrats, forcing the two parties into a rare power-sharing arrangement. A day after Simonds emerged as the victor of a recount, a three-judge panel ruled that a disputed ballot should be counted for Yancey. That decision left the two candidates tied with 11,608 votes each in a district that encompasses the shipping hub of Newport News in southeastern Virginia, setting the stage for the equivalent of a coin toss to pick a final winner. Simonds asked a state court to reconsider on Tuesday, arguing that the disputed ballot was wrongly included. An image filed in court showed that the ballot had bubbles filled in beside both names, with a slash mark by Simonds' name. The voter selected Republicans for other offices. Simonds told reporters that the case had implications not only for her contest but for the integrity of state elections as a whole, saying that without a court ruling in her favor, "recounts would become a never-ending spiral of courtroom challenges." The chairman of the Virginia Board of Elections, James Alcorn, said in a statement that while holding a lottery would be in keeping with state law, such a move should be considered "an action of last resort." He added: "Any substantive concerns regarding the election or recount should be resolved before a random drawing is conducted." Yancey's campaign did not immediately respond to requests for comment. The Virginia House Republican Caucus said in a statement that it was reviewing the new court filings. "We believe the court acted appropriately and that the integrity of the process is without question," spokesman Parker Slaybaugh said. Virginia Department of Elections spokeswoman Andrea Gaines said in an email that no new date for a drawing has been set. Democrats notched historic gains in Virginia's statehouse elections last month, part of the party's first big wave of political victories since Republican Donald Trump won the White House last year. Before the Nov. 7 general election, Virginia Republicans held 66 seats to the Democrats' 34 in the House of Delegates, along with a majority in the state Senate.

LR Prediction: Real News
DT Prediction: Real News
GBC Prediction: Real News
RFC Prediction: Real News

```

Figure 9 : Detecting result of Real news.

### 4.3 Discussion

I have taken two different csv files as input. After merging and saving, with the choice of self-verification, above mentioned models assessed its credibility.

## **CHAPTER 5**

### **Impact on Society, Environment and Sustainability**

#### **5.1 Impact on Society**

It's only normal that when we come across misleading information on social media, we feel compelled to point it out or disagree with it. However, based on my research, this could do more harm than good. It might seem counterintuitive, but the best way to respond to fake news – and mitigate its effect – may be to do nothing. On social media, false information is a major issue. Online disinformation, according to a UK parliament committee, is a challenge to "the very fabric of our democracy." It has the potential to manipulate and deepen social tensions. People should try to reduce the exposure of false information to reduce its effects. All should try to stay away from spreading false information. That means that instead of only adding a warning label, social media platforms should consider fully eliminating false information. And it suggests that the safest course of action for individual social media users is to avoid engaging with fake information altogether. So, we can say, Fake News Detection does a big positive impact for a peaceful society.

#### **5.2 Impact on Environment**

Climate change is one topic where fake news is especially prevalent and effective. Messages casting doubt on the reality of anthropogenic climate change (Oreskes 2011) [16], (Farrell et al 2019) [17]. So, we can say, Fake News Detection does a big positive impact for a nice environment.

#### **5.3 Ethical Aspects**

Spreading fake news and rumor, both are crime. Sharing fake news lack of understanding some people engaged with this crime. Before sharing now using Fake News Detection they can decide news as real or fake.

## **5.4 Sustainability Plan**

We need updated dataset. New news publishes regularly. For consistency for the application we have need to update dataset regularly and then train the dataset again.

## CHAPTER 6

### Summary, Conclusion, Recommendation and Implication for Future Research

#### 6.1 Summary of the Study

Study is the thing which has no end. Previously there many study of research about Fake News Detection I got. Because, it's a very important research topic. People all time read news over online. But sometimes wrong information spreading by fake news as rumor. I try to solve and reduce the problem doing the study. Relevant clarifications are needed to help guide the potential directions of fake news detection study. Social media has proven to be an effective tool for spreading fake news. In social media, there are several emerging trends that can be used to identify fake news. A basic understanding of the state-of-the-art fake news detection methods can be gained by reviewing current fake news detection methods in different social media scenarios. Fake news identification on social media is still in its infancy, and there are still a slew of problems that need to be investigated further. It's important to talk about future research avenues for improving fake news identification and mitigation.

#### 6.2 Conclusions

The combination of TF-IDF Vectorizer and four different classifiers: Logistic Regression, Decision Tree Classifier, Gradient Boosting Classifier & Random Forest Classifier improved the efficiency of our model than before. I got 98-99% accuracy.

#### 6.3 Implication for Further Study

Different types of data characteristics, such as dataset, temporal, and psychological, are being studied in data-driven fake news detection. I demonstrated that there is no current benchmark dataset that includes tools to extract all applicable features from a dataset perspective. The development of a comprehensive and large-scale fake news benchmark dataset, which researchers can use to promote more research in this field, is a promising path. From a temporal perspective, the spread of false news on social media exhibits

©Daffodil International University

distinct temporal trends from that of real news. Along these lines, one intriguing issue is early fake news identification, which seeks to provide early warnings of fake news during its distribution.

**Appendix:**

Full code you can found on github repository. [18] There I used only jupyter notebook.

## Reference:

- [1] Learn about Wikipedia/Ebola, available at << <https://en.wikipedia.org/wiki/Ebola>>>, last accessed on 01-05-2021 at 12:00 PM.
- [2] Learn about Pew Research Center, available at << <https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/>>>, last accessed on 01-05-2021 at 12:00 PM.
- [3] Learn about Kaggle/Datasets, available at <<<https://www.kaggle.com/clmentbisailon/fake-and-real-news-dataset?select=Fake.csv>>>, last accessed on 01-05-2021 at 12:00 PM.
- [4] Learn about politico, available at << <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535>>>, last accessed on 01-05-2021 at 12:00 PM.
- [5] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives*, vol. 31, no. 2, 2017, pp. 211–36.
- [6] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, “Novel visual and statistical image features for microblogs news verification,” *IEEE transactions on multimedia*, vol. 19, no. 3, 2016, pp. 598–608.
- [7] Y. Liu and Y.-F. B. Wu, “Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks,” in *Thirty-Second AAAI Conference on Artificial Intelligence, 2018*.
- [8] Wang, W. Y. (2017). “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 422–426, Vancouver, Canada, July. Association for Computational Linguistics.
- [9] Tacchini, E., Ballarin, G., Vedova, M. L. D., Moret, S., and de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. CoRR, abs/1704.07506.
- [10] Abu Salem, F. K., Al Feel, R., Elbassuoni, S., Jaber, M., and Farah, M. (2019). Fa-kes: A fake news dataset around the syrian war. Proceedings of the *International AAAI Conference on Web and Social Media*, 13(01):573–582, Jul.
- [11] Santia, G. C. and Williams, J. R. (2018). Buzzface: A news veracity dataset with facebook user commentary and egos. In Twelfth International AAAI Conference on Web and Social Media.
- [12] Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., and Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989.
- [13] Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal, “On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification”, arXiv:1712.03935v1 [cs.CL] 11 Dec 2017
- [14] Shivam B. Parikh and Pradeep K. Atrey, “Media-Rich Fake News Detection: A Survey”, *2018 IEEE Conference on Multimedia Information Processing and Retrieval*, p.p. 436-441.



- [15] Pranav Ashtaputre, Ashutosh Nawale, Rohit Pandit, “A Machine Learning Based Fake News Content Detection Using NLP”, *International Journal of Advanced Science and Technology*, Vol. 29, No. 7, (2020), pp. 11219-11226.
- [16] Anker, P. (2011). Naomi Oreskes;, Erik M. Conway. Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. 355 pp., bibl., index. New York: Bloomsbury Press, 2010. \$27 (cloth). Isis, 102(3), 589–590. doi:10.1086/663066
- [17] Farrell, J., McConnell, K., & Brulle, R. (2019). Evidence-based strategies to combat scientific misinformation. *Nature Climate Change*. doi:10.1038/s41558-018-0368-6
- [18] Learn about github code, available at <<<https://github.com/mesbahhsabbir/FakeNewsDetection>>>, last accessed on 01-05-2021 at 12:00 PM.

# PLAGIARISM REPORT

fake\_news\_detection.docx

## Turnitin Originality Report

Processed on: 02-May-2021 10:50 +06  
ID: 1575617635  
Word Count: 3017  
Submitted: 1

fake\_news\_detection By Hasna Hena

Similarity Index	Similarity by Source
24%	Internet Sources: 13% Publications: 12% Student Papers: 12%

3% match (Internet from 25-Mar-2021) <a href="http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5250/162-15-7955%20%287_%29.pdf?isAllowed=y&amp;sequence=1">http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5250/162-15-7955%20%287_%29.pdf?isAllowed=y&amp;sequence=1</a>
3% match (Internet from 18-Jul-2020) <a href="https://slidelegend.com/fake-news-detection-on-social-media-arxiv_5ac317711723dd46ae1d66a6.html">https://slidelegend.com/fake-news-detection-on-social-media-arxiv_5ac317711723dd46ae1d66a6.html</a>
3% match (Internet from 20-May-2020) <a href="https://deepal.org/publication/r-fakeddit-a-new-multimodal-benchmark-dataset-for-fine-grained-fake-news-detection">https://deepal.org/publication/r-fakeddit-a-new-multimodal-benchmark-dataset-for-fine-grained-fake-news-detection</a>
2% match (student papers from 11-Mar-2021) <a href="#">Submitted to De Montfort University on 2021-03-11</a>
1% match (student papers from 08-Oct-2020) <a href="#">Submitted to QA Learning on 2020-10-08</a>
1% match (Internet from 13-Apr-2021) <a href="http://ns2.thinkmind.org/download_full.php?instance=IMMM+2019">http://ns2.thinkmind.org/download_full.php?instance=IMMM+2019</a>
1% match (student papers from 29-Apr-2021) <a href="#">Submitted to University of North Texas on 2021-04-29</a>
1% match (student papers from 18-Apr-2018) <a href="#">Submitted to Universiti Teknologi MARA on 2018-04-18</a>

1% match () <a href="#">Hamidian, Sardar, Diab, Mona T. "Rumor Detection and Classification for Twitter Data". 2019</a>
1% match (student papers from 05-Apr-2021) <a href="#">Submitted to University of Greenwich on 2021-04-05</a>
1% match (student papers from 07-Oct-2017) <a href="#">Submitted to The University of Memphis on 2017-10-07</a>
1% match (Internet from 07-May-2016) <a href="http://ijicic.org/ijicic-14-03010.pdf">http://ijicic.org/ijicic-14-03010.pdf</a>
< 1% match (Internet from 07-Apr-2021) <a href="http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5207/162-15-7914%20%2825_%29.pdf?isAllowed=y&amp;sequence=1">http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5207/162-15-7914%20%2825_%29.pdf?isAllowed=y&amp;sequence=1</a>
< 1% match (Internet from 07-Apr-2021) <a href="http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5149/162-15-7878%20%2818_%29.pdf?isAllowed=y&amp;sequence=1">http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5149/162-15-7878%20%2818_%29.pdf?isAllowed=y&amp;sequence=1</a>
< 1% match (Internet from 07-Apr-2021) <a href="http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5158/152-15-6060%20%2817_%29.pdf?isAllowed=y&amp;sequence=1">http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5158/152-15-6060%20%2817_%29.pdf?isAllowed=y&amp;sequence=1</a>
< 1% match (Internet from 04-Aug-2020) <a href="http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4092/P15438%20%286_%29_.pdf?isAllowed=y&amp;sequence=1">http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4092/P15438%20%286_%29_.pdf?isAllowed=y&amp;sequence=1</a>
< 1% match (Internet from 08-Apr-2018) <a href="https://arxiv.org/pdf/1708.01967.pdf">https://arxiv.org/pdf/1708.01967.pdf</a>
< 1% match (publications) <a href="#">L.E. Reed, "Performing a literature review". FIE '98. 28th Annual Frontiers in Education Conference. Moving from 'Teacher-Centered' to 'Learner-Centered' Education. Conference Proceedings (Cat. No.98CH36214). 1998</a>
< 1% match (publications) <a href="#">Deepak P. Tanmoy Chakraborty, Cheng Long, Santhosh Kumar G. "Data Science for Fake News". Springer Science and Business Media LLC, 2021</a>
< 1% match (Internet from 15-Jul-2020) <a href="https://au.answers.yahoo.com/question/index?qid=20101014130025AAV8bIA">https://au.answers.yahoo.com/question/index?qid=20101014130025AAV8bIA</a>