Fake News Detection

Manisha Gahirwal
manisha.gahirwal@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Sanjana Moghe
sanjana.moghe@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Tanvi Kulkarni
tanvi.kulkarni@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Devansh Khakhar
devansh.khakhar@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

Jayesh Bhatia
jayesh.bhatia@ves.ac.in
Vivekanand Education Society's Institute of Technology, Mumbai, Maharashtra

ABSTRACT

Fake news, one of the biggest new-age problems has the potential to mould opinions and influence decisions. The proliferation of fake news on social media and Internet is deceiving people to an extent which needs to be stopped. The existing systems are inefficient in giving a precise statistical rating for any given news claim. Also, the restrictions on input and category of news make it less varied. This paper proposes a system that classifies unreliable news into different categories after computing an F-score. This system aims to use various NLP and classification techniques to help achieve maximum accuracy.

Keywords: *Stance Detection, Natural Language Processing (NLP), Random Forest.*

1. INTRODUCTION

Fake news has been around for decades and is not a new concept. However, the dawn of the social media age which can be approximated by the start of the 20th century has aggravated the generation and circulation of fake news many folds. Fake news can be simply explained as a piece of article which is usually written for economic, personal or political gains. Detection of such bogus news articles is possible by using various NLP techniques, Machine learning, and Artificial intelligence.

In 2016, the Prime Minister of India, Mr. Narendra Modi declared that most of the cash that people possessed had become worthless and in the span of one month all this old currency had to be deposited in the banks. This led to a chain reaction of a series of fake news being published used mainly for click bait and political gains. News about the new paper bills having a GPS tracker or the daily limit of the amount that can be deposited in banks has increased, were spreading like wildfire. Now, this may not seem like a huge thing but the impact of such articles was so much that there was a point where the Ministry of the Finance had to officially release statements assuring citizens that what they were reading was false information. This is just a small instance of how the spread of false news can impact a much greater audience than it may seem.

This paper provides an insight into the procedure of detecting fake news, it's implementation and its results. In order to reach a conclusion on the authenticity of the news article we first take the text article, analyze the data and then use various classification algorithms to classify the news as legitimate or fake.

2. LACUNA IN THE EXISTING SYSTEM

I. BS Detector

BS Detector is a plug-in used by Mozilla and Chrome browsers to detect the presence of fake news sources and to alert the user accordingly.It works by searching through web pages references of links which have already been flagged unreliable in their database. BS Detector has been used by Facebook to solve their proliferation of fake news problem.But lately, they blocked the extension stating that they have been working on their own technique to curb the problem. BS Detector just states a warning message if the article is found to

be fake.It does not specify the percentage of error and neither does it classify news into levels of "truthfulness" or "fakeness".

II. Politi Fact

PolitiFact is a fact-checking US-based website used by editors and writers which gives the credibility of claims by US officials involved American politics.This system places judgment in the form of Truth-O-Meter which is a measure of the accuracy of a statement.These people first decide on which news to evaluate depending upon some characteristics like significance and worthiness of the claim.After that, the Truth-O-Meter is generated and a panel of more than two people thoroughly go through it to evaluate final rating of the claim.

The drawback of this system is that human intervention is required.Secondly, it works only for US politics.Also, every claim is not being fact-checked by them. The choice of evaluation depends upon them.

III. Flock Fake News Detector

Fake News Detector was a feature added by Flock-a new generation messaging and collaborative platform. Whenever links are being sent to each while chatting, FND algorithm gets activated.It checks the content of links to their databases of websites computed according to rankings. It gives a statistic rating and generates a warning message if the source is not found to be reliable.Flocks database has more than 600 news URL'S that are fact-checked.The drawback of this system is that their database for fact checking is less in number and chances of hoaxes still not being determined are high.

3. LITERATURE SURVEY

While there are some existing applications like BS Detector and Politifact which to some extent help users to identify misleading news but it requires human intervention and also the domain is limited in case of BS Detector which does not give the user the extent of any article to be fake.

In [1], they are using linguistic cues approaches and network analysis approaches to design a basic fake news detector which provides high accuracy in terms of classification tasks. They propose a hybrid system whose features like multi-layer linguistic processing, the addition of network behavior are included.In [2], they propose a method to detect online deceptive test by using a logistic regression classifier which is based on POS tags extracted from a corpus deceptive and truthful texts and achieves an accuracy of 72% which could be further improved by performing cross-corpus analysis of classification models and reducing the size of the input feature vector.
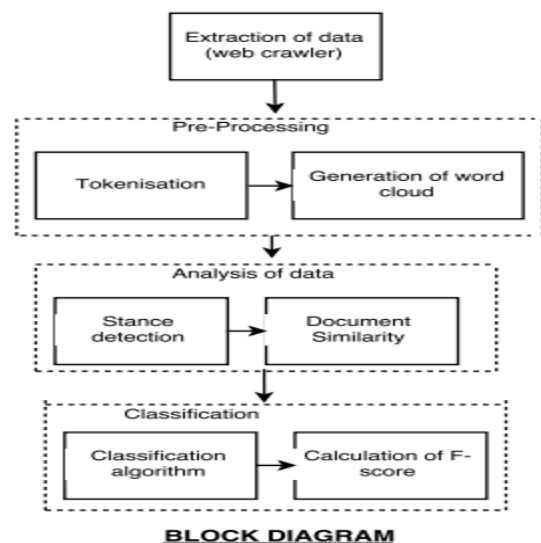
To detect fake news on social media, [3] presents a data mining perspective which includes fake news characterization on psychology and social theories. This article discusses two major factors responsible for widespread acceptance of fake news by the user which are Naive Realism and Confirmation Bias. Further, it proposes a two-phase general data mining framework which includes 1) Feature Extraction and 2) Model Construction and discusses the datasets and evaluation metrics for the fake news detection research. In [4], they propose an SVM-based algorithm with 5 predictive features i.e. Absurdity, Humour, and Grammar, Negative Affect, and Punctuation and uses satirical cues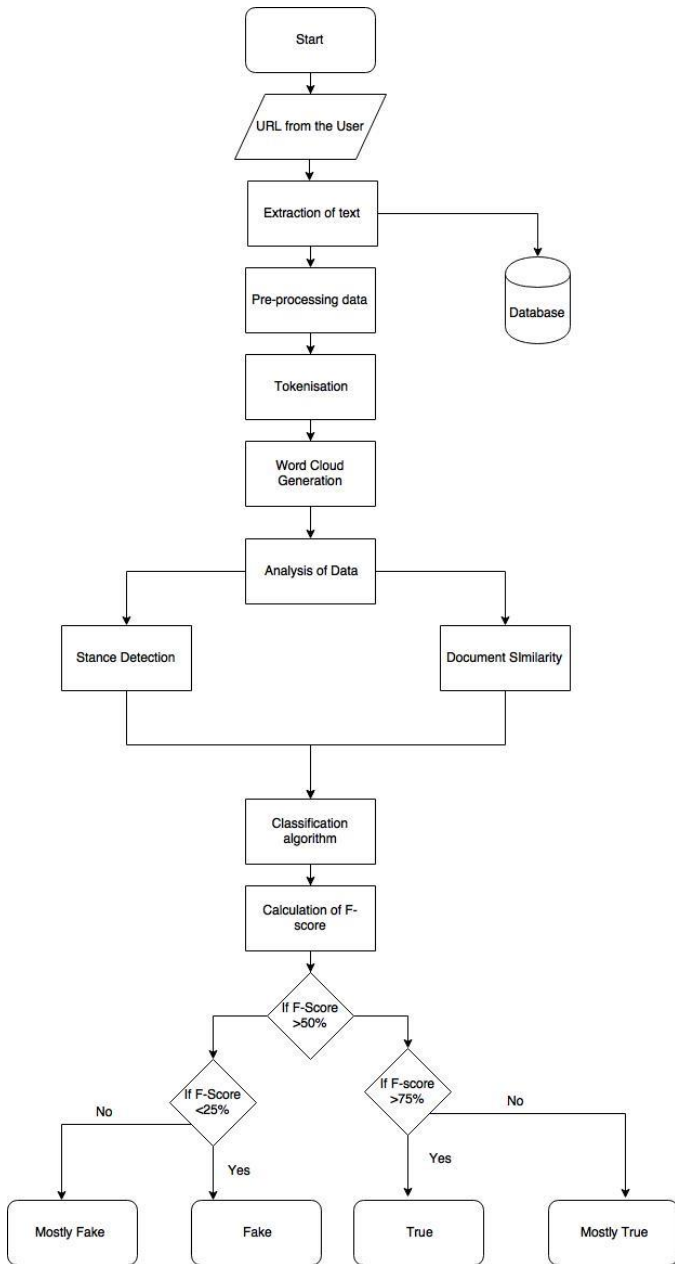 to detect misleading news. The paper translates theories of humor, irony, and satire into a predictive model for satire detection with 87% accuracy.

The purpose of this paper is to propose a new model for fake news detection which is using Stance Detection and IF-TDF method for analyzing the data which is taken from various datasets of fake and legitimate news and Random Forest classifier for classifying the output into four classes namely: True, Fake, Mostly True, and Mostly Fake. Using Random Forest gives us an advantage of handling binary features and moreover, they do not expect linear features.

4. METHODOLOGY

The aim of this project is to accurately determine the authenticity of the contents of a particular news article. For this purpose, we have devised a procedure which is intended to fetch favorable results. We first take the URL of the article that the user wants to authenticate, after which the text is extracted from the URL. The extracted text is then passed on to the data pre-processing unit. The data preprocessing unit consists of various processes like the Tokenization and Generation of the word cloud. The outputs from these processes play an important role in further analyzing the data. The core deciding factors that we use to determine the output of our project ie if a particular news article is fake or not are the stance of the article and comparison of the article with top google search results. The first method is by using stance detection to in order to analyze the stance of the author. Stance is a mental or an emotional position adopted by the author with respect to something. Stance detection is an important part if NLP and has wide applications.The stance of the author can be divided into various categories like Agree, Disagree, Neutral or Unrelated with respect to the title. Giving each of these categories weights can help us in the final conclusion of whether a news article is fake or not. The second method is to use document similarity or if-idf to know how similar a document is to top search results. This too can give us an insight into the authenticity of a news article. Next, we need to classify the output into various output classes for which we can use classification algorithms or regression models. The output classes can be true, mostly true, false, and mostly false or we can just present it with a number. For example, 68% true or the score is 7 out of 10 where 1 is completely true and 10 is completely false.

BLOCK DIAGRAM

Flowchart

are taking their steps towards preventing the spread of fake news. Our systems take input from an URL or an existing database and classify it to be true or fake.To implement this, various NLP and Machine Learning Techniques have to be used.The website can be used by any user having a browser and an internet connection to check the veracity of the new article.

6. REFERENCES

[1] Conroy, Niall & Rubin, Victoria & Chen, Yimin. (2015). Automatic Deception Detection: Methods for Finding Fake News. . USA
[2] Ball, L. & Elworthy, J. J Market Anal (2014) 2: 187. https://doi.org/10.1057/jma.2014.15
[3] Lu TC. Yu T., Chen SH. (2018) Information Manipulation and Web Credibility. In: Bucciarelli E., Chen SH., Corchado J. (eds) Decision Economics: In the Tradition of Herbert A. Simon's Heritage. DCAI 2017. Advances in Intelligent Systems and Computing, vol 618. Springer, Cham
[4] Rubin, Victoria & Conroy, Niall & Chen, Yimin & Cornwell, Sarah. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. . 10.18653/v1/W16-0802.

5. CONCLUSION

In the 21st century, the majority of the tasks are done online. Newspapers who were earlier preferred as hard-copies are now being substituted by applications like Facebook, Twitter, and news articles to be read online. The growing problem of fake news only makes things more complicated and tries to change or hamper the opinion and attitude of people towards use of digital technology. When a person is deceived by the real news two possible things happen.People start believing that their perceptions about a particular topic are true as assumed. Another problem is that even if there is any news article available which contradicts a supposedly fake one, people believe in the words which just support their thinking without taking in the measure the facts involved.

Thus, in order to curb the phenomenon, Google and Facebook