Fake News Detection

[1]Ishita chakraborty, [2]`Sadaf hasan, [3]Prof.Mohd Tahir, [4]Dr S.R.Ahmed
[1]UG student, [2]UG student, [3] Assistant Professor, [4]Associate professor
[1234]Department of computer science and engineering,
[1234]Anjuman College Of Engineering and Technology, Nagpur, Maharastra, India

Abstract

The rapid increase in number false information in social media feeds, news blogs, and online newspapers have made it challenging for us to identify which news are real and which news are falsely fabricated.Thus,in this paper, we focus on the identification of fake news from the real one. First,we extract the fake news from a relevant social media sites,then we compare it to a verified site named smhoaxslayer which is affiliated by the government to perform tasks regarding the identification,correction of the fake news.Our project compares the fake news from this website and gives an estimated idea about how close is the fake news to the real news.The Neighbour's algorithm is used to compare the real news with the fake news.

I. INTRODUCTION

These days fake news is being a rising dilemma due to the spreading of sarcastic articles and other fake news regarding political and other social issues. One of the earliest recorded fake news that came into account was spreaded by Octavian against Marc Anthony which led to the final war of the Roman Republic (32 BC to 30 BC). There were many parallels with the elections of today. Instead of using Twitter or other social accounts, Octavian used short slogans written on coins that for which Anthony was criticized unfairly for being a puppet of Egypt, disloyal to Rome, a philanderer and a drunk. Anthony was declared a traitor and Cleopatra was named the queen of Egypt.Fake news and lack of trust in the social media has been a growing problems which have unpleasant consequences in our society. The term 'fake news' became common word for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper,it is seeked to produce a model that can accurately predict the likelihood of a certain fake news article.

1.1 Fake news and Types

A further problem is that Fake News never looks the same and it becomes difficult to identify which type of fake news we wish to deal with.In fact we may need different implementations to deal with different kind of fake news.In order to solve this problem, it is necessary to have an understanding on what Fake News is. Fake News can be classified as:[6]

1) Click Bait: Shocking Headlines meant to generate clicks to increase ad revenue. Often times these stories are highly exaggerated and are totally false.
2) Sloppy reporting that fits an agenda – news that contains some grains of truth that are not fully verified, which are used to support a certain position or view.
3) Misleading news that's not based on facts, but supports an on-going narrative – news where there is no established baseline for truth, often where ideologies or opinions clash and unconscious biases come into play. Conspiracy theories tend to fall here!
4) Intentionally deceptive – news that has been fabricated deliberately to either make money through number of clicks, or to cause confusion or discontent or as sensationalist propaganda. These stories are tend to be distributed through fake news sites which are designed to look like 'real' news brands.They often employ videos and graphic images that have been manipulated in some way.

1.2 Dealing with Fake news

Machine learning algorithms, are a part of Artificial Intelligence, they have been successful for decades in dealing with spam email, by analysis of messages and texts and by determining how likely it is that a particular message is a real communication from an actual person – or a mass-distributed solicitation for some pharmacy firms or other misleading companies.

Based on this type of analysis in spam fighting, AI systems can evaluate how well a certain article's text, or a headline, by comparing it with the actual article which someone is sharing online. Another method is to examine similar articles to see whether other news media have any differing facts. Similar systems can also identify specific type of accounts and source websites that spread fake news.The best way to combat the spread of fake news also depends upon human at some extent. The societal consequences of fake news are greater political polarization, increased partisanship, and decrease of trust in mainstream media and government – are significant. If more people knew the importance of real news and its effects were that high, they might be more wary of information, particularly if it is more emotionally based, because that's one of the most effective way to get people's attention.When someone sees an enraging post, that person should better investigate the information, rather than sharing it immediately. The act of sharing also lends credibility to a post: When other people see it, they register that it was shared by someone they know and automatically trust at least a bit, and are less likely to notice whether the original source is questionable or in simpler sense -fake.A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum.

Fake news and lack of trust in the media are growing problems with huge unpleasant consequences in our society. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper,it is seeke to produce a model that can accurately predict the likelihood that a given article is fake news..

II. LITERATURE REVIEW

2.1 Overview

In this chapter, we would discuss and present the current state of the art in research for analyzing and detecting trustworthy information from Twitter (Social Networking Site). Section 2.1.1 discusses the research work to assess,detect, measure, quantify good quality news content from Twitter. In section 2.1.2, we would present the research done to characterize that how important is the role of Twitter during real world events. In the last section, we summarize the implications and research gaps in analyzing trustworthy real informations from Twitter during real-world events.

2.1.1 Quality Assessment of Content Posted on OSM

This section presents the research and analytical work done in the space of extracting and analyzing trustworthy and credible information from Twitter during real world events. One major challenge in consuming content from Twitter is that it is difficult to filter out good quality content from the large volume of content created and which can be fake. The quality of content on Twitter is polluted with the presence of phishing, spam, advertisements, fake images, rumors and inflammatory content.

Media such as Twitter, which is a micro-blog is more suited for dissemination and sharing news based information, since it is mostly public, and gives a bigger range of audience for the content posted. Hence, most of the work discussed in this survey, is centered around Twitter. Researchers have used various classical computational techniques such as classification, ranking, characterization and conducting user studies, to study the problem of trust on Twitter. Some of the researchers who applied variouskinds of classifiers (Naive Bayes, Decision Tree, SVM) to identify spam, phishing and not credible content on Twitter, using message, user, network and topic based features on Twitter.

a) Emergence of Twitter as a News Media

Computer science community dealing with the research has done critical and major analysis regarding the relevance of online social media, in particular Twitter, as news fabricating agent. Kwak et al. showed that how prominent was Twitter as a newsmedia,they showed that more than 85% topics discussed on Twitter are related to news. Their works highlighted the relationship between user specific parameters v/s the tweeting activity patterns,like analysis of the number of followers and followees v/s the tweeting (re-tweeting) numbers. Zhaoet al. in their work, used unsupervised modeling to compare the news topic from Twitter versus that on New York Times (a traditional news medium) . They showed that Twitter users are relatively less interested in real world news, still they are active in spreading news of important world events. Lu et al. showed how tweets related to the actual news event on Twitter can be mapped . The methods proposed act like novel event detection techniques.

2.1.2 Analyzing Twitter Data during Real-World Events

The posts and different activities on Twitter, impacts a larger audience and plays a vital role in various real world events. Role of a social media handle like Twitter has been constantly analysed by computer scientists, psychologists and sociologists for their actual impact in the real-world. Twitter has progressed from being merely a medium to share different users opinions; to an information sharing agent; to propagation and coordination of relief and response efforts. Some of the popular case studies analyzed by computer scientists have been, Twitter activities during elections, natural disasters (like hurricanes, wildfires, floods,earthquakes etc.), political and social uprisings (like Libya and Egypt crisis,the Brexit) and terrorist attacks (like Mumbai triple bomb blasts,Phulwama attack). Content and user activity patterns of Twitter during events have been analyzed for both positive and negativeopinions as per various aspects. Some of the problems studied that result in bad quality of data, presence of spam and phishing posts, content spreading rumors / fake news, privacy breach of users via the content shared by them and use of Twitter for propagation and hate among people. Researchers have used various machine learning, information retrieval, social network analysis and image and video analysis for the purpose of analyzing and characterizing Twitter usage during real-world events which helps to identifying a fake news from the real one.

III. IMPLEMENTATION DETAILS

Fake news analyzer tool intends to eradicate the spreading of Fake News circulating in the social media at some extent. Mainly the Fake News Analyzer Tool has being prepared by keeping machine learning on mind with the use of basic Web development methods like Html,CSS,PhP, JavaScript etc.Our main aim is to implement Neighbor's algorithm for the detection of the Fake News found on the social media handles.

Fake News Analyzer tool is basically a testing site which is connected with a HTML site and Smhoaxslayer i.e. a website dedicated to critically analyze and debunk false stories spread by social Medias. The First API is connects the Html page and the "Fake News Analyzer Tool" testing site. The second API works as a bridge between the Testing site and Smhoaxslayer.

In the project the news that is expected as to be fake is derived from a particular social media account like Twitter, Facebooketc. The data fetched from this sites are then fed into the database where the news are stored and compared.So,if data is required again it can be reused without wasting physical memory. Then, the data is then forwarded through an API to the Smhoaxslayer site.

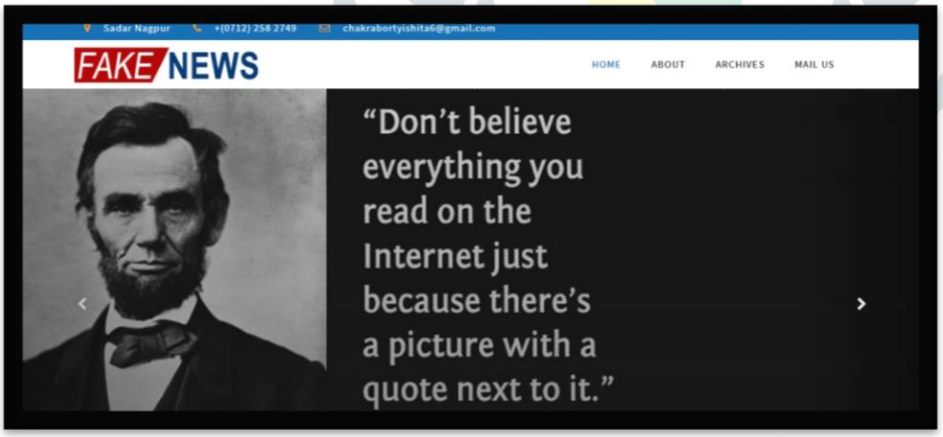
After that the database of the testing site, which has the news to be analyzed are compared with the Smhoaxslayer database .The data is then listed on a table to analyze whether it is 100% True or not.

ID	Date	News Headline	Content	% of the news to be true	True/False
123	12/22/34	Modi,Xi meets 6 times in 24 hours	Prime Minister of India Narendra Modi and Chinese president Xi will meet 6 times in 24 hours to discuss bilateral ties with China as reported by TOI.	40%	False

The modules associated with the project are based on basic web development technology which includes the implementation of HTML, PHP, CSS. The website consists of various details regarding fake news. They are the home page, archives, about us and contact details. There are four panels which holds the content details of the fake news.

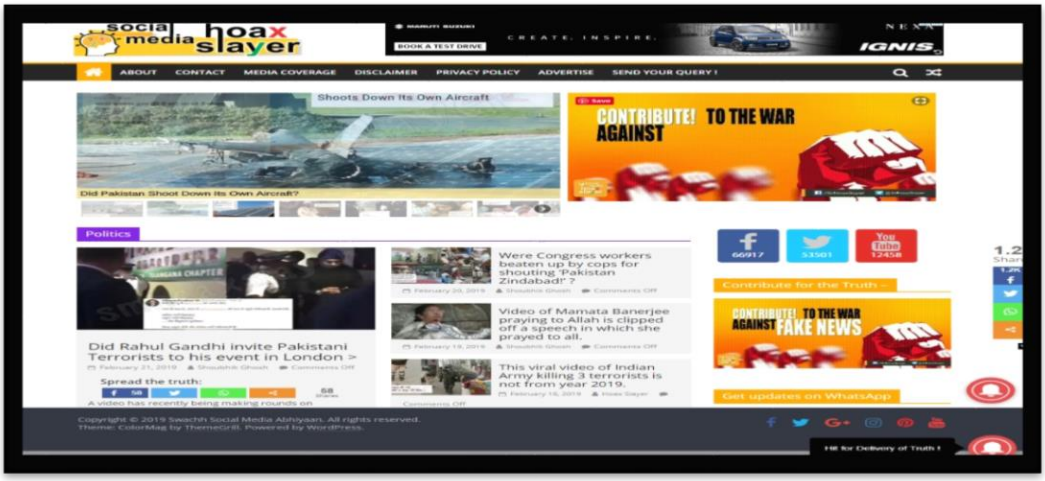
Module 1:

- The first module consists of the development of website.The website development along with the contents associated with the website.
- It consists of different tabs like Home,About,Archives and Mail Us.
- The different tabs perform different operations along with the information regarding the fake news.



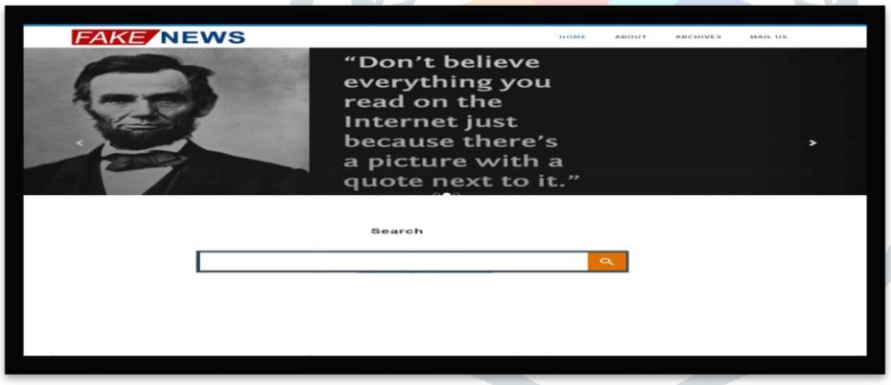
Module 2:

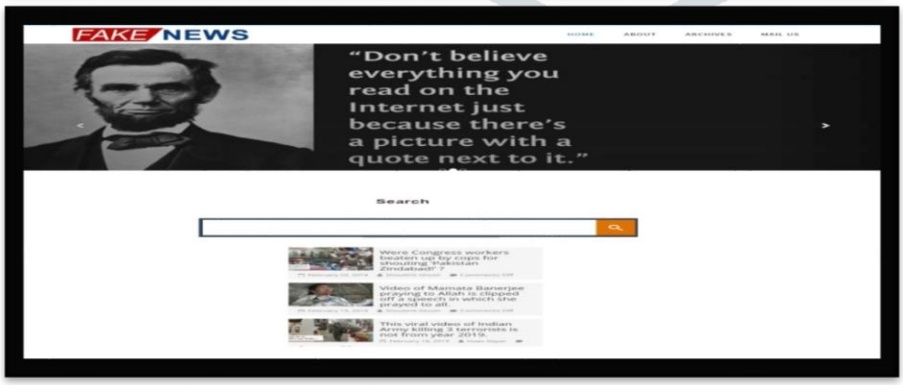
- The second module consists of the API generation i.e. taking the API of an authenticated site(in our project its smhoaxslayer site)
- The API generation and implementation is carried out.



Module 3:

- The third and the final module consists of the database and implementation of Neighbour's Algorithm for searching and comparing fake news with the real news.
- Implementation of Neighbour's Algorithm detecting fake news. In this project the Neighbour's Algorithm is used to compare the fake news with the real news.As all the data in the internet are uploaded mostly by humans and it can have human negligience,No news can be treated as 100% true until any witness visits the spot and verify the news(for this different group are assigned to visit the site of news and verify it).The nearest Neighbour's Algorithm can only detect how much close is the fake news to the real news.





V.CONCLUSION

In this paper we give a short but precise analysis towards understanding of fake news. Fake news is an integral part of the world and an important concept which may have serious real world consequences. Even though the scope of different type of fake news (ex: including satire or rumors as fake news), the challenges exist for the automatic detection of the actual type of fake news for all. The elimination mechanisms of fake news is an important step towards understanding and preventing the spread of misinformation. The importance of social media in the spread of fake news should not be underestimated as it has the maximum number of users. Deeper understanding of how a human psychology actually works on fake news could be helpful to develop tools for detection and prevention of misinformation. The existing methods for automatic fake news detection are mostly based on linguistic and machine learning techniques. In addition to these methods image analysis is applied. With the drastic increase in the popularity of the term fake news, the research towards automatic detection also has seen a rapid increase. The manual fact checking done by professional journalists give the researchers opportunity to understand the nature of misinformation and work more importantly towards the automatic detection of fake news.

REFERENCES

Journals:

1.Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1):1{4, 2015.

2. Wingfield, Nick, Mike Isaac, and Katie Benner. "Google and Facebook Take Aim at Fake News Sites." The New York 11 (2016)

Website:

1. https://en.wikipedia.org/wiki/Fake_news

2.https://www.kdnuggets.com/2017/10/guide-fake-news-detection-social-media.html.

3. https://www.cbsnews.com/pictures/dont-get-fooled-by-these-fake-news-sites/

4. https://www.fastcompany.com/section/fake-news

5. https://www.bbc.com/news/entertainment-arts-42242630