A Model to Measure Credibility of Information of Social Platform

Prajakta Vilas Chiddarwar[1], Prof. Dr. J. Naveen Kumar[2], Prof. Dr. M. S. Bewoor[3]
[1]Mtech student, [2]Assistant Professor, [3]Associate Professor,
[123]Bharati Vidyapeeth (deemed to be), University, College of Engineering, Pune.

Abstract- Social Media has grown rapidly through last 20 years, due to increase in involvement in online web sites. Because of this credibility of information is introduced in past decade. Fake news, data Misinformation, and incorrect data are crucial issues nowadays and these issues are affecting due to social interest and ease of access to social media. The growth in social media like Facebook, Twitter, Whatsapp, Blogs and many more , due to this people are getting attracted and the dissemination of information lead people to search more news. The dissemination of fake news may have low quality data or irrelevant data or any fake news which is spread over a link or via message to one another. This fake news can be intensely spread to create vulnerable environment or to damage the status of Politician. Online social media is suffering from number of fake accounts and dissemination of fake news from fake accounts. In the recent era of social media platform there are tremendous number of users and the usage of platforms. On the social media the receiver user is unaware about validity and credibility of news is being shared and unknowingly this misinformation can spread over the network. In this paper, I have survey research papers on the social, media scams and fake news and review to get better solution over the problem.

Keywords- fake news, user profile, trust analysis

I. INTRODUCTION

The credibility of information is a central issue of modern social platform. In recent years, the spreading of misinformation is growing rapidly primarily on politicians and social media influencers. The impact of fake news on social media is very problematic as we know that in 2016, in US Presidential Election the dissemination of fake news leads to most diverse decision, analyzed by the most viral news and fake message spread over a social media. After the survey it was viewed as the dissemination of fake news was more than the true and relevant news that has to be spread over the social media. It seems that the receivers are getting the advantages of web news in wrong manner. Rather to take it as a beneficiary the services are been use in wrong manner.

However, as we know that the providing news on online social media is cheap and much faster. When a source sends a message or a news on social media like whatsapp the receiver at another end is unaware about the relevance of news. These fake news and altered messages manipulates receivers to believe on it, also it changes the way people read and respond on real news.

Detecting fake news is on social media poses a new challenge. Sometimes the fake news is spread over a whatsapp group or a small platform which directly or indirectly bothers the credibility of information on social media. For example, if the sender sends a message of any politician or a social media influencer which is not true and the message spread from one end to another which leads to big issue on big platform. Because of this fake news detection concept is proposed to make the news unique and trustworthy at all end. To detect fake news different machine learning techniques are used.

1.1 Credibility of Information

Credibility refers to the term believability, how much the context is trustworthy and also the quality and accuracy of context to be shared over a social platform. Information can be called as trustworthy when it seems to be accurate, unbiased and fair. The information shared from source to destination should be constant in nature and should not be allow manipulation or alteration of information throughout the communication. Credibility of information depends on three factors that are source of message, the message itself and the platform shared the communication.

Previously, credibility was only check on offline context but in modern era, the social media providing premises to share information very easily and affordably which leads to dissemination of information widely over network. Also ease of manipulation and alteration of information leads information to be vulnerable. Information that shared on social platform can be of organizational, personal, political, because of this accuracy and expertise analysis is required to maintain credibility of context. Credibility refers to measure the accuracy and quality of context to be shared over a platform.

II. LITERATURE SURVEY

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.

"In this section, we briefly review the related work on fake news detection system and their different techniques.

In this paper [1], the results of a fake news identification study that documents the performance of a fake news classifier are presented. The Textblob, Natural Language, and SciPy Toolkits were used to develop a novel fake news detector. Advantages- 1. Used natural language processing2. Fake news detection based on attribute classification Disadvantages- Time consuming process.

This paper [2] introduce the datasets which contain both fake and real news and conduct various experiments to organize fake news detector. Advantages is 1. Used Natural Language Processing, Machine learning and deep learning techniques to classify the datasets 2.Accuracy is better and disadvantages is use Limited dataset.

This paper [3] proposed a distributed framework to implement the proposed truth discovery scheme using Work Queue in an HTCondor system. Advantages is 1. Find trustworthy information on Social media 2.Proposed truth discovery scheme using Work Queue in an HTCondor system and disadvantages is Accuracy is low

This Paper [4] Studied various detection techniques i.e. content based, social context based and hybrid based. Advantages is Proposed content-based, social context-based and hybrid-based methods and disadvantages is only survey state of the methods.

This paper [5] Present a new fake news detection model using unified key sentence information which can efficiently perform sentence matching between question and article by using key sentence retrieval based on bilateral multi perspective matching model. Advantages is Implement natural language processing using key sentence retrieval and disadvantages is Fake news detection accuracy is low.

This Paper [6] classifies fake news messages from Twitter posts using hybrid of convolutional neural networks and long-short term recurrent neural network models. Advantages is Implement hybrid CNN and RNN Models and Accuracy is much better. Disadvantages is only consider tweet headlines.

This paper [7] Compare news to other sources in 2016 year. Advantages is 1.detect 2016 election fake news spread through social media 2. Goal in this paper is to offer theoretical and empirical background to frame this debate. Disadvantages is 1. Limited dataset used 2. Limited to 2016 news only.

This paper [8] shows a new approach for fake news detection using naive Bayes classifier. Use Implement naïve bayes machine learning algorithm but accuracy is low.

This paper [9] introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. Advantages is in this paper, they explored the fake news problem by reviewing existing literature in two phases i.e. characterization and detection but on Use static data.

This study [10] contributes to the scientific knowledge regarding the influence of the interaction between various types of media use on political effects. Advantage is Used multiple news sources for fake news detection and disadvantage is Focus on only political data.

Sr No	Paper Name	Year	Author	Advantages	Disadvantages	Refer Ideas
1	Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator	2019	Terry Traylor, Jeremy Straub, Gurmeet, Nicholas Snell	1. Used natural language processing 2. Fake news detection based on attribute classification	Time consuming process	In this paper, the results of a fake news identification study that documents the performance of a fake news classifier are presented. The Textblob, Natural Language, and SciPy Toolkits were used to develop a novel fake news detector
2	Fake News Detection Using A	2018	Rohit Kumar Kaliyar	1. Used Natural Language Processing,	Limited dataset	This paper introduce the datasets which contain both fake

	Deep Neural Network			Machine learning and deep learning techniques to classify the datasets 2.Accuracy is better		and real news and conduct various experiments to organize fake news detector
3	On Scalable and Robust Truth Discovery in Big Data Social Media Sensing Applications	2018	Daniel (Yue) Zhang, Dong Wang, Nathan Vance, Yang Zhang, and Steven Mike	1. Find trustworthy information on Social media 2.Proposed truth discovery scheme using Work Queue in an HTCondor system	Accuracy is low	a distributed framework to implement the proposed truth discovery scheme using Work Queue in an HTCondor system
4	Fake News on Social Media: Brief Review on Detection Techniques	2018	ZaitulIradahMahid, SelvakumarManickam, Shankar Karuppayah	Proposed content-based, social context-based and hybrid-based methods	Only survey state of the are methods	Studied various detection techniques i.e. content based, social context based and hybrid based
5	FAMOUS: Fake News Detection Model based on Unified Key Sentence Information	2018	Namwon Kim, DeokjinSeo, Chang-Sung Jeong	Implement natural language processing using key sentence retrieval	Fake news detection accuracy is low	present a new fake news detection model using unified key sentence information which can efficiently perform sentence matching between question and article by using key sentence retrieval based on bilateral multi perspective matching model
6	Fake News Identification on Twitter with Hybrid CNN and RNN Models	2018	OluwaseunAjao, DeepayanBhowmik, ShahrzadZargari	Implement hybrid CNN and RNN Models Accuracy is much better	Only consider tweet headlines	classifies fake news messages from Twitter posts using hybrid of convolutional neural networks and long-short term recurrent neural network models
7	SOCIAL MEDIA AND FAKE NEWS IN THE 2016 ELECTION	2017	Hunt Allcott Matthew Gentzkow	1.detect 2016 election fake news spread through social media	1. Limited dataset used 2. Limited to 2016 news only	Comparing news to other sources

				2. Goal in this paper is to offer theoretical and empirical background to frame this debate		
8	Fake News Detection Using Naive Bayes Classifier	2017	MykhailoGranik, VolodymyrMesyura	Implement naïve bayes machine learning algorithm	Accuracy is low	This paper shows a new approach for fake news detection using naive Bayes classifier
9	Fake News Detection on Social Media: A Data Mining Perspective	2016	Kai Shu , Amy Sliva , Suhang Wang , Jiliang Tang , and Huan Liu	In this paper, we explored the fake news problem by reviewing existing literature in two phases i.e. characterization and detection	Use static data	In the characterization phase, we introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction.
10	When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism	2014	MeitalBalmas	Used multiple news sources for fake news detection.	Focus on only political data	This study contributes to the scientific knowledge regarding the influence of the interaction between various types of media use on political effects.

III. PROPOSED WORK

As social media usage is rapidly increasing, credibility of news gather over the social media is main aspect to be solved. News collected on social media can have difference source path such as twitter, Whatsapp, Facebook, social web sites and many more which doubted the accuracy of news received at one end. To maintain the credibility of information over social media there are special annotators who are expert in Natural Language Processing and Text Extraction so that the news can be extracted and the resulted news can be compare with real and genuine news. Annotator analysis the context and check for authorized sources. If any context seems to be fake annotators follows various steps to segregate fake news and genuine news:

There are four steps for fake news detection:

i. Data Collection
ii. Data Preprocessing
iii. Feature Extraction
iv. Classification

In this process, whenever news is send from source all data are collected from the message and verify for further preprocessing technique where the data collected from the source is analyze and the messages is splits into separate words using stopword removal, stemming and tokenization, after preprocessing the resulted text follows feature extraction and required algorithm is applied to text for classification.

News that is spread over the whatsapp is set in a special block of information in which the message is set in 3 categories:

i. Source
ii. Headline
iii. Body Text
iv.

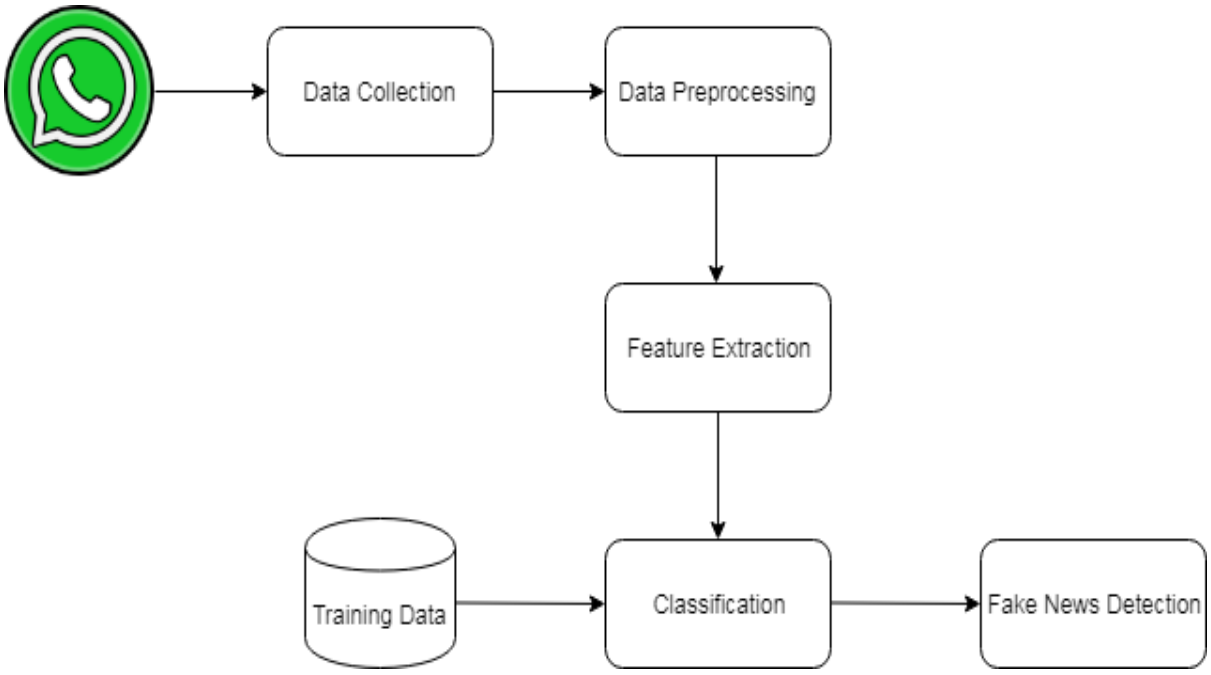

Fig.1:

IV. CONCLUSION

Growing popularity of social media, more and more people consume social media news instead of traditional media. However, social media have also been used to disseminate false news, which has strong negative impacts on individual users and the wider society. Here to explore the problem of false news by reviewing existing literature in two phases: characterization and detection. In Stage I, I have done characterization phase, in this phase analyze the basic concepts and principles of false news in both traditional media and social media. Also survey different relevant research papers and analyze the research gap and future direction for my project.

V. REFERENCES

[1]. Terry Traylor, Jeremy Straub, Gurmeet, Nicholas Snell" Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator"2019.

[2]. Rohit Kumar Kaliyar" Fake News Detection Using A Deep Neural Network"2018.

[3]. Daniel (Yue) Zhang, Dong Wang, Nathan Vance, Yang Zhang, and Steven Mike" On Scalable and Robust Truth Discovery in Big Data Social Media Sensing Applications" 2018.

[4]. ZaitulIradahMahid,SelvakumarManickam,ShankarKaruppayah" Fake News on Social Media: Brief Review on Detection Techniques" 2018.

[5]. Namwon Kim, DeokjinSeo, Chang-Sung Jeong" FAMOUS: Fake News Detection Model based on Unified Key Sentence Information" 2018.

[6]. OluwaseunAjao, DeepayanBhowmik, ShahrzadZargari" Fake News Identification on Twitter with Hybrid CNN and RNN Models"2018.

[7]. Hunt Allcott Matthew Gentzkow" SOCIAL MEDIA AND FAKE NEWS IN THE 2016 ELECTION" 2017

[8]. MykhailoGranik, VolodymyrMesyura" Fake News Detection Using Naive Bayes Classifier"2017

[9]. Kai Shu , Amy Sliva , Suhang Wang , Jiliang Tang , and Huan Liu" Fake News Detection on Social Media: A Data Mining Perspective"2016

[10]. MeitalBalmas" When Fake News Becomes Real: Combined Exposure to Multiple News Sources and Political Attitudes of Inefficacy, Alienation, and Cynicism" 2014

[11]. Daniel (Yue) Zhang, Chao Zheng, Dong Wang, Doug Thain, Chao Huang, Xin Mu, Greg Madey ."Towards Scalable and Dynamic Social Sensing Using A DistributedComputing Framework."Department of Computer Science and EngineeringDepartment of Aerospace and Mechanical Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2017.

[12]. Daniel (Yue) Zhang, Dong Wang, Hao Zheng, Xin Mu, Qi Li , Yang Zhang." Large-scale Point-of-Interest Category Prediction Using Natural Language Processing Models." Department of Computer Science and Engineering Department of Aerospace

and Mechanical Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2017.

[13]. Marco Viviani and Gabriella Pasi. "On Credibility in social media: opinions, news, and health information—a survey" Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(5), e1209, 2017.

[14]. Daniel (Yue) Zhang, Rungang Han, Dong Wang, Chao Huang." On Robust Truth Discovery in Sparse Social Media Sensing." Department of Computer Science and Engineering University of Notre Dame Notre Dame, IN, USA IEEE 2016.