INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

Fake News Detection using Machine Learning Framework

Dipali Ghodke[1] , Rutuja Borate[2] , Amruta Kadam[3] , Pratima Rupanwar[4] , Prof. G.M.Bhandari[5]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[1]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[2]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[3]

U.G. Student, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[4]

HOD, Department of Computer Engineering, BSIOTR, Pune University, Pune, India[5]

ABSTRACT: social media as a source for news it's a matter of judgement. Social media news is so accessible and easy to get that people search for it and consume it regardless of its price. on the other hand, fake news (i.e. low-quality news with intentionally false information) can be widely disseminated There is the potential for extremely negative impacts to individuals and society from the widespread dissemination of false news. Because of this, recent research into finding false news in social media has become an emerging field of study that is receiving a great deal of attention. An irresponsible and groundless report of Traditional media has old detection algorithms that are ineffective when applied to social media. First, the intent of information on the internet, called 'fake news,' is to deceive readers into believing information that is actually incorrect, which makes it more complex to uncover based on news content; thus, we must include information that can aid in the verification. The data from the social commitments of users with false news is large, incomplete, unstructured, and noisy. This survey was conducted in order to continue facilitate research on the problem of trying to find false news in social media. In this survey, we cover everything from characterizations of fake news on psychology and social theories to current algorithms from a data mining perspective. Furthermore, we cover other related research topics, open problems, and future directions for social media fake news detection.

KEYWORDS-Fake News, User Profile, Trust Analysis, machine learning, Social Media.

I. INTRODUCTION

The spread of reliable information on the Internet (the WWW) is an important issue in today's society. Many publications have recently taken note of the growing problem of misinformation and fake news on the internet. It is having a significant impact on political and social realities. To illustrate, he showed how fake news had an enormous impact on the 2016 U.S. presidential election. He further investigated the most viral tweets from the Boston Marathon blasts in 2013 and found that there was an increased amount of misinformation compared to the amount of truth. More and more people now look to social media rather than traditional news organisations for information. These social media platforms are built with this trend in consumption behaviour in mind: I news stories on social media are easier to consume, such as through Facebook or Twitter, as opposed to traditional media such as newspapers or television; and

,

(ii) the capacity to share, comment on, and discuss the news is easier on social media, too. For example, 62% of adults in the United States get their news on social media in 2016, compared to 49% in 2012. In addition, it was found that social media has surpassed television as the leading source of news. Even though social media provides significant advantages, news quality on social media is lower than news organisations that use traditional media.

While it is easy and inexpensive to disseminate news online, large volumes of fake news are produced online for various reasons, such as financial and political gain. Pizzagate has been estimated to have been tweeted about 1 million times "This could be finished by the end of the presidential election. We know this phenomenon exists because of the abundance of this newer form of media "This is especially impressive because it was the first time the Macquarie Dictionary had selected a word of the year. Fake news can be dangerous to individuals and society because of its widespread distribution. Fake news is dangerous for the media ecosystem, because it threatens the authenticity of news items. An example of this is shown by the wide dissemination of false news on Facebook during the 2016 U.S. presidential election, as opposed to the wide dissemination of genuine mainstream news. Another consideration is that fake news is made to purposefully convince consumers to accept biassed or false beliefs. Manufactured "fake news" is usually used by propagandists to get their political messages across, or to influence others. In other words, it has been reported that Russia has developed false accounts and social bots to disseminate false stories. The third point is that fake news can influence the way people process and respond to the news. Let's take a look at some fake news as an example. It was created to purposely disturb people and confuse them, thereby limiting their ability to tell what is true from what is false. We have to find a way to automatically detect fake news on social media.

There are many different, new, and difficult problems that arise when it comes to detecting fake news on social media. The spread of fake news on social media has made fake news more powerful, and traditional journalistic norms are under attack. This problem contains multiple characteristics that are hard for automated detection to handle. The first is that fake news is written with the intent to mislead readers, making it nearly impossible to identify based on the quality of the news content. It's difficult to identify an exact type of fake news because fake news content varies by topic, style, and media platform. Fake news seeks to distort truth using various language styles while also poking fun at true news. A fake news storey might cite verifiable evidence to make an incorrect claim. Fake news detection requires data-specific textual features and hand-crafted ones, and the existing body of text-based features does not meet this requirement. Detectability must also be enhanced with auxiliary information, such as the use of knowledge bases and social engagements with users. Lastly, in order to effectively use this auxiliary information, the quality of the data itself is of utmost importance. There is a low probability of duplicated claims, because new, time-sensitive events do not yet have an established body of knowledge to corroborate them. However, along with sharing fake news with their social circles, people's social interactions produce large, incomplete, unstructural, and noisy data. There is still much work to be done in discovering effective means of differentiating credible users; additional investigations are needed.

II. RELATED WORK

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 classificationDisadvantages- 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 datasets2.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 is1. 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

III. PROPOSED APPROACH:-

If you want to get online news, you can go to various news websites, such as government press releases, search engines, and social media sites. Manually determining the truth of news stories, however, is a difficult challenge, typically requiring experts with special knowledge who thoroughly assess claims and evidence as well as context and multiple reports from authoritative sources. The following ways can be used to gather news data with annotations: Brawny brawns, Brainy brains, Scanning software, and Subscription services for gleaning data from the general public.

Meta information related to a piece of news is reported in news content. Below is a list of all news attributes we will include in our API.

The source of the news article is the author or publisher of the article.

Title: A brief description of the main topic of the article that's designed to catch the attention of readers

(Large Text: The main text of the news storey which elaborates on the details of the storey, and this highlights a specific claim, which influences the point of view of the media)
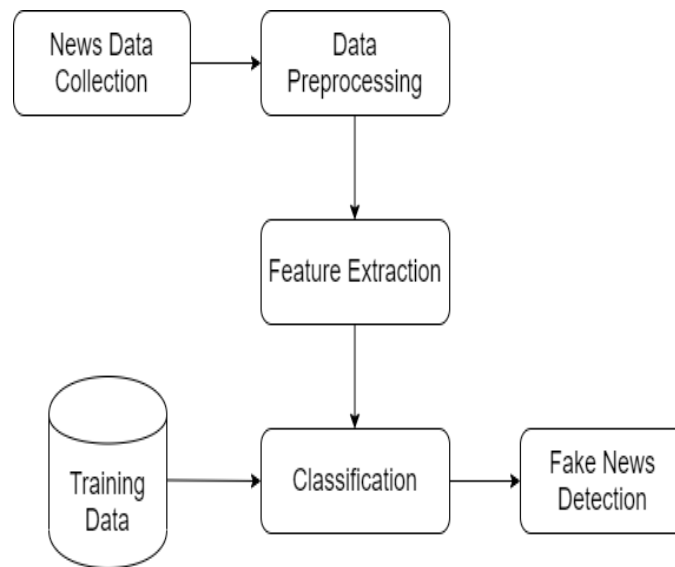
System Diagram:

Fig 1. System Architecture

Algorithm:

Naive Bayes

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A=number of objects of class A/total number of objects
   Calculate the prior probability of class B=number of objects of class B/total number of objects
3. Find NI, the total no of frequency of each class
   Na=the total no of frequency of class A
   Nb=the total no of frequency of class B
4. Find conditional probability of keyword occurrence given a class:
   P (value 1/Class A) =count/ni (A)
   P (value 1/Class B) =count/ni (B)
   P (value 2/Class A) =count/ni (A)
   P (value 2/Class B) =count/ni (B)
   ………………………………..
   ………………………………..
   ………………………………..
   P (value n/Class B) =count/ni (B)
5. Avoid zero frequency problems by applying uniform distribution
6. Classify Document C based on the probability p(C/W)

   a.   Find P (A/W) =P (A)*P (value 1/Class A)* P (value 2/Class A)……. P(value n /Class A)

    b.   Find P (B/W) =P (B)*P (value 1/Class B)* P (value 2/Class B)……. P(value n /Class B)

7.   Assign document to class that has higher probability.

IV. RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. Some functions used in the algorithm are provided by list of jars like Twitter-core and Twitter-stream jars etc.

Naïve Bayes Performance:
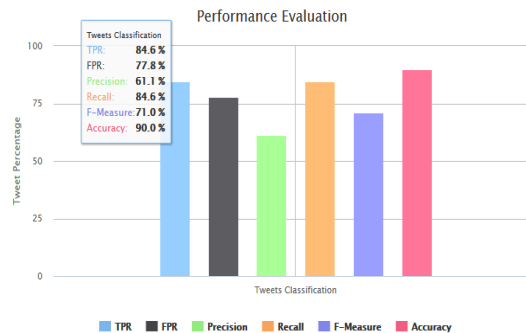
Parameters	Percentage
TPR	84.6
FPR	77.8
Precision	61.1
Recall	84.6
F-Measure	71.0
Accuracy	90.0

V. CONCLUSION

Social media is getting more and more popular, and more and more people are turning to social media for their news. However, false news can be spread via social media, and this can have a wide range of harmful effects on individuals and society as a whole. We will perform a literature review in two phases: characterization and detection to help elucidate the problem of false news. We define basic concepts and principles of false news in both traditional media and social media during the characterization phase. We investigated the current false news detection techniques, including feature extraction and model building, in the detection phase. Additionally, we explore evaluation metrics, as well as future promising fake detection research paths, and we increase the reach of our fake detection area to various applications.

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] Oluwaseun Ajao, Deepayan Bhowmik, Shahrzad Zargari" 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, Volodymyr Mesyura" 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 Distributed Computing Framework." Department of Computer Science and Engineering Department 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] 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

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi® crossref

ISSN
INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 🟢 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details