IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

Machine Learning Based Fake News Detection

Samarendra Yadav, Rashmi Khade, Sukanya Konale, Sanjeev Yadav
Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India
Savitribai Phule Pune University Maharashtra, Pune, Maharashtra, India

Abstract: *Consumption of news via social media, It is a two-edged sword. On one hand, consumers use social media to look for and consume news because of its low cost, ease of access, and rapid transmission of information. On the other side, it facilitates the widespread distribution of fake news, which is low-quality news that contains purposely misleading material. False news that is widely disseminated has the potential to have extremely negative consequences for both individuals and society. As a result, the identification of false news in social media has recently become a hot topic of research. False information about social media detection has unique characteristics and obstacles that render classic media detection algorithms ineffective or inapplicable. First, false news is purposefully designed to fool readers into believing false information, making it difficult to deconstruct and discover based on news content; as a result, we need to integrate auxiliary information, such as users' social commitments in social media, to aid in making a conclusion. Second, utilising this auxiliary data is difficult in and of itself, as users' social commitments to false news result in data that is huge, fragmentary, unstructured, and loud. We undertook this poll to continue to promote research on the subject of detecting false news in social media, which is both complex and relevant. We provide a complete analysis of detecting false news on social media in this survey, which includes fake news characterizations based on psychology and social theories, existing data mining techniques, evaluation criteria, and sample datasets. We also talk about adjacent study topics, open difficulties, and future research paths for social media fake news identification.*

Keywords: Fake News, User Profile, Trust Analysis, Machine Learning, Social Media

I. INTRODUCTION

The trustworthiness of information disseminated on the World Wide Web (WWW) is a major concern in modern culture. In recent years, the spread of misinformation and fake news on the Internet has gotten a lot of attention, and it's gotten to the point where it's altering political and social realities substantially. For example, studied the most viral tweets related to the 2013 Boston Marathon bombings, finding that the percentage of rumours and bogus content was larger than the amount of accurate information. People are increasingly seeking out and consuming news through social media platforms rather than traditional news organizations, since we spend an increasing amount of our life communicating online through social media platforms. The reasons for this shift in consumption habits are inherent in the nature of these social media platforms: (I) It is often more timely and less expensive to consume news on social media than it is to consume news through traditional news media, such as newspapers or television; and (II) it is easier to share, comment on, and discuss the news with friends or other readers. For example, in 2016, 62 percent of adults in the United States received news via social media, compared to only 49 percent in 2012. It was also discovered that, as a key news source, social media currently beats television. Despite the benefits of social media, the quality of news on these platforms is inferior to that of established news organizations.

Large volumes of fake news, i.e. news pieces with purposely misleading material, are created online for a variety of reasons, including financial and political advantage, because it is inexpensive to provide news online and much faster and easier to disseminate through social media. Over 1 million tweets have been attributed to the false news scandal known as “Pizzagate” By the end of the presidential election, three people would have been elected. Given the ubiquity of this new issue, “fake news” is a good term to use. In 2016, the Macquarie dictionary named it the word of the year. Fake news widespread dissemination has the potential to harm both individuals and society. For starters, false news has the potential to disrupt the news ecosystem's authenticity balance. During the 2016 presidential election in the United States, for example, it is clear that the most popular false news was shared even more extensively on Facebook than the most popular legitimate mainstream news. Second, fake news is designed to induce customers to believe in biased or incorrect information. Propagandists frequently use fake news to spread political messages or exert influence. According to certain reports, Russia

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

has developed phoney accounts and social bots in order to distribute incorrect information. Third, bogus news alters how people perceive and react to actual news. For example, some fake news are manufactured solely to instil distrust and perplexity in individuals, obstructing their ability to discern what is true from what is not. It is vital that we create tools to automatically detect fake news on social media to limit the harmful consequences of fake news, both for the benefit of the public and the news ecosystem.

Detecting fake news on social media presents a number of novel and difficult research challenges. Though fake news is not a new problem—nations and groups have been utilizing the news media to carry out propaganda and influence operations for centuries—the rise of web-generated news on social media has amplified the potency of false news, challenging traditional journalistic conventions. There are various elements of this problem that make automated detection particularly difficult. (I) False news is purposefully created to deceive readers, making it difficult to detect based on news substance alone. Fake news has a wide range of topics, styles, and platforms, and it uses a variety of language approaches to distort the truth while mocking actual news. Fake news, for example, may use accurate evidence in the wrong context to support a false claim. As a result, existing hand-crafted and data-specific textual features are often insufficient for detecting fake news. Other auxiliary information, such as a knowledge base and user social engagements, must also be used to improve detection. (II) Utilizing this supplementary data raises another significant issue: the data's quality. Fake news is usually associated with recent, time-sensitive events that, due to a lack of corroborating evidence or claims, may not have been properly verified by existing knowledge bases. Furthermore, users' social interactions with bogus news generate large, incomplete, unstructured, and noisy data. Effective approaches to distinguish credible users, extract relevant post attributes, and harness network interactions are still a work in progress that requires more research.

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 Text blob, Natural Language, and SciPy Toolkits were used to develop a novel fake news detector. Advantages are: 1. Used natural language processing 2. Fake news detection based on attribute classification Disadvantages: Time consuming process.

This paper [2] introduces the datasets which contain both fake and real news and conduct various experiments to organize fake news detector. Advantages are: 1. Used Natural Language Processing, Machine learning and Deep Learning techniques to classify the datasets 2. Accuracy is better and Disadvantage 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 are: 1. Find trustworthy information on Social media 2. Proposed truth discovery scheme using Work Queue in an HTCondor system and Disadvantage is: Accuracy is low

This Paper [4] Studied various detection techniques i.e. content based, social context based and hybrid based. Advantages are: Proposed content-based, social context-based and hybrid-based methods and the only Disadvantage is: 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 are: Implementation of 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 are: Implementation of hybrid CNN and RNN Models and Accuracy is much better. Disadvantage is: it only considers tweet headlines.

This paper [7] Compare news to other sources in 2016 year. Advantages are: 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 are: 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. Used Implementation of 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

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

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 user 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 are 1. Used multiple news sources for fake news detection and Disadvantage are: Focus only on political data

III. PROPOSED APPROACH

Online news can be collected from different sources, such as news agency homepages, search engines, and social media websites. However, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs careful analysis of claims and additional evidence, context, and reports from authoritative sources. Generally, news data with annotations can be gathered in the following ways: Expert journalists, Fact-checking websites, Industry detectors, and Crowd-sourced workers.

News content features describe the Meta information related to a piece of news. A list of representative news content attributes is listed below:

- Source: Author or publisher of the news article.
- Headline: Short title text that aims to catch the attention of readers and describes the main topic of the article
- Body Text: Main text that elaborates the details of the news story; there is usually a major claim that is specifically highlighted and that shapes the angle of the publisher

3.1 System Diagram

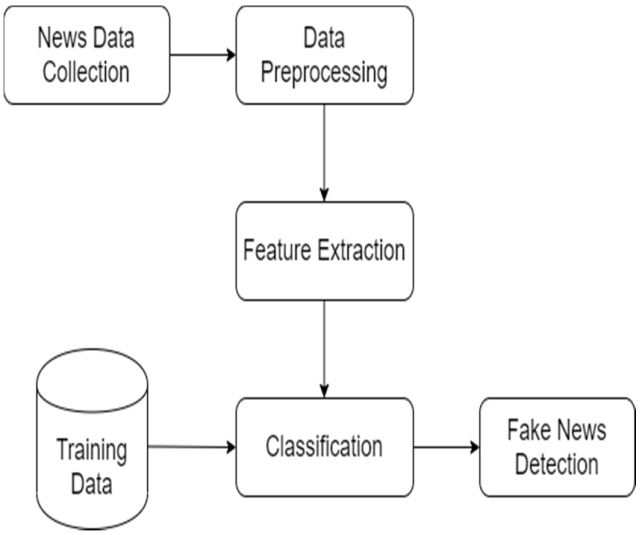


Fig 1. System Architecture

3.2 Algorithm

A. 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)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

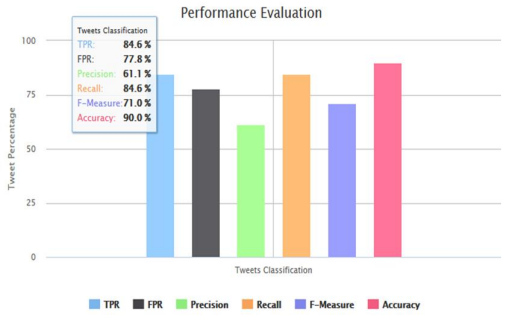
P (value 1/Class B) =count/ni (B)
P (value 2/Class A) =count/ni (A)
P (value 2/Class B) =count/ni (B)
……..
……..
……..
P (value n/Class B) =count/ni (B)

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.

4.1 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

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 the characterization phase, we introduce the basic concepts and principles of false news in both traditional media and social media. In the detection phase, we reviewed the current false news detection approaches from a machine learning perspective, including feature extraction and model building. We also discuss evaluation metrics, and future promising directions in fake detection research and expand the field to other applications

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

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]. Zaitul Iradah Mahid,Selvakumar Manickam,Shankar Karuppayah "Fake News on Social Media: Brief Review on Detection Techniques" 2018.

[5]. Namwon Kim, Deokjin Seo, 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]. Mykhailo Granik, 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]. Meital Balmas "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