A Comparative Analysis on Fake News Detection Methods

Md. Badhon Miah
American International University-Bangladesh
Dhaka, Bangladesh
badhon1512@gmail.com

Tawhid Khondakar
American International University-Bangladesh
Dhaka, Bangladesh
tawhid.khondakar18@gmail.com

Atanu Saha
American International University-Bangladesh
Dhaka, Bangladesh
atanu.saha415@gmail.com

Arrafi Ahsan Somudro
American International University-Bangladesh
Dhaka, Bangladesh
arrafeeahsan@gmail.com

Dr. Dip Nandi
American International University-Bangladesh
Dhaka, Bangladesh
dip.nandi@aiub.edu

ABSTRACT

In present era, believing a news is as easier that it is also hard to believe. It is highly possible that people who came across a news don't justify it before believing. Majority of people believe a news on the basis of its spreading. But believing a news which is fake impacts the world including politics, economics, education, health, business, communications and many more domains of life. To bounce back from these impacts, it is necessary to know in depth of fake news and its properties. Thus, in this paper we discuss the real definitions of fake news, how a fake news is made, and its impact on life and the science behind fake news. As currently people are more interested to research on fake news, we discuss and analyze existing detecting methodologies, dividing each into certain criteria. Finally, we conclude by giving a table which contains comparative analysis of detecting methods and suggested some future work.

KEYWORDS

Fake news, Fake News Detection Method, Social Media, Satire, Hoaxes

ACM Reference Format:
Md. Badhon Miah, Tawhid Khondakar, Atanu Saha, Arrafi Ahsan Somudro, and Dr. Dip Nandi. 2018. A Comparative Analysis on Fake News Detection Methods. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The term 'Fake News' is a new phenomenon that has got a huge attention worldwide in recent decades. Fake means 'a thing that is not genuine' and News means 'newly received or noteworthy information, especially about recent events. 'Fake News' is a term which often discussed in both media and research and in the era of internet, fake news is one of the considerable problems in social media in recent years, which can cause serious problems in society. Macquarie dictionary had chosen 'Fake News' as 'The Word of the Year' in the year 2016 [15]. When authoritative structures and civilians take standard decisions, the overall nationality becomes secure and well built [10]. But it creates problem when those citizens start to believe a news which is merely true. Anyone can easily get access to news information sources because of the internet and social media [3]. In this present era, a news which is not true, can easily make a person's mind to believe it as true in that much which provokes the person to share it with other people.

This article describes a depth portion of the properties of fake news along with existing detection methods which are mostly in the region of social media and online news article. The goal of the research paper is to help people both from non-research background and research background to understand and work with fake news. The research paper gives an overview of fake news including its history, impacts, sources of fake news and how fake news are made which is enough for one to understand the wholesome of fake news. As recently, research on fake news is getting interested by research community, we also worked to help them by discussing and analyzing existing works and come up with a comparative table so that is becomes easy to pick one for any research purpose.

2 THE SCIENCE OF FAKE NEWS

People nowadays mostly relies on social media as a news source nowadays because of mainly two reasons – 1. Getting news from social media is less expensive and it requires less time comparing to traditional news sources. 2. It is easy to share, comment and interact with news sources in social media [17]. The rate has increased from 49 percent to 62 percent for the adult users in U.S. from 2012 to 2016 [18] [21]. In Facebook, like-minded people often create groups where they promote, consume and share fake contents among themselves. They also assume this news to be true due to the lack of proper evidence. This is called the Echo Chamber Effect [18]. They belief such things because of two psychological reasons-

1. Social credibility: People tends to perceive a source as credible when they find other people perceiving the source as credible even though there is no evidence to support the truthfulness of these sources.

2. Frequency heuristic: If a large number of people supports an idea or holds a belief, then it influences others to support that idea.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

Also, due to the echo chamber effect, people are more likely to create communities with specific ideologies and spreading them among themselves.

3 IMPACTS

Research has shown that people prefer to belief information that satisfies with their pre-existing thought process which means people are emotionally biased towards accepting information weather it is a superstition or mutual belief in a conservative community [6][1]. Social media platforms use complex statistical model to predict and show relevant information that matches the interest of the particular user [6]. But these models often ended up providing them the false information that they were seeking for. Human intelligence is incapable to distinguish between real and fake news. There are two psychological factors that makes them vulnerable towards fake news [15] –

1. Naïve Realism: This is a term of social psychology that refers the human tendency to believe that their own perception of the world around them are the only accurate views and those whoever disagrees with them considered as biased, irrelevant and uninformed.

2. Confirmation bias: This is the tendency to accept and search for information that supports one's prior beliefs and values. This also impact how people gather and recall information. For example, a person will seek information that supports or oppose a particular issue. They will also interpret that information in a certain way that upholds the ideas which they support.

These psychological biased factors increase misperceptions among ideological groups. Also, the desire for maximizing reward in social network by getting acceptance and response from the other users is responsible for spreading fake news [15].

4 DEFINITION

So, what really makes 'Fake News' misleading? To know the answer, one should know about the definition of Fake News. There is not yet an agreed definition for the term but there some key properties of fake news which is widely adopted in many literatures and can be combined to use as a definition [15]. The word fake is related to the words copy, forgery, counterfeit, and inauthentic. Many authors have given the definition of fake news in their own way. Some of them are given below:

1. 'Fake news is an allegation that some story is misleading – it contains significant omissions – or even false – it is a lie – designed to deceive its intended audience' [2].

2. Another definition by [9] refers to fake news as ' any false information that is deliberately meant to be wholly or largely false or misleading, spread through online social media, but occasionally finding its way to mainstream traditional print and broadcast news media'.

3. [11] assumes ' Fake news articles are intentionally fabricated to be deceptive and can be proven that they are false'.

4. [14] defines fake news as the news which are deliberately and verifiably false and could misguide the consumers while the broad definitions focus on the authenticity and intent of the news content.

5. In another definition, 'fake news can be represented as the online publication of intentionally or knowingly false statements of fact' [5].

6. The denouncing of media and journalism, using derogatory terms can be regarded as part of a well-known strategy predating the popularization of the term fake news [12].

Satire news with proper context, there are also some rumors that does not belong to any news portal or conspiracy theory is difficult to verify whether it is real or fake. Unintentionally created misinformation, and hoaxes that are used for fun are not considered as fake news [15]. In a nutshell, Fake news is the propaganda of spreading wrong information which are done intentionally to achieve financial profits and goals.

5 SOURCES

Fake News can spread in various ways which indicates that it can contain many sources for-example oral message, print media, news sites, social and news media. Fake news is false news intended to deceive people and does so by trying to look like real news [17]. Nowadays online platforms create a huge opportunity for non-journalists to make or spread fake news [18]. A study shows that 44% of the population in the USA gets their news from social media [18]. Fake news is mainly manipulated and created by some propagandists to convey political message or to achieve business profits through advertising [15]. There are some malicious accounts on social media that is fully autonomous and can interact with systems and users. Facebook, Google, Internet news sites and Twitter are the major sources of fake news, while social media acts as a minimal cost channel of distribution, where fake data if published then it spreads throughout the web as wildfire to mass through retweets, likes and sharing [10]. Among these mentioned digital media, majority people of any country consume news through social media because a news which has more likes, comments and shares receives huge attention and people tends to like, comment and share that again. A few famous websites of fake news include "The Onion", a sarcastic publication with a readership of 4.3 trillion, Facebook, and Monticello [9]. There are some more web pages that publish fake news purposely (For example, Denvergurdian.com, wtoe5news.com, and ABCnews.com.co). These webpages publish fake news like legit news organizations and intentionally produced to distribute hoaxes, advertising, and false news to gain financial and political benefits without the reader's awareness [20].

6 GENERATORS AND BOTS

To increase the spreading ratio of fake news, the creator often uses Online Fake News Generation Tool. "Break your own News", "ClassTools Breaking News Generator", "The Fake News Generator" are the most popular tool for fake news generation which only needs to upload picture, textual content, author and an eye-catching title to create a Fake News. After this, the creator tends to spread the Fake News by posting it on Facebook, Twitter, Tumblr etc. [10]. There are some malicious accounts on social media that is fully autonomous and can interact with systems and users. These accounts are intentionally created and programmed to spread a particular news or advertisement to a group of people. These accounts are called bot. Based on sharing behavior, number of ties and linguistic

features it is estimated that around 9-15% of active twitter accounts are bots and about 60 million of Facebook accounts that are bots [6]. Fake news is widely spread by bots [16]. The scope and variety of online fake news are understood by a few important aspects for defining fake news in Fig. 1, in the onion-shaped graph the term fake new is the core of it, and it holds four main components: Creator/ Spreader, Target Victims, News Content, and Social Context. Every component is in the first inner layer around "Fake News" [20].

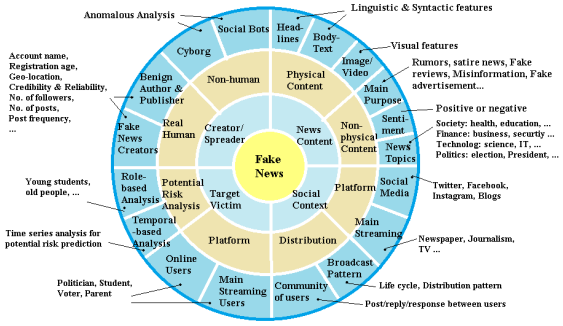

Figure 1: All the factors related with fake news

7 METHODS

Fake news detection is a comparison between some authentic newspapers with social news content [21]. Till now, authors have discovered some reliable methods which are able to detect a news as Fake News. It depends on style-based, propagation-based, credibility-based, and so on. Each identical maintains its own quality accessing datasets, different detection tricks in data mining, machine learning, many natural sources, and local search. All of these sources are finally generating a unified framework for the exploration of fake news. Some authentic detection methods are describing below:

7.1 Propagation Path Classification

7.1.1 Domain. [Weibo, Twitter] [7]

7.1.2 Theory. Author Yang Liu and Yi-Fang Brook Wu has denoted their model as "PPC" (Propagation Path Classification), also as "PPC_RNN+CNN" which also includes a reduced version of the main method. Propagation path is a model which is form with each news story as a two or more embedded series of time with adding tuple on which each tuple is a numbered container shows habitual behavior of a user who has a connection directly or indirectly with spreading fake news. Along with this, recurrent and convolutional networks related time series were added to capture user characteristics and detect fake news [7].

7.1.3 Main Method. The fake news detection method proposed by Yang Liu et al is a combination of four components including, 1) Propagation Path Construction and Transformation: From a news story a propagation path was constructed by identifying users whose are involve in propagating the news. Propagation path AKA variable-length multivariate time series was constructed from relevant user profiles. For transforming the length of a variable from multi organized series of time into a constant series of time sequence an algorithm was used to get a vector representation which basically encodes a global variation of user-characteristics.

2) RNN-Based Propagation Path Representation: Gated Recurrent Unit (GRU) which is a variation of RNN was utilized for each of the transformed propagation path and then by applying mean pooling it was possible to reduce the sequence of the output vectors which is the final vector representation of the general deviation of the characteristics of a user.

3) CNN-Based Propagation Path Representation: To adapt and equip another delegation of vector, CNN was used for each of the transformed propagation path to generate final vector representation that encodes the local variation of user-characteristics by applying mean pooling.

4) Propagation Path Classification After getting global variation of user characteristics from RNN and local variation of user-characteristics from CNN, a single vector was concatenated which represents the transformed propagation path. The final output brings probability distribution for the corresponding propagation path.

A combination of three real world dataset Weibo, Twitter15, Twitter16 were used to evaluate the proposed model where it was estimated that for the validation of model it is necessary to hole 10% of the dataset, rest of the dataset were split with the ratio of 3:1 for training and testing purpose. The proposed model was implemented using Keras (https://keras.io/) [7].

7.1.4 Accuracy. By evaluating the model through three real world dataset, it was confirmed that the proposed model which is based on Propagation Path Classification can find out a news as a fake news with an extreme accuracy of 85% in Twitter and 92% in Weibo in 5 minutes after it starts to spread.

7.2 Naïve Bayes

7.2.1 Domain. Facebook [3]

7.2.2 Theory. Mykhilo Granik and others provided a simple fake news detection approach. They used Naïve Bayes classifier to complete the whole procedure execution successfully. Mainly, the approach they proposed was implemented as a software system. They collect a labeled dataset which contains Facebook posts and tested against it which gave them a decent accuracy [3].

7.2.3 Main Method. Authors tried to figure out similarities between Spam message and the articles which contains fake news and discovered that there are many significant landmark which intersects between them. Inspired by these properties, authors tried to use similar approach for fake news detection. Identify and get information from each word of a news article was the key idea.

As it is often recognized as True Positive that the news articles which uses same words more often are categorized as fake. Though claiming a news fully fake is impossible but we can see that the number of similar type of words encounters maximum time which changes the statistical probability. Formula for calculating conditional probability of this fact using Bayes theorem is:

(1) $Pr(F|W) = Pr(W|F) \cdot Pr(F) / (Pr(W|F) \cdot Pr(F) + Pr(W|T) \cdot Pr(T))$

Pr(F|W) is known as the probability which represents every single words belonging to a news article. After that it is necessary to confirm that given news article is fake by using these probabilities using Spam filtering formulas, given below: (2) p1 = Pr(F|W1)·...·Pr(F|Wn), (3) p2 = (1 − Pr(F|W1))·...·(1 − Pr(F|Wn)), (4) p = p1 / (p1 + p2),

For calculating the conditional probabilities of finding specific word it was necessary to use a dataset of fake news contents labelled as either True or False. For this BuzzFeed news dataset was used. After filtering dataset, 1771 news articles was obtained and it was divided into 3 categories named as Training dataset which is used to train the naïve bayes classifier, Validation dataset which is used to tune the generic attributes of the classifier and Test dataset which is used to get the detached measurement of well performing. After that necessary calculation were done with the equations provided above, to get the classification accuracy for news article by using,

$$Precession = tp/(tp + fp) \quad (1)$$

$$Recall = tp/(tp + fn) \quad (2)$$

Table 1: Accuracy of Naive Bayes

New article types	Total number of news in the test dataset	Number of correctly classified news	Classification accuracy
True	881	666	75.59%
Fake	46	33	71.73%
Total	927	699	75.40%

7.3 Content Based

The content-based method is one of the automatic detection techniques which use content cues for detecting fake news [11]. There are two types of content available which are textual content and non-textual content. Textual contents, generally analyze the text for detecting fake news. In the modern news article, it is not only the collection of various texts but also a combination of text, images, audio, and videos. Other type non-textual content is an analysis of Images, URLs, sound, video, metadata of existing elements, web track information, image captions. The content-based method analyzes the contents of news like text, image, video, or sound and uses various machine learning techniques to analyze the content. The content-based method incorporates Linguistic features, Writing style features, Semantic features, Sentiment features, VisualBased features [4]. The basic language element analysis and sentence structure analysis are the significance of Linguistic features. For detecting fake news semantic features is one of the most important features. Sentiment analysis can detect the emotions and judgment from any news article and it is very helpful for detecting false information. There are several methods available for utilizing the sentiment of news articles such as arousal valence dominance score, happiness detection and analysis, emotion analysis, and analysis of polarization and strength. Sentiment analysis can also detect fake bot accounts on social media. The visual-based feature analyzes the news visual contents: the number of images and videos, multi-image frequency, popular image frequency, and so on [4].

7.4 User Based

Social media users are doing lots of activities on social media, these activities can give us strong clues about false information [11]. Authors [4], divide the user-based method into 3 sub-categories such as user profiling features analysis, posting behavior and temporal features analysis, and credibility features analysis. User profile features contain fundamental user information such as the number of followers, friends, number of statuses, location, profile creation time, profile creation date, verification information of users, etc.. By analyzing the information of the user, it can be possible to detect the activities of the user and this analysis is extremely helpful to detect fake or bot accounts on social media. It can be effective for detecting fake information because many fake users are spreading fake news on social media. Fake accounts such as bots and cyborgs post fake news within a specific certain time that creates a pattern. Posting behaviors show the social media users posting pattern at a certain amount of time, time pattern between two successive posts, sharing, mentioning, etc. User interaction on specific news is very important to detect whether it is fake news or not, and this analysis is called credibility features analysis. If the news has got more interaction such as likes, shares and comments from unreliable users then there is a high chance that it is fake news. Research [11] stated that any kind of user interaction analysis takes place in this category including user comments, Facebook likes etc. Social context relies on the demography of the user. Different attributes of a user like their age, gender, education, and affiliations are observed to standardize the user into a particular category [15]. Features of the user profile can be categorized as explicit and implicit features. Previously mentioned attributes are the implicit features and user's row metadata are the explicit features such as their information credibility classification [15].

7.5 Geometric deep learning

7.5.1 Domain. Twitter

7.5.2 Theory. [8] The approached model is based on geometric deep learning. There are multiple deep learning methods that has a great influence on other domains. Convolutional neural networks and other popular deep neural models are completely based on classical signal processing theory. These models are organized as a grid-structured data. Recently, the technique is becoming so popular to convert deep learning models as grid-structured non-Euclidean data. Since then, geometric deep learning is popular as a term similar to non-Euclidean deep learning approaches.

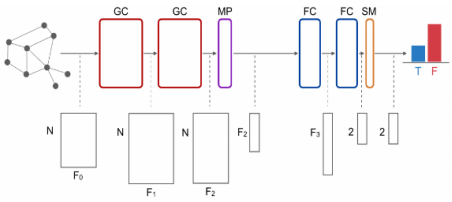

Figure 2: The architecture of neural network model [8]

7.5.3 Main Methods. For input, a URL u will be given or a cascade c which is a link of u. URL u will be described as a graph Gu which will have a set of tweets Tu = t1u, t2u, …., tNu. If i and j are two nodes and edge (i, j) Gu iff one of the following conditions occurs: 1. The person who wrote tweet i is following the person who wrote tweet j. 2. The person who wrote tweet j is following the person who wrote tweet i. 3. News spread occurred from i to j 4. News spread occurred from j to i. Here a graph CNN will be used with two convolutional layers and two fully connected layers to predict probabilities. The figure 3 of the model is given below:

Figure 3: Neural network model designing

Here, GC = Graph Convolution MP = Mean Pooling FC = Fully Connected SM = SoftMax layer. The method has examined in two different conditions: URL wise detection and cascade wise detection. The true positive rate and false positive rate for both conditions are given in figure 4

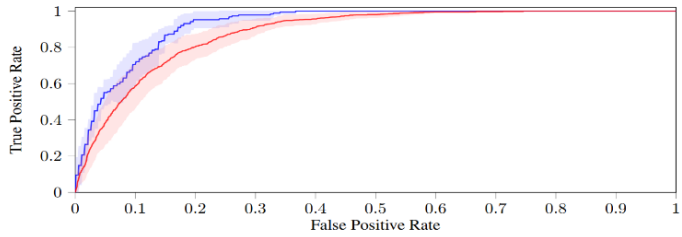

Figure 3: True positive and false positive rate [8]

7.5.4 Accuracy. The deep learning model was trained by supervised learning on a very large scale of fake and real news stories published on twitter within the period of 2013-2018. Also, the model has been observed in different challenging settings. The model performs extremely well and shows accuracy of 93% ROC AUC within just a few hours of the propagation.

7.6 Multiple Feature based solution

7.6.1 Domain. Facebook. The solution provided here will analyze some features associated with Facebook accounts.

7.6.2 Theory. [13] Fake news detection approach provided in this paper requires multiple classifiers. The description of each classifier given below: A. Machine learning related classifiers: a. KNN: This technique previously stores some predefined conditions and classifies and it produces a new condition by measuring similarities. It classifies an object based on the popularity index of other members in its neighborhood. The distance is calculated using Euclidean, Manhattan and Minkowski functions. b. Logistic regression: This is the most appropriate method if dependent variable is in binary form. Here, an output Y is predicted using some input parameters where each parameter has different weights with some coefficient values. The standard equation for this technique is given below:

$$Y = e(b0 + b1X)/(1 + e(b0 + b1X)) \quad (3)$$

c. Decision tree: In this technique, all the classification model forms a tree structure and all the data is divided into sub segments. Every decision is indicated by an edge and the resultant output for that decision is presented by a node. Principle of entropy and Information gain is used to build a decision tree. d. Naïve Bayes: This model is based on probabilistic principle. The following equation is used for classification which is based on Bayes theorem-

$$P(A|B) = (P(B|A)P(A))/P(B) \quad (4)$$

B. Deep learning related classifiers: In this proposed solution, LSTM has been used which is a short form of 'Long Short Term Memory'. LSTM: It is a recursive neural network model that handle long time dependencies. It mainly remember information for a longer period of time. The flow of information within LSTM is operated by multiple gates which are composed of sigmoid gate. It is also called forget gate. The output in this gate will be either 0 or 1.

7.6.3 Main Methods. The solution is based on deep learning. Here, a chrome extension is designed which will analyze the user's profile as well as the other networks related to news dissipation. It uses both machine learning and deep learning method to extract necessary data from the profiles. This approach consists of three different phases. i. In phase 1, Information about a user is collected using crawler and Facebook API. Also, the collected data is preprocesses and refined in that phase. Here, crawler extracts the public information available on a user's profile. Public information includes friends, followers, number of likes and choice of interests and other information. Crawler also uses a Facebook API to access private information on request. Crawler uses chrome environment and uses the following algorithm to extract information:

Algorithm: Collected dataset based on the extracted feature (UCF features and NCF features) Input: Facebook users list Output: Generation of a labeled dataset for each Facebook profile in the list do Crawler extracts Facebook profile features, Store features in an EXCEL file; Extraction of new features by using Facebook API and crawler is done based on the contents available; for each UCF by multiple profiles i.e. Post shared do Extract set of the crude dataset from multiple features like Messages, Sharedmsg, Comment, Likes, Multimedia content shared, Follower, Followings from FBp as an input parameter to the crawler

Store features in an EXCEL file; End for each NCF by multiple profiles i.e. News Post shared NPshared, do Extract set of the crude dataset from multiple features like Source, Bodytext, Headlines, Texts, Wordcount,Sentence from FBp as an input parameter to the crawler Store features in an EXCEL file. End Parse EXCEL files and calculate the total and average values for each user post Performance are analyzed based on the stored files. End

Executing the following algorithm, the crawler stores all the information of the user in a XLS file. The XLS file gets updates each time when the user visits any other profiles. ii. In phase 2, all the collected data passes through deep learning and machine learning based classifiers. iii. In the phase 3, the most suitable classifier is selected for fake news detection.

7.6.4 Accuracy. Only using machine learning algorithms doesn't provide a better solution for detecting fake news. But using deep learning methods combined with machine learning extends the accuracy to a huge rate. Here, the confusion matrix for the resulted experiment is given below:

Accuracy Classification in the form of actual output Fake News Real News

Fake news Real News True positive False Negative False Positive True Negative

Table 2: Confusion matrix for the resulted experiment(Sahoo S.R., Gupta B.B., 2020)

Based on the experimental data of real-world information suggests that the proposed solution here has a higher accuracy than the existing state of art techniques. The approach is a combined solution of machine learning and deep learning techniques. Here the accuracy is given for specific algorithms:

Table 2: Accuracy of multiple features based solution

Algorithm	Accuracy
KNN	99.3%
SVM	99.3%
Logistic regression	99.0%
Decision tree	99.1%
LSTM	99.4%

In machine learning environment, SVM provides better accuracy than the remaining algorithms.

7.7 Unsupervised Learning Framework (UFD)

7.7.1 Domain. Twitter (in twitter there are many tweets in news category which can be noticed and gathered. For example, by using the advanced search API of Twitter with the title 'news').

7.7.2 Theory. Proposed detection method usually works with fake news which are unsupervised on Twitter by utilizing user's erratic involvement facts. The framework proposed in this paper called UFD (Unsupervised Learning Framework) that uses a probabilistic graphical model in order to framework truthiness and credibility. For solving the inference issue Gibbs sampling method is suggested. Two real world social media datasets (LIAR and BuzzFeed) used to conduct experiments. Firstly, authors have introduced the hierarchic social engagement framework, secondly presented difficulty features and lastly created a Bayesian network by formalize the problem.

Probabilistic Graphical Model: Probabilistic graphical model is a framework to represent complex domains using probability distributions, with numerous applications in machine learning. Gibbs Sampling: Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution. Bayesian Network: Bayesian network is a probabilistic graphical model to represent knowledge about uncertain domain where every node corresponds to a random variable and every edge represents the conditional probability for the corresponding random variables.

7.7.3 Main Method. The collected information from the news tweets holding the contents of the tweet like title, a link, a photo and user's own textual contented. News tweet also has likes, retweets, and replies from second-level users. Authors have divided two groups of users as verified and unverified. It has three-fold benefit. Firstly, it can be beneficial to reduce the noise of data as there are so many unverified accounts those are not participate in so many social engagements. Secondly, an implied supposition be forced in the paper which indicates verified users those who have huge number of followers, effects, and more standing become more reliable in differentiating fake or real news. Finally, reduction of the model is the third advantage proposed by authors. By concentrating on a small portion of data, simplification of the problem would be easier as the data collected from the social media is complicated, noisy, and incomplete. For the Problem Model section authors denoted N as set of news, and M and K respectively verified and unverified users. Every given news i N, and authors have collected users tweets for the following content who are verified. Mi M indicate the set of users who are verified and posted tweets for the news. The tweet of every verified user j Mi, authors have collected the unverified user's communal activities. Let Ki, j K indicate the set of users who are unverified and participate in the tweet. For every news i, authors have used a l potential arbitrary variable xi 0, 1 to indicate truthiness, for example, fake news (xi = 0) or true news (xi = 1). Even if the news can be true or fake, users' opinions is needed to be extracted from their engagement behaviors on the news.

Fake news detection is straightforward with the collapsed Gibbs sampler having update rule. Firstly, authors estimated the news truth by initializing randomly to either 0 or 1, and compute the sum up of every verified and unverified user depending on previous estimated truth result. Afterwards, the authors perform sampling procedure repeatedly. In each iteration, authors estimated every news sampling from its distribution conditioned on present approximation of every other news stated and accordingly update the sum up of every user.

Table 3: Accuracy for both LIAR and BuzzFeed dataset

Methods	Accuracy on LIAR	Accuracy on BuzzFeed
Majority Voting	0.586	0.556
TruthFinder	0.634	0.554
LTM	0.641	0.465
CRH	0.639	0.562
UFD	0.759	0.679

7.7.4 Accuracy. For the experiment authors have used two datasets which are public. For example, LIAR (Wang 2017) and BuzzFeed News for evaluating performance of the method. One of the largest fake news datasets is LIAR, holding above 12,800 minified news. 1,627 news articles are being hold by BuzzFeed connected with U.S. election from Facebook. Authors have used the consecutive benchmarks to evaluate execution of the detecting method: accuracy, precision, recall, and F1-score, could be commonly used for evaluating performance of classification methods. The proposed algorithm is compared with four other unsupervised fake news detection benchmarks which are Majority Voting, TruthFinder [19], LTM [21], CRH [**?**]. Table 4 and Table 5 shows experiment outcomes of LIAR and BuzzFeed Datasets respectively.

8 DISCUSSION AND CONCLUSION

As most of the fake news region are based on social media and online news article, we analyzed both areas detection articles and

come across the below comparative table. All of the mentioned methodologies belong to social media region but only no. 3 also belongs to online news article. It can be assumed from the table that on social media fake news mostly spread through either Facebook or Twitter as the number of users in these two platforms are huge. So, all the articles that we analyzed are tested one of these media platform. We find that for detection of fake news, linguistic, data science and machine learning techniques are the most commonly used. Na¨ıve bayes, Multiple Feature which uses NLP, Random Forest, SVM and ML + Deep Learning accordingly have an accuracy of 75.4% and 99%. So, it is obvious that if one wants to work with Facebook originate news, he/she can work with these methods but for verifying a news based on its accuracy Multiple Feature methodology will be the best. For twitter, we analyzed Propagation path classification, Content based method and Unsupervised learning. Propagation path classification uses convolutional network carrying an accuracy of 85%, Geometric deep learning is based on deep learning methods including Euclidean convolutional network having an accuracy of 93% and lastly, we have unsupervised learning which has a highest accuracy of 75.9%. Research work on twitter-based news can be done by using these methods efficiently but for verifying a news it is suggested to go for Geometric Deep Learning methods as it gives highest accuracy. A comparative table is given in the next page for better clarification:

Table 4: Comparison between various fake news detection methods

No	Method name	Domain	Area	Accuracy
1	Propagation path classification	Convolutional Network	Weibo, Twitter	85% on Twitter 92% on Weibo
2	Naïve Bayes	Naïve Bayes	Facebook	75.40%
3	Content Based	Natural Language Processing, Random Forest, SVM (Support Vector Machine)	News Article, Twitter	Non textual: 73%
4	User Based	N/A	Social Media	N/A
5	Geometric Deep Learning	Deep Learning, Euclidean convolutional network	Twitter	93% ROC
6	Multiple Feature Based	Machine Learning + Deep Learning	Facebook	99%
7	Unsupervised learning	Unsupervised learning framework	Twitter	75.9 % (LIAR Dataset) 67.9 % (Buzzfeed Dataset)

Among all the approaches given in the following table, user-based approach does not have any actual experimental accuracy value as it is still under development and not yet have been implemented. So, further research is needed to use this as a fake news detection method. Also, all the approaches given here works differently based on the situation and accuracy which also varies in different platforms. So, there is an opportunity to work with an AI based solution which can actually work uniformly in all social media accounts by learning all the characteristics all by itself.

REFERENCES

[1] Jonathan Albright. 2017. Welcome to the era of fake news. *Media and Communication* 5, 2 (2017), 87–89.
[2] Matthew RX Dentith. 2016. The problem of fake news. *Public Reason* 8, 1-2 (2016).
[3] Mykhailo Granik and Volodymyr Mesyura. 2017. Fake news detection using naive Bayes classifier. In *2017 IEEE first Ukraine conference on electrical and computer engineering (UKRCON)*. IEEE, 900–903.
[4] Adnan Hussein, Farzana Kabir Ahmad, and Siti Kamaruddin. 2019. Content-Social Based Features for Fake News Detection Model from Twitter. *International Journal of Advanced Trends in Computer Science and Engineering* 8 (12 2019), 2806–2810. https://doi.org/10.30534/ijatcse/2019/20862019

[5] David O Klein and Joshua R Wueller. 2018. Fake news: A legal perspective. *Australasian Policing* 10, 2 (2018).

[6] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[7] Yang Liu and Yi-Fang Brook Wu. 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Thirty-second AAAI conference on artificial intelligence*.

[8] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673* (2019).

[9] Jane Wambui Waweru Muigai. 2017. Understanding fake news. *Washington Post* (2017).

[10] Bhawna Narwal. 2018. Fake news in digital media. In *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. IEEE, 977–981.

[11] Özlem Özgöbek and Jon Atle Gulla. 2017. Towards an understanding of fake news. In *CEUR workshop proceedings*, Vol. 2041. 35–42.

[12] Thorsten Quandt, Lena Frischlich, Svenja Boberg, and Tim Schatto-Eckrodt. 2019. Fake news. *The international encyclopedia of journalism studies* (2019), 1–6.

[13] Somya Ranjan Sahoo and Brij B Gupta. 2021. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing* 100 (2021), 106983.

[14] Kai Shu and Huan Liu. 2019. Detecting fake news on social media. *Synthesis lectures on data mining and knowledge discovery* 11, 3 (2019), 1–129.

[15] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[16] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding user profiles on social media for fake news detection. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 430–435.

[17] Edson C Tandoc Jr. 2019. The facts of fake news: A research review. *Sociology Compass* 13, 9 (2019), e12724.

[18] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. 2018. Defining "fake news" A typology of scholarly definitions. *Digital journalism* 6, 2 (2018), 137–153.

[19] Xiaoxin Yin, Jiawei Han, and S Yu Philip. 2008. Truth discovery with multiple conflicting information providers on the web. *IEEE Transactions on Knowledge and Data Engineering* 20, 6 (2008), 796–808.

[20] Xichen Zhang and Ali A Ghorbani. 2020. An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management* 57, 2 (2020), 102025.

[21] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 836–837.