iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4507

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

A Survey on Identification of Fake News

Shubham Gupta[1], Dishant Grover[2], Anushka Gupta[3], Deepanshu Kapoor[4], and Narina Thakur[5]
[1, 2, 3, 4, 5]*Bharati Vidyapeeth College of Engineering, New Delhi*

Abstract: Fake News is a major problem faced by many agencies and organisations which may seem like a corporate issue but is a threat to the very idea of democracy and every citizen's right to information. It is one of the major false practice issue growing at an enormous rate on online social media platforms disrupting the journalistic practices of spreading the truth to the people unaware of the background of news. Stanford University, in one of its recent studies, has provided a definition of fake news, "We define fake news to be news articles that are intentionally and verifiably false, and could mislead readers". News sensationalism, falsification, and exaggeration are the first steps in spreading these pieces of propaganda. It is one of the biggest problems for news organisations which are trying to report truth to the average citizen but are slowed down in the process due to some culprits for their own gains and compromising all the journalistic effort put in by all the people behind the news cycle. Machine Learning and Data Science are contemporary fields of computer science which are able handle complex problems and are tackling those problems head on. Ours is an effort to use these vast technologies to solve the problem of fake news on social platforms like Twitter.

I. INTRODUCTION

Twitter is a widely used online social platform used by people of every generation and every field. Due to its widespread reach of more than 330 million users, it accumulates all forms of data on its website. Due to this, it is also considered as the 21st-century newspaper because of every users' ability to report major events and incidents across the world. People share articles, videos, photographs, and news through Tweets. Almost 85% of the topics discussed on Twitter are related to news. This paper primarily focuses on the spread of Fake News through short posts called Tweets. Twitter is a free social networking micro-blogging platform that allows users to broadcast Tweets. Twitter has a small size of stories of 280 characters (previously 140) which allows to skim a lot of them fast and hence compensates for them not being ranked by quality. Also, the retweet mechanism is efficient for redistribution of Tweets. The hashtags and trending topics make the search works easier and also help people to look for the events happening [24]. However, this freedom to the user to post anything leads to spread of fake information also [25]. People are also less likely to check the news shared by their friends and hence it leads to spreading of false misleading information at a faster rate [26].

Fig.1 Some examples of fake news

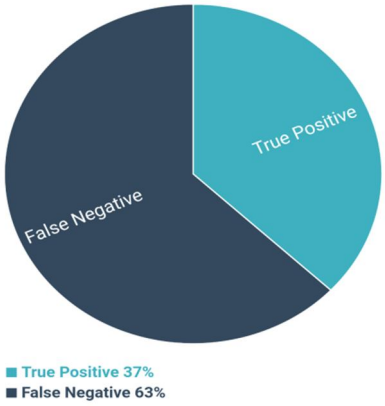
Fake news can lead to chaos and worriment among people. Fake images also cause the same problem. A recent study showed that Tweets containing fake images were mostly retweets (86%). Very few users posted original Tweets with fake images. URL shorteners are generally used to spread fake news.

It is important to think and evaluate the news critically and it can be done by encountering a claim in the news by checking it against reliable sources and everyone should do their part in not spreading it further through social media, e-mails or conversation. Given

the prevalence of this new phenomenon, “Fake news” was even named the word of the year by the Macquarie dictionary in 2016. Many incidents have been previously reported of Fake News causing chaos and troubles in certain parts of India and even at an International level. Last year, an angry citizen attacked a restaurant in daylight with a rifle because he read a Tweet which reported that, in that restaurant, some children was harbored as sex slaves, as part of a child-abuse ring led by a major political representative. Luckily no one was harmed. A Tweet by the Swarajya columnist Shefali Vaidya said that a foreign national, was a member of the National Advisory Council in the UPA government. The Tweet went viral but its allegations weren't new. Then the respected development economist, and activist had to write a letter clarifying that he's an Indian citizen. This, despite living in India for more than three decades and being a citizen of India for 15 years. However, a year later, the fake news is still doing the rounds. These incident and many others are examples of order of chaos, one simple Tweet, a Facebook post or a blog post can cause. This is undermining every great innovation happened in online space with social networks. In the following sections, we propose how we plan to do it and how others have done it in the past and how by learning from their efforts we are taking a step to prevent this.

A. Survey

We conducted a survey among students in the age group of 18-21. The students were provided a verified fake news on twitter and they were asked to categorize it as true or false i.e. true news or fake news.
Results are given below



From the results, we infer that with a probability of 0.37 students could identify the fake news, which is a very low probability and hence the need to identify fake news.

II. LITERATURE REVIEW

Social media is not just used to share opinions but also news, as well as rumors and fake news. As there is no strict check on tweets, people tend to post fake news for their own benefit. A lot of people retweet it and it spreads like an epidemic. The spread of fake news and propaganda has been a bane to the industries while causing a trouble to journalists and different social media platforms trying to use their work for better causes but they’re being disrupted by the horrendous people present on the Internet. Many people have been in the loop for trying to lessen the harmful effects it is causing. Fette et al. used machine learning to classify an email as phishing or not by using features such as age of URL, number of dots in URL and HTML content of email while obtaining a high accuracy of 99.5% [1]. Bajaj et al. used Two-Layer Feed forward Neural network, Recurrent neural network, Convolution neural network, Long-Short term memories, Gated recurrent units and observed that the RNN architecture with GRUs outperformed one with LSTM cells for the detection of fake news. The feed forward network performed well as a classifier but the convolution network didn’t perform well [4]. Yang et al. presented a trailblazing vision on how criminal accounts tend to be socially connected forming a small world network. Criminal accounts randomly follow accounts, expecting them to follow back. There are three types of criminal supporters found namely, social butterflies, social promoters and dummy [2]. Conroy et al. used linguistic approach (Both words and syntax) as well as network approach which says that message metadata or structured knowledge network queries can be harnessed to provide aggregate deception measures. They used Centering Resonance Analysis(CRA), a mode of network-based text analysis [3]. Gupta et al. presented a path breaking work by characterizing and identifying fake images on Twitter during Hurricane Sandy. They said that 86% of the Tweets containing fake images were retweets. There were very few instances of users posting original Tweets with fake images. URL shorteners were used to spread the fake news (bit.ly). They mentioned that the best

indicators of credibility were URLs, mentions, retweets and Tweet length. User Level features (number of friends, followers and status messages of the user) and Tweet level features (words, URLs, hashtags, is Tweet a reply or a retweet) were used and best results were obtained from Tweet level features [5]. In a follow-up work, Aditi Gupta wrote on credibility ranking of Tweets during high impact events as 30% of the Tweets posted about an event contained situational information about the event while 14% was spam. They used linear logistic regression, SVM ranking algorithm and supervised learning. They also formed a chrome extension for Twitter which gives a score(out of 7) to each Tweet based on its credibility [9]. Shao et. al. investigated the temporal relationship between the spread of misinformation and fact checking using a public tweet dataset . The dataset was formed by collecting tweets that contained URLs from fact checking domains as well as fake news domains. This research supports that URLs are the most common method to share news articles on platforms that limit the amount of user expression (such as Twitter's imposed 140 character limit) . Their implemented system, Hoaxy, provides an automated solution for identifying fake-news stories by comparing them against known news sources and fact checking websites. The research found there are a few highly active users responsible for fake news on social media [6].

Shu et al. said that malicious accounts can be easily and quickly created to boost the spread of fake news, such as social bots, cyborg users, or trolls. The existing algorithms for detection of fake news are either i). News Content Based or ii). Social Context Based. News content based approaches focus on extracting various features in fake news content, including knowledge-based and style-based. Social context based approaches aim to utilize user social engagements as auxiliary information to help detect fake news [7]. Another related work was done by Kang et al. who evaluated effects of individual factors and said that metadata and image type elements are the strongest influencing factors in credibility assessment. They surveyed 183 people through MTurk and did an analysis through various graphs namely Likert scale, Hinton map, and heat map. They included 11 factors, however, did not consider personal factors [14]

Janze et al. trained a variety of machine learning classifiers suitable for binary classification problem specially Logistic regression, Support vector machines, Random forest and extreme gradient boosting. They evaluated the classification models (LOG, SVM, DTR, RFO and XGB) via different metrics which are based on a stratified 10-fold cross validation approach. They divided the data set of n=460 posts into 10 equally sized folds containing the same amount of fake and non-fake observations randomly selected from the total sample. Then, they took out one fold and trained the models with the nine remaining folds. They obtained an accuracy of nearly 80% [8]. Galuba et al. focussed on characterizing and modeling the information cascades formed by individual URL mentions in the Twitter follower graph. They constructed two models of information propagation in social networks namely At-least-one (ALO) model and Linear Threshold model. The linear threshold model was able to correctly predict almost half of the URL mentions [10]. Aggarwal et al. formed an automatic real-time phishing detector on Twitter using random forest classification on certain fields of Twitter together with URL and WHOIS features and obtained an exemplary accuracy of 92.5%. One of the major contributions of their work was the Chrome Extension they developed and deployed for real-time phishing detection on Twitter [11]. Cook et al. examined the online fake electoral personas in Australia. He derived a nine way test that includes : entropy test, Spam and miscreant test, account properties, account created date (if it was near election date), inactivity before the election, inactivity after the election, follower alignments, mass retweets on policy specific days and times and Discrimination analysis (combining entropy, spam and account, inactivity, alignments and mass retweets) [12]. Allcott et al. drafted that fake news spread during 2016 US elections was all in favor of electing Donald Trump as the president of the US. They conducted a survey to collect the data. The data obtained said that Trump might not have won if there was not so much fake news about him. The websites spreading fake news tend to be short-lived and the major reason behind producing fake news was to generate advertising revenues when news goes viral and some ideological reasons. A US company called Disinfomedia owns many fake news sites. Unintentional reporting mistakes, rumors, conspiracy theories, satire and false statements by politicians were all ruled out from their model. They mainly used linear regression and prediction [13] .

Benevenuto et al. classified real YouTube users, as spammers, promoters, and legitimates. They used techniques such as supervised machine learning algorithms to detect promoters and spammers; they achieved higher accuracy for detecting promoters; the algorithms were less effective for detecting spammers [15]. Ghosh et al. characterized social farming on Twitter, and also proposed a methodology to combat link farming [16]. Pisarevskaya et al. aimed to reveal fake and truthful news using markers from different linguistic levels. They used POS tags, length of words, sentiment terms and punctuation on the lexics level. They also used Rhetorical Structures Theory(RST) relations as markers on the discourse level [17]. Yardi et al. published that spams can be filtered by examining the content and tracking IP address. They found that spammers have the high following to friends ratio, post more Tweets per day(spammer mean= 8.66) as compared to legitimate accounts (legitimate mean= 6.7) and use more hashtags. A total number of followers and friends for spammers was three times that of legitimate users. Popularity and legitimacy are indicated by

high indegree and spam is indicated by high outdegree. However, spammers may try to game the system by auto-following users and then unfollow them to invert their followers/friends ratio [18]. In another follow-up work, Yardi studied Twitter's behavioral and structural properties and how they make it a fertile breeding ground for spammers to proliferate. She collected 17,803 Tweets from 8,616 unique users using the #robotpickuplines and was the first one to examine Twitter memes and spam based on network and temporal properties. She found out that the spam accounts are not significantly newer and exhibit slightly higher retweet and reply properties. Spammers tweeted and replied slightly more frequently than legitimate users in her dataset. Spammers were likely to be found at the edges of the Twitter graph rather than the center. Chen et al. said that the line between traditional media content and user generated content is being blurred. The rumor is being presented as truth in headline by formulating it as a question and advertised content is presented as native content [19]. Lazer et al. identified the source of fake news and offered feedback to users that a particular news may be fake. They detected information being promoted by bots and 'cyborg' accounts . Also, they found that older and more extreme individuals on the political spectrum appear to share fake news more than others [20]. Sakaki et. al analyzed Twitter logs of people's actions in a time of a catastrophic disaster and attempted to extract what happened in the emergency situation like The Great Eastern Japan Earthquake, which struck Japan on March 11,2011 [21]. Zubiaga et al. also studied about the real and fake images that spread during various natural disasters including the Hurricane Sandy [22]. Castillo et al. used Decision Tree classification to find out the newsworthiness of published content and for credibility, J48 method was used [23].

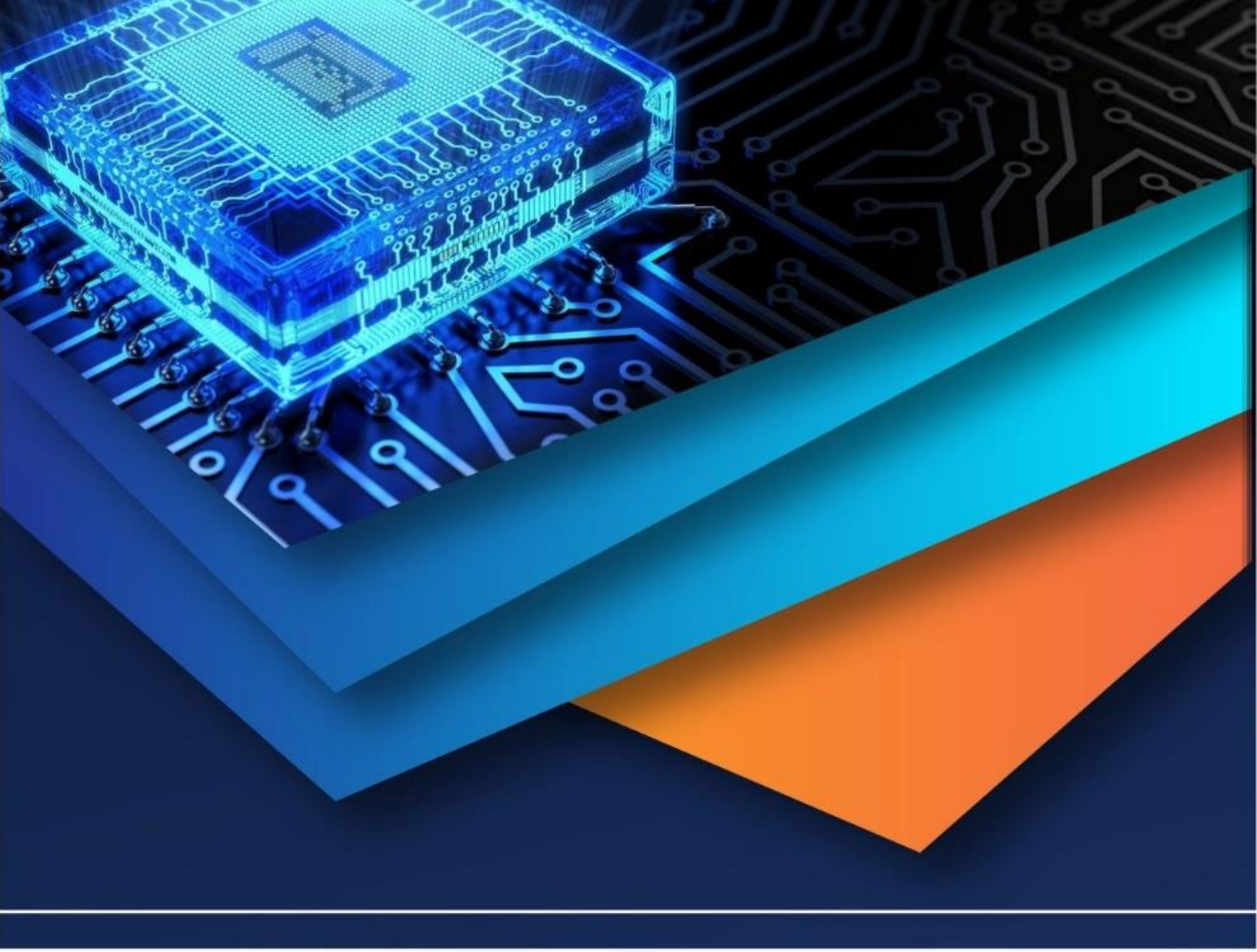
III. CONCLUSION AND FUTURE SCOPE

In the course of our study, we realised that people have blind trust on posts published on social media and hence it is very important to eliminate the problem of fake news from social media. In this survey, we presented an overview of different models and algorithms employed for identification of fake news on social media. Based on the knowledge extracted from the studied research papers and some more techniques, we also proposed a system to identify fake news on Twitter. We plan to work on detection of fake news on Twitter using Natural language processing, divide and conquer approach and fuzzy matching technique. However, apart from fake news, rumors and clickbaits are also used for spreading misinformation.

REFERENCES

[1] Fette, N. Sadeh, and A. Tomasic, "Learning to detect phishing emails," in Proceedings of the 16th international conference on World Wide Web. ACM, 2007, pp. 649–656.
[2] Chao Yang, Robert Harkreader, Jialong Zhang, Seungwon Shin, and Guofei Gu. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In Proceedings of the 21st international conference on World Wide Web, WWW '12, 2012.
[3] Niall J.Conroy,Victoria L.Rubin, Yimin Chen. Automatic deception detection : Methods for finding fake news. Crossref,2015.
[4] Bajaj, Samir. ""The Pope Has a New Baby!" Fake News Detection Using Deep Learning."
[5] Ponnurangam Kumaraguru, Aditi Gupta, Hemank Lamba. Faking sandy : Characterizing and identifying fake images on twitter during Hurricane Sandy. ACM,2013.
[6] C. Shao and G. L. Ciampaglia. "Hoaxy: A platform for Tracking Online Misinformation. 4 Mar. 2016.
[7] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD Explorations Newsletter19.1 (2017): 22-36.
[8] Janze, Christian, and Marten Risius. "Automatic Detection of Fake News on Social Media Platforms." (2017).
[9] Aditi Gupta and Ponnurangam Kumaraguru. Credibility ranking of tweets during high impact events. In Proceedings of the 1st Workshop on Privacy and Security in Online Social Media, PSOSM '12, pages 2:2–2:8, New York, NY, USA, 2012. ACM.
[10] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer. Outtweeting the Twitterers - Predicting Information Cascades in Microblogs. In Usenix Workshop on Online Social Networks, (WOSN'10), 2010.
[11] Ponnurangam Kumaraguru Anupama Aggarwal, Ashwin Rajadesingan. Phishari: Automatic realtime phishing detection on twitter. 7th IEEE APWG eCrime Researchers Summit (eCRS), 2012.
[12] Cook, David M., et al. "Twitter deception and influence: Issues of identity, slacktivism, and puppetry." Journal of Information Warfare 13.1 (2014): 58-IV.
[13] Hunt Allcott, Matthew Gentzkow. Social media and fake news in the 2016 election. NBER, January 2017.
[14] Byungkyu Kang, Tobias Hollerer, John O'Donovan. Believe it or not? Analyzing information credibility in microblogs. ASONAM'15 Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015
[15] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on Twitter. In CEAS, 2010.
[16] Saptarshi Ghosh, Bimal Viswanath, Farshad Kooti, Naveen Kumar Sharma, Gautam Korlam, Fabricio Benevenuto, Niloy Ganguly, and Krishna PhaniGummadi. Understanding and combating link farming in the twitter social network. In Proceedings of the 21st international conference on World Wide Web, WWW '12, 2012
[17] Pisarevskaya, Dina. "Deception Detection in News Reports in the Russian Language: Lexics and Discourse." Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism. 2017
[18] Sarita Yardi, Daniel Romero, Grant Schoenebeck, and Danah Boyd. Detecting spam in a Twitter network. First Monday, 15(1), January 2010
[19] Yimin Chen, Niall J.Conroy, Victoria L.Rubin. News in an online world : The need for an automatic crap detector. Crossref,2015
[20] David Lazer, Matthew Baum, Nir Grinberg, Lisa Friedland. Combating fake news : An agenda for research and action. 2017.

[21] Sakaki, Fujio Toriumi, and Yutaka Matsuo. Tweet trend analysis in an emergency situation. In Proceedings of the Special Workshop on Internet and Disasters, SWID '11, pages 3:1–3:8, New York, NY, USA, 2011. ACM.
[22] Arkaitz Zubiaga, Heng Ji. Tweet, but verify : epistemic study of information verification on Twitter. Springer Vienna, 2014.
[23] Carlos Castillo, Marcelo Mendoza, Barbara Poblete. Information credibility on Twitter. WWW'11 Proceedings of the 20th international conference on World wide web, April 2011
[24] Haewoon Kwak, Changhyun Lee, Hosung Park and Sue Moon. What is twitter, a social network or a news media? Raleigh, NC, USA, 2010.
[25] Chris Grier, Kurt Thomas, Vern Paxson, Michael Zhang. @spam : The underground on 140 characters or less. 17th ACM conference on Computer and communications security, 2010
[26] Chris Grier, Kurt Thomas, Vern Paxson, Michael Zhang. @spam : The underground on 140 characters or less. 17th ACM conference on Computer and communications security, 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY