A Review on Rumor Detection Methods on Social Media: Issues and Challenges

Swati Goswami[1], Chetan Agrawal[2], Bhavana verma[3],

[1, 2, 3]Radharaman Institute of Technology & Science, Bhopal, India

Abstract- *Web based life stages have been used for information and news gathering, and they are completely huge in various applications. Regardless, they in like manner lead to the spreading of gossipy goodies, Rumors, and fake news. Various undertakings have been taken to perceive and uncover reports employing long-range interpersonal communication media through analyzing their substance and group environment using ML (Machine Learning) procedures. This paper gives a diagram of the continuous examinations in talk discovery. The task for talk recognition intends to recognize and portray tattle either as self-evident (authentic), false (nonfactual), or dubious. This can massively benefit society by thwarting the spreading of such mixed-up and misguided information proactively. This paper is a prologue to gossip acknowledgment employing person-to-person communication media which presents the fundamental phrasing and sorts of pieces of talk and the nonexclusive methodology of gossip identification. A state of the art depicting the use of coordinated ML algorithms for talk identification employing Social media is presented.*

Keywords:- Rumor Detection Issues, Rumor Classification Challenges, Social Media, Machine Learning.

Introduction

With the inception of Web 2.0 and the increasing ease of access methods and devices, more and more people are getting online, making the Web indispensable for everyone. The focal point of innovation of Web 2.0 is social media. Active participation is a key element that builds social media. Numerous social networking platforms as Twitter, YouTube, and Facebook have become popular among the masses. It allows people to build connection networks with other people and share various kinds of information in a simple and timely manner. Today, anyone, anywhere with an internet connection can post information on the Web. But like every coin has its two sides, this technological innovation of social media also has some good as well as bad aspects. We are benefited from social media but we cannot oversee its negative effects on society. The majority of citizens esteem it as an innovator discovery and a few seem to receive it as an unenthusiastic bang on the civilization. As a positive case, these online communities facilitate communication with people around the globe regardless of your physical location. The perks include building connections in society, eliminating communication barriers, and helping as effective tools for promotion, whereas on the flip side, privacy is no more private when sharing on social media.

Due to the ubiquitous and overdependence of users on social media for information, the recent trend is to look and gather information from online social media rather than traditional sources. But there are no means to verify the authenticity of the information available and spreading on these social media platforms thus making them rumor breeding sources. The standard definition of the rumor is: any section of data publicize in a community lacking adequate facts and/or confirmation to sustain it thus putting a query on its legitimacy. It

might be accurate, bogus, or indeterminate and is generated purposely (awareness seeking, self-objectives, finger-pointing someone, hoax, to extend terror, and disgust) or by accident (mistake). Further, these might be private as well as commercial. Knapp [1] classified rumors into three categories, namely pipe dream, bogy, and wedge driving for describing intentional rumors.

Rumors are circulated and believed overtly. And due to the increasing reliance of people on social media, it is inevitable to detect and stop rumors from spreading to reduce their impact. It gets only a few minutes for a single tweet or post to go viral and affect millions. Thus, rumor detection and mitigation have evolved as a recent research practice where the rumor has to be recognized and its source has to be identified to limit its diffusion. It is essential not just to detect and deter, but to track down the rumor to its source of origin. Various primary studies with promising results and secondary studies [2, 3] have been reported in this direction. The work presented in this paper is a primer on rumor detection on social media to explicate the what, why, and how about rumor detection on online social media. The intent is to aid novice researchers with a preliminary introduction to the area and at the same time, offer background work to the experts. The types of rumors and the typical process of rumor detection are discussed followed by a state-of-the-art review of supervised ML-based rumor detection on online social media. The research gaps have been identified as issues and challenges within the domain which make it an active and dynamic area of research.

The rest of the paper is organized as following: section 2 explains how rumors can detect on social media with its types, section 3 describes various methods of rumor detection like machine learning and deep learning methods, section 4 describes various challenges and issues during the rumor detection, section 5 explains open future research directions on which current research is going on or can be done in near future, finally we conclude our work in section 6 following with references used in this work.

II. Rumor Detection on Social Media

Social media has the power to make any information, be it true or false, go viral, and reach and affect millions. Due to the speed of information spread, even rumors are spread. Hence, it is necessary to detect and restraint these rumors before they have a serious impact on people's lives.

2.1 Types of Rumors

A rumor is defined as information whose veracity is doubtful. Some rumors may turn out to be true, some false, and others may remain unverified. Not all false information can be classified as a rumor. Some are honest mistakes by people and are referred to as misinformation. On the other hand, there may be intentional rumors put to mislead people into believing them. These are labeled as disinformation and are further classified based on the intent of the originator. The following Fig. 1 depicts the classification of rumors. We define a rumor as any information put out in public without sufficient knowledge and/or evidence to support it. It is misleading, either intentionally or unintentionally.

If some information has been put out in public erroneously without authentic or complete information with no ulterior motive of hurting or causing any disturbance to anyone whatsoever, it is called misinformation. It is an honest mistake. Disinformation, on the other hand, is information that is intentionally put out in public view to mislead people and start a false rumor. Disinformation depending on the motive of the writer and nature of the post can be classified as a humorous, hoax, finger pointing, tabloids, and yellow press. The most harmless type of rumor is the humorous ones. Sources spreading this type of information fabricate news and stories to give it an amusing side. The motive is usually to entertain people. The information is pre-declared to be false and intended only for comical purposes. The best examples of such sources include news satires and news game shows. The next form of disinformation is a hoax. A hoax is intentional fake news spread to cause panic among people and cause trouble to people at whom it is aimed. A hoax can also be an imposter. Examples include fabricated stories, false threats, etc. In 2013, a hoax stating Hollywood actor "Tom Cruise to be dead" started doing the rounds. Social messaging apps like WhatsApp worsen the situation when it comes

to hoaxes. A currency ban of Indian rupees 500 and 1000 was done in November 2016. Soon after a hoax message went viral on WhatsApp stating that the government will release a new 2000 rupee denomination that would contain a GPS track able nanochip that would enable to locate the notes even 390 feet buried underground. The government and bank spokespersons had to finally issue an official statement stating it was false. Still, many people found the official statement hard to believe as they were so brainwashed by the hoax message.
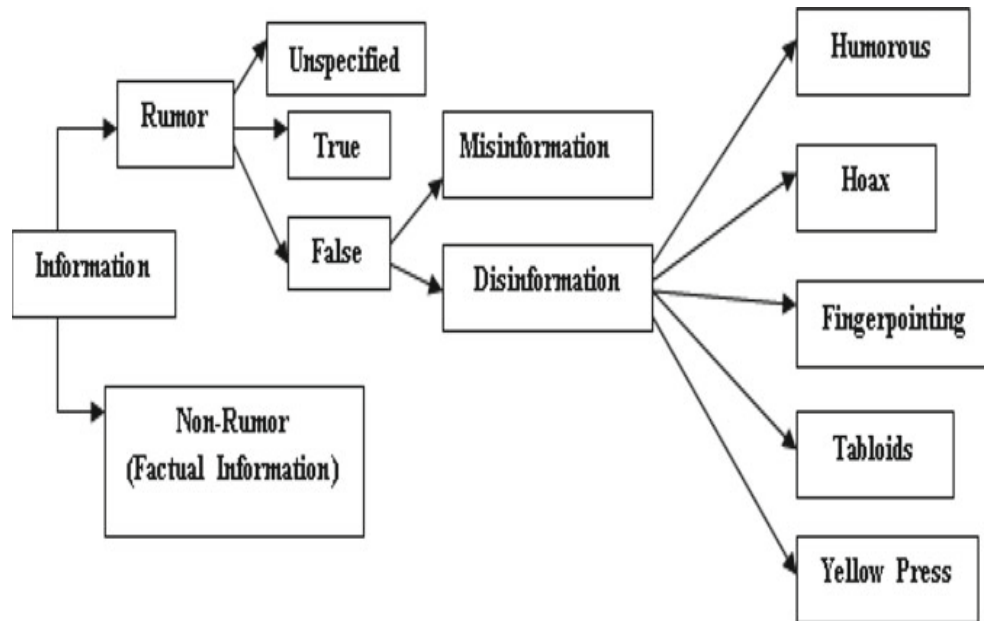
Fig 1: Classification of rumors.

Another form of disinformation is finger pointing. Finger-pointing always has an associated malicious intent and personal vested interest. It blames a person or an organization for some bad event that is happening or happened in the past. It aims at a political or financial gain by tarnishing the image of the target person/ organization/party/group, etc. Tabloids have had a bad name for spreading rumors since when they started. It is the type of journalism that accentuates sensational stories and gossips about celebrities that would amount to spicy page 3 stories. Yellow press journalism is a degraded form of journalism that reports news with little or no research at all. Journalists' only aim is to catch attention using catchy headlines with no regard whatsoever to the authenticity of the news. They do not bother to delve deep into a story but just publish it to sell as many stories as possible and make money. It is the most unprofessional and unethical form of journalism.

III. Rumor Detection Approaches

There have been various efforts in the field of rumor detection and mitigation. Many authors have used simple cue-based, network-based, Psycho and social theory-based approaches whereas many other have used machine learning approaches. Many other studies have incorporated different aspects and their methodology is an amalgamation of various techniques. There has also been a debate around which features are most important in detecting a rumor. This has led to a new approach of deep learning where feature selection is not required for the efficient performance of the framework. Here, we discuss variously supervised, unsupervised, and other machine learning approaches, as well as the deep learning-based approaches in the field of rumor detection.

3.1. Machine Learning-Based Approaches

There have been various efforts in information credibility analysis in online social networks. As the dataset is an important characteristic of any problem to be solved in a Machine learning scenario, the early works were more focused on feature engineering. In one of the early works using machine learning, Castillo et al. [4] use algorithms including Decision trees like J48 and Random Forest, Support Vector Machines (SVM), and Bayes networks for evaluation of the credibility of a tweet. The input features to these algorithms were based on the characteristics of users, messages, propagation dynamics, and the topic in question. Based on the results obtained, they concluded that the topics of news (chats and opinions excluded) that are credible, are mainly single-sourced or a few-sourced; propagated through authors who have a long history of propagating like messages [4].

Another work on rumor analysis and detection by Yang et. al proposed two more features: 'client program used' and 'event location' [5]. They performed two experiments on data of Sina Weibo to study any improvement in the efficiency through the introduced features. One experiment was performed on the existing set of four features and the other was performed with the augment of the two features. The study concluded that the augmentation of the proposed features improved the overall accuracy of SVM from 72.5 % to 77 %. As the study done by Yang et. al [5] was carried out on data of Weibo, its validity on Twitter data was still a question due to the different nature of both the platforms. The need to create a standard benchmark dataset for rumor detection was felt and many researchers devoted their studies to the same.

In this pursuit, Qazvinian et al. [6] released an annotated dataset of Twitter microblog for rumor detection. This dataset contains tweets of five established rumors being investigated. This dataset, among many researchers, was also used by Hamidian and Diab [7] for rumor detection, by employing a multi-staged strategy (3-class classification followed by a 4-class classification) with a varying set of features and different pre-processing tasks. They added two Twitter and network features: Replay time (network-based) and time of posting the tweet (Regular day or busy day). They also added three pragmatic features: Named Entity recognition, Emoticon, and Sentiment. They used the J48 decision tree algorithm on Weka to carry out the experiments. The method that they used was different in the sense that in the common 6-class classification, a single step was involved in the detection and classification of the rumors while as in the 3:4 class classification, the detection was followed by classification. They reported that their 2-staged strategy (each for detection and classification) outperformed the single-stage strategy with a 14% increase in F1 Score on the Obama dataset.

There have been certainly contradicting conclusions with studies like ones carried out by Sahana et al. [8] stating that user-based features have very little significance or no correlation with the rumor detection while as some studies like one conducted by Castillo et al. [4] show that user based features enhanced the performance of the rumor detection system. Castillo et al. [4] also identified word frequencies as an important feature for rumor detection whereas Sahana et al. [8] stressed that content-based features are important for rumor detection. They reported an accuracy of 87.9% in their approach, using the J48 algorithm with 10-fold cross-validation for 10 iterations. The dataset they used, was based on tweets and retweets about the London riots. They also conclude that most active users are prone to rumor propagation as they retweet without establishing the credibility of a tweet. Another study on rumor detection was carried out by Kwon et al. [9]. They examine different rumor characteristics over a varying time window. They employed the variable selection process using the Random forest algorithm proposed by Genuer et al. [10] for selecting temporal, linguistic, user, and network-based features. The temporal window was kept as 3, 7, 14, 28, and 56 days from the onset of rumors. The authors proposed two algorithmic approaches, one with the user and linguistic features and the other with all of the features. It was observed that the user and linguistic features perform better to detect rumors at the onset whereas the structural and temporal features were beneficial in telling rumors from non-rumors.

Takahashi and Igata [11] explore the essence of among many features, a feature "retweet ratio" in rumor

detection. They conclude that although it remains inconclusive in the investigated sample, it may be beneficial for a large sample size. Another study presented by Jain et al. [12] to detect misinformation on Twitter uses mismatch ratio as the threshold for detecting whether a topic constitutes a rumor or not. The basic assumption in their study was that the verified news channels on Twitter would be very less prone to spreading rumors than any other user. Based on this assumption, they create two sets of tweets relating to a topic and calculate sentiment and contextual mismatch between them. If the value of the calculated function of mismatch (which is a ratio) is more than a threshold, they label the topic as 'rumor' and if the value is less, it is labeled as a 'nontumor. The authors concluded that the results were better if the tweets were less subjective and more objective. Chang et al [13] used a cluster-based approach for political rumor detection on the dataset consisting of two sets of tweets. One set consisted of tweets about Barack Obama in September 2015 and the other contained tweets related to Hillary Clinton, posted in August 2015. They identified 'extreme users', the ones tending to tweet false news and rumors. These users were identified on some features such as 'high tweeting frequency', 'a huge number of followers', 'use of extreme keywords in tweets' and over- enthusiasm' about the topic [13]. They use cosine similarity to club the clusters discussing the same news, after clustering the tweets containing the same URL as a link. They reported that the best rule derivations are subjective and thus differ from one case to another, as a function of the dataset.

3.2. Deep Learning-Based Approaches

Deep learning has proven to be very advantageous over traditional machine learning in various problems since it is almost immune to the feature selection problem. Deep neural networks need no fewer features to work efficiently and rather can perform well on unsifted features.

Ruchanski et al. [14] propose a three-module hybrid model for fake news detection. Their model is based on three steps or modules. The devised model focuses on textual, user response-based, and Source-based features. The module named 'capture' leverages Long Short Term Memory networks (LSTM). This module captures the temporal text and temporal activity of a user on a given article. The second module, Score, focuses on the source characteristics of the behavior of a user. It assigns a score to a user based on his tendency of participation in a particular source promotion group. The third module combines the result of the first two modules into a vector for classifying an article as fake or not-fake and is thus named as 'Integrate'. Ma et al. [15], in one of the earliest works of rumor detection with the aid of neural networks, apply recurrent neural networks to detect rumors. Based on their observation that a rumor is initiated from an original post (source) and a series of re-posts, relating posts and comments follow the original post, they utilize the time series concept to model rumor data. They treat a batch of posts falling in the same time interval as a single unit in time series and model the data using a recurrent neural network (RNN) sequence. For each interval, top-K values on term frequency-inverse document frequency (tf-idf) of vocabulary were taken as an input. Their model performs well than the contemporary manually selected feature methods.

Chen et al. [16] also use recurrent neural networks for the early detection of rumors. They use what they call an attention mechanism, in their models to understand the particular words that are important for a particular rumor category. They create batches of posts according to the time intervals and use tf-idf as the input representation. They conclude that the attention mechanism is efficient in detecting rumors and it results in ignoring unrelated words while giving less weight to the event-related words but more weight to the words expressive of a user's doubts and anger relating to the rumor. Yu et al. [17] propose a convolutional neural network (CNN) for misinformation detection. Based on the observation that RNN is incapable of detecting tumors at an early stage due to its bias towards the temporal sequence of input, they split a rumor into different phases. Then, they use doc2vec for vector representation generation subsequently used as an input to a two-layered CNN. Nguyen et al. [18] propose a model based on CNN and RNN for early detection of rumors. Apart from the time-series-based classification model, they use event credits for the prediction of rumors. In the proposed model, CNN is used to learn the hidden representations of specific tweets by

extracting a sequence of high-level phrase representations as input to LSTM, providing the tweet representation as an output. The output of this model (CNN+RNN) is then combined with a dynamic time series-based rumor discrimination model to get the final output. The authors report improved efficiency in classifying rumors in the early hours of the spread of a rumor.

IV. Issues and Challenges

Rumor detection comes with its share of issues and challenges. The main challenge for carrying out the rumor detection task is the collection of data. Even the most popular social media sites, namely Twitter and Facebook do not provide full freedom to users for extracting data. Most of the data posted on Facebook is private, hence inaccessible. Only data posted on Facebook pages can be collected. Twitter, on the other hand, these days do not allow data older than seven days to be fetched. Another issue faced by researchers in the detection of new rumors from real-time data. It is easier to detect old posts regarding a rumor that we know of because we know the keywords. But with emerging rumors, we are in a fix as we do not know what to look out for. Also, some rumors remain unspecified and there is no confirmation or debunking for them. Hence, detecting rumor veracity is very challenging. Another aspect that needs to be taken care of is the detection of the origin of a rumor as it is difficult to identify the user who started a particular rumor. These issues need to be addressed to improve the quality and speed of rumor detection.

V. Future Research Scope

Even though noteworthy advances have been made in exposing bits of Rumor through Social networking media, incidentally, there stay numerous difficulties to survive. Because of the survey of past examinations and our encounters in both research and down-to-earth framework execution of Rumor Detection, here we present a few bearings for future rumor discovery inquire about.

- **Knowledge Base:** Knowledge Base (KB) is useful for bogus news discovery [19]. There have been a few examinations on utilizing KB for bogus news discovery, however not many or none on rumor detection over online life. One explanation is that for bits of rumor via social networking media, we as of now have a lot of data, particularly the social media data, to misuse and do inquire about on. Another explanation is that, contrasted with bogus news recognition which chiefly manages news stories, bits of rumors through social networking media are about different subjects, and it is difficult to manufacture proper KBs that spread them. Along these lines, most past examinations on rumor recognition have not focused on abusing KB for exposing bits of rumors.
- **The target of User Response:** Client reactions are very instructive for rumor identification. Normally, bogus bits of rumor will get progressively negative and addressing reactions, which can be utilized for tumor detection. Each source message (gossip guarantee) has numerous answers, and they are either immediate answers or answers to different messages in the change string. The structure of the transformation string is significant for understanding the genuine position of the client of an answer. For instance, given the message "This is phony" and an answer to it "I concur" if we don't consider that the answer is towards "This is phony", at that point we will give an off-base position name, "support", to this answer. However, this reaction is denying the rumor guarantee. Although the neural system models dependent on engendering investigation may somewhat become familiar with this data, we think expressly handle this circumstance would improve rumor discovery execution.
- **Cross-domain and Cross-language:** Most past examinations stress recognizing bogus rumor tidbits from reality with trial settings that are commonly constrained to a particular internet-based life stage, or certain point areas, for example, legislative issues. Breaking down bits of rumor across points or stages would let us increase a more profound comprehension of bits of rumor and find the extraordinary qualities that can additionally help to expose them across areas (subject and stage).
- **Explanatory Detection:** Most rumor identification moves toward just foresee the veracity of rumor, and next to no data are uncovered why it is a bogus rumor. Finding the confirmations supporting the

forecast and introducing them to clients would be exceptionally gainful since it causes clients to expose bits of rumor without anyone else. Making the outcome informative has pulled in inquire about in different regions, for example, illustrative proposal, yet it is as yet another point in the rumor identification field. This may get more in-depth as more models are utilizing profound learning procedures these days. Be that as it may, as AI methods are utilized in more applications, the requests for result clarification from clients are additionally expanding.

- **Multi-task Learning:** Studies as of now show that together learning of stance identification and rumor discovery improves the exhibition of rumor identification [20, 21]. In the rumor identification work process, contingent upon the calculations, the accompanying errands may be included: client believability assessment, source validity assessment, information extraction, and so on. On the off chance that there are fitting datasets with explanations for these information types, one research heading is to investigate perform multiple tasks learning for these errands, notwithstanding the stance identification and rumor identification undertakings. We expect it will profit from the rumor identification forecast task.

- **Rumor Early Detection:** rumor early identification is to distinguish rumor at its beginning time before it wide-spreads via social networking media with the goal that one can take fitting activities prior. Early detection is particularly significant for a constant framework since the more rumor spreads, the more harm it causes, and more probable for individuals to confide in it. This is an extremely testing task since at its beginning period rumor has little proliferation data and not very many client reactions. The calculation needs to depend on substance and outer information, for example, KB. A few examinations have tried their calculations on the beginning period of bits of gossip [9, 22] investigated highlight strength after some time and detailed that client and etymological highlights are better than organized and proliferation highlights for deciding the veracity of rumor at its beginning time. Although there are as of now a few investigations toward this path, more research endeavors are as yet required, because of their significance in the genuine frameworks.

VI. Conclusion

This paper introduced the essential ideas of talk location. However much web-based media has turned into an important hotspot for sharing ongoing and vital data, it is additionally a reproducing stage for reports. Opportune gossip recognition is fundamental to forestall freeze and keep up with harmony in the public eye. This paper clarifies the talk recognition cycle and audits the examination completed for gossip location utilizing different ML methods. The extent of this audit is restricted to a solitary level arrangement task where we foresee if given online data is talk. This undertaking can be stretched out to a staggered, fine-grain grouping where bits of hearsay can be identified for being falsehood or disinformation, deceptions, and so forth Different novel and cross breed AI strategies, for example, fluffy, Neuro-fluffy can likewise be utilized for recognizing bits of hearsay.

References:-

[1.] Knapp R (1944) A psychology of rumor. Public opinion. Quarterly 1:22–37

[2.] Serrano, Emilio, Carlos A. Iglesias, and Mercedes Garijo. "A survey of Twitter rumor spreading simulations." In Computational Collective Intelligence, pp. 113-122. Springer, Cham, 2015.

[3.] Zubiaga, Arkaitz, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. "Detection and resolution of rumours in social media: A survey." ACM Computing Surveys (CSUR) 51, no. 2 (2018): 1-36.

[4.] Castillo, Carlos, Marcelo Mendoza, and Barbara Poblete. "Information credibility on Twitter." In Proceedings of the 20th international conference on World wide web, pp. 675-684. 2011.

[5.] Yang, Fan, Yang Liu, Xiaohui Yu, and Min Yang. "Automatic detection of rumor on Sina Weibo." In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, pp. 1-7. 2012.

[6.] Qazvinian, Vahed, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. "Rumor has it: Identifying misinformation in microblogs." In Proceedings of the conference on empirical methods in natural language processing, pp. 1589-1599. Association for Computational Linguistics, 2011.

[7.] Hamidian, Sardar, and Mona T. Diab. "Rumor detection and classification for Twitter data." arXiv preprint arXiv:1912.08926 (2019).

[8.] Sahana, V. P., Alwyn R. Pias, Richa Shastri, and Shweta Mandloi. "Automatic detection of rumoured tweets and finding its origin." In 2015 International Conference on Computing and Network Communications (CoCoNet), pp. 607-612. IEEE, 2015.

[9.] Kwon, Sejeong, Meeyoung Cha, and Kyomin Jung. "Rumor detection over varying time windows." PloS one 12, no. 1 (2017).

[10.] Genuer, Robin, Jean-Michel Poggi, and Christine Tuleau-Malot. "Variable selection using random forests." Pattern recognition letters 31, no. 14 (2010): 2225-2236.

[11.] Takahashi, Tetsuro, and Nobuyuki Igata. "Rumor detection on Twitter." In The 6th International Conference on Soft Computing and Intelligent Systems, and The 13th International Symposium on Advanced Intelligence Systems, pp. 452-457. IEEE, 2012.

[12.] Jain, Suchita, Vanya Sharma, and Rishabh Kaushal. "Towards automated real-time detection of misinformation on Twitter." In 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2015-2020. IEEE, 2016.

[13.] Chang, Cheng, Yihong Zhang, Claudia Szabo, and Quan Z. Sheng. "Extreme user and political rumor detection on Twitter." In International Conference on Advanced Data Mining and Applications, pp. 751-763. Springer, Cham, 2016.

[14.] Ruchansky, Natali, Sungyong Seo, and Yan Liu. "CSI: A hybrid deep model for fake news detection." In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797-806. 2017.

[15.] Ma, Jing, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. "Detecting rumors from microblogs with recurrent neural networks." (2016): 3818.

[16.] Chen, Tong, Xue Li, Hongzhi Yin, and Jun Zhang. "Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection." In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 40-52. Springer, Cham, 2018.

[17.] Yu, Feng, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. "A convolutional approach for misinformation identification." (2017).

[18.] Nguyen, Tu Ngoc, Cheng Li, and Claudia Niederée. "On early-stage debunking rumors on Twitter: Leveraging the wisdom of weak learners." In International Conference on Social Informatics, pp. 141-158. Springer, Cham, 2017.

[19.] Hassan, Naeemul, Fatma Arslan, Chengkai Li, and Mark Tremayne. "Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster." In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1803-1812. 2017.

[20.] Gorrell, Genevieve, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. "SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours." In Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 845-854. 2019.

[21.] Ma, Jing, Wei Gao, and Kam-Fai Wong. "Detect rumor and stance jointly by neural multi-task learning." In Companion Proceedings of The Web Conference 2018, pp. 585-593. 2018.

[22.] Liu, Yang, and Yi-Fang Brook Wu. "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. 2018.