South Asian Journal of Research in Engineering Science and Technology
Volume 3, Issue **12**, **December 2018**

www.sajrest.com
SAJREST©2018 e-ISSN: 2455-9261
pp. **29–35**

REAL TIME DETECTION OF MISINFORMATION THROUGH TEXT ANALYSIS IN A SOCIAL MEDIA ENVIRONMENT.

E.V.R.M Kalaimani[1], Indhumathi[2], Iyer Janani Sethuraman[3], Jayamary S[4],
Ph.D-Head of Department, UG Scholar,
Computer Science and Engineering,
Arasu Engineering College,
Kumbakonam, Tamil Nadu.
hodcse@aec.org.in

Abstract— The spread of malicious or accidental misinformation in social media especially in time-sensitive situations such as real-world emergencies can have harmful effects on individuals and society. To address this issue, a novel classification approach is proposed to check whether the message is credible or rumor. The proposed work involves extraction and analysis of live streaming tweets from most popular social media-Twitter. Along with this segregation of tweets takes place as they are from verified news channel or general user. Also calculation and comparison of these tweets by providing them similarity with semantic analysis and sentiment analysis. SVM classifier is used to differentiate the misinformation from the credible message. The ability to track rumor and predict the outcome of the project is recall, precision and accuracy that may have practical application for news consumer, financial markets and emergency services to minimize the impact of false information on social media.

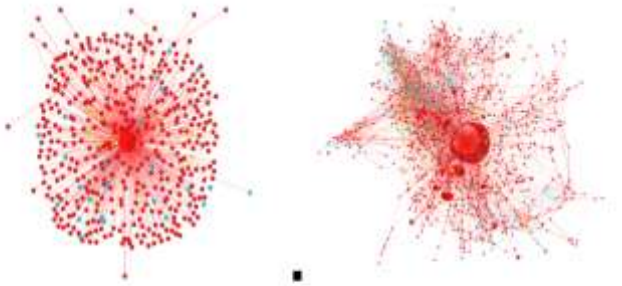
Keywords— Misinformation, Tweets, Semantic and sentiment analysis, SVM.

I. INTRODUCTION

The rise and emergence in popularity of social media and networking services such as Twitter, Facebook and Google+ have greatly affected the news reporting and journalism landscapes. While social media is mostly used for everyday chatter, it is also used to share news and other important information. Due to the rapid development of social media sites and social media networks, user-generated message can reach many audience easily. Such a potential for rapid and far-reaching information propagation in social media brings unprecedented challenges in information quality assurance and management. A recent study revealed that social media activity increases up to 200 times during major events like elections, sports, or natural calamities. This swollen activity contains a lot of information about the events, but is also prone to severe abuse like spam, misinformation, and rumor propagation, and has thus drawn great attention from the computer science research community. Since this stream of information is generated and consumed in real time, and by common users, it is hard to extract useful and actionable content, and alter out unwanted feed. During real world events, researchers have found that Twitter is mainly used actively by many users. Among them, Twitter has evolved into a source of news for many users around the world. It has become a fast, efficient and easily accessible source for news-enthusiasts all over the globe. People are turning to Twitter to seek information about emergency situations and daily events.

A. OVERVIEW OF RUMOR IDENTIFICATION

It dissects how information spread, how it was filled with rumors, and where many of these rumors contained mis-information. The team's exploratory research examines three claims, later demonstrated to be false, that circulated on Twitter in the aftermath of the bombings. The findings revealed within this article suggest corrections to the misinformation emerge but are muted compared with the propagation of the misinformation. Of particular note were the similarities and differences observed in the patterns of the misinformation and corrections contained within the stream over the days that followed the attacks. This article runs on the notion that earlier studies have suggested that crowd sourced information flows can correct misinformation, and this research investigates this proposition, showing that mis-information continues to persist in figure 1.1. This source is useful in analyzing how terrorist attacks are covered on social media as they occur and presents a foundation for researchers to begin thinking about how to address the issue of separating useful intelligence from misinformation. Traditionally television, radio channels, and newspapers were the only news sources available. They are still the top trusted news sources but there is a large new trend toward digital sources. A considerable ratio of newspaper readers now read them digitally and the number of people relying on social media as a news source doubled since 2010. Social media helps you post your news online by a single click, this feasibility leads novel breaking news to show up first on micro blogs. Twitter is one of the most popular micro blogging platforms with more than 250 million users. Accessibility, speed and ease-of- use have made Twitter a valuable platform to read and share information. However, the same features which make Twitter or any micro blogging platform a great resource, but combined with lack of supervision make them fertile grounds for malicious or accidental misinformation in social media. Accordingly, this can lead to harmful incidences especially in sensitive circumstances, which then could cause damaging effects on individuals and society. There are many information seekers who do not rely on a single source to get information, but this is not always a good solution since even other news outlets sometime rely on social media when it comes to novel breaking news. Smart phones enable everyone to capture and tweet every single moment hours before TV cameras arrive. Considering that, social media is an appealing option for those who crave novel tempting news but on the other hand, could deceive anyone by well-structured and formatted rumors. In this study we work on a standard dataset of rumors collected by Qazvinian et al. (Qazvinianet al., 2011). In their work, the definition of rumor is defined as a statement whose truth value is unverifiable or deliberately false. We are using the same definition and not investigating the stimulus behind rumors creation. We investigate the problem of detecting rumors in Twitter data. We start with the motivation behind this research, and then the history of similar studies about rumors is overviewed. Then the overall pipeline is exposed, in which we adopt a supervised machine learning framework, and then we investigate the belief change for president Obama rumors in three years, and finally, we compare our results to the current state of the art performance on the task.



(a) 60 seconds after the hacked twitter account sent out the White House rumor there were already sufficient enquiry tweets (blue nodes).

(b) Two seconds after the first denial from an AP employee and two minutes before the official denial from AP, the rumor had already gone viral.

Figure 1.1 Example for Rumor Spread during bomb explosion.

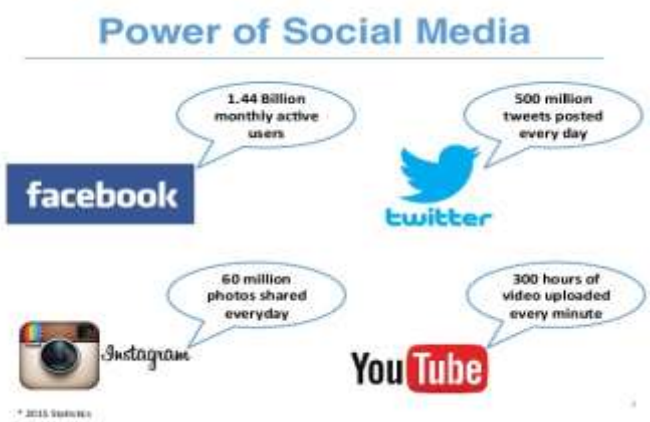


Figure 1.2 Usage of social media in day to day life

II. LITERATURE REVIEW

Many researchers are working to address the problem of credibility of information on Twitter and other platforms in timely manner. Still today many research are taking place in detecting misinformation and through this section we have listed few works that have taken place.

In *Automatic detection and verification of rumors on Twitter* by Vosoughi and Soroush (2015) , the person tried to solve the issue by classifying and clustering assertions made about that event through a speech-act classifier and then evaluating the veracity by examing aspects of information spread : linguistic style used to express rumors , characteristics of people involved in propagating information and network propagation dynamics. In *Deception Detection and Rumor Debunking for Social Media* by Victoria L. Rubin (2017), the person provided a tool for filtering deceptive information.

In *Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts* (2016), where rumors are detected with help of signal tweets, that is, tweets that contain skeptical enquiries: verification questions and corrections. In *How Information Snowballs: Exploring the Role of Exposure in Online Rumor Propagation* by Ahmer Arif et. Al. (2016) made use of mixed method approach in order to demonstrates the importance of rumor content and people content.

Table 1 REVIEW SUMMARY

S NO	Author	Methodology	Points Focussed
1.	Ricardo Buettner	Kuran's theory	Excluded negative and non-conformatory results.
2.	Kate Starbird	Supervised machine learning approach	Explore rumors stance.
3.	Li Zeng	Fleiss kappa statistic	Implies on crisis informatics.

III. METHODOLOGY

A. *PROPOSED ARCHITECTURE*

The proposed architecture is denoted in figure 3.1 that has four main modules as : data collection , user query processing, sentiment analysis, classification and rumor detection.

In this phases Data collection is the first phase where we are able to collect the dataset from particular platform which consist of mixed form as image , video , and some text in form of tweets. In this our work is to collect and annotate a large dataset that includes all the tweets that are written about a rumor in a certain period of time. To overcome the rate limit enforced by Twitter, we collected matching tweets once per hour, and remove any duplicates.

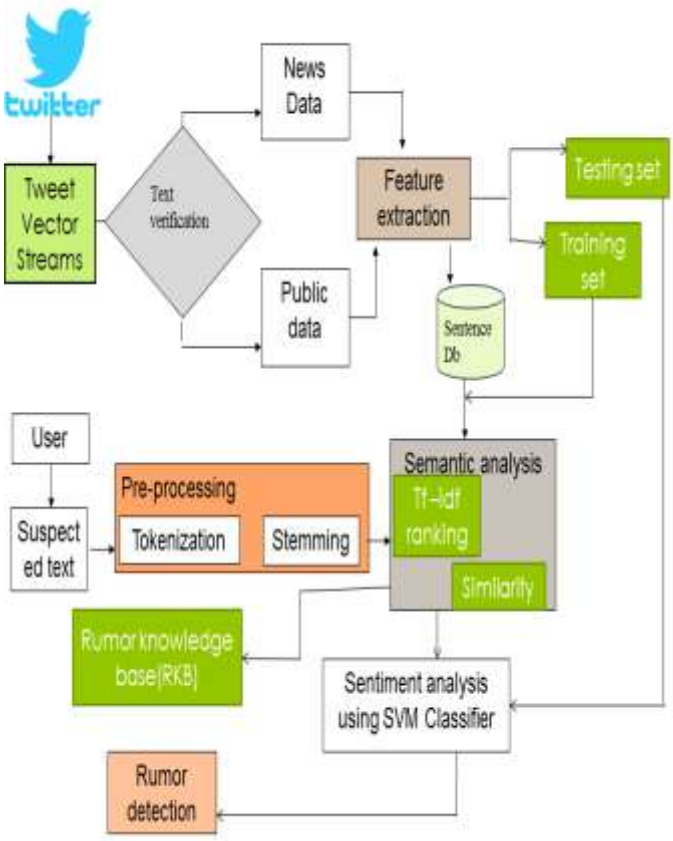


Figure 3.1 Proposed architecture for rumor detection

In phase of user query processing, processing of tweet made by user are made in this phase and also filtering of text from the data store are stored. The text from various category are undergoing process of tokenization , removal of punctuation, stemming. Query processing enables the automated enhancement of user queries. Great queries generally produce satisfying results. Query processing takes the user's query, and depending on the application, the context, and other inputs, builds a better query automatically and submits the enhanced query to the search engine on the user's behalf.

In phase of sentiment analysis, it is a type of data mining that measures the inclination of people's opinions through natural language processing (NLP), computational linguistics and text analysis, which are used to extract and analyze subjective information from the Web - mostly social media and similar sources. The analyzed data quantifies the general public's sentiments or reactions toward certain products, people or ideas and reveal the contextual polarity of the information.

In last phase of rumor detection we are able to provide an accuracy as the message from particular situational tweet is credible or misinformation with help of support vector machine.

A. Semantic and Sentiment Analysis

Semantic analysis is made in order to identify the difference in various words having same identical meaning. For this we are making use of rumor knowledge base(RKB). This knowledge base helps in order to identify the words and they are collected and stored. Consider an example as if the tweet consist of line as "The apple is worst", here the word apple has two meaning as apple may be an fruit or it can be the company. In order to avoid this confusion we make use of knowledge base and with of help of this we detect the sentence is describing which form.

Sentiment analysis is described with help of figure 3.2 which provides an idea as how the process is taking place.

The figure describes how to detect rumor on basis of weight. Here weight describes the number of retweet present about a particular topic. Among the total number of tweets the tweets providing positively are taken count as positive tweets and similar for negative tweets. If the value exceeds as positive more than negative indicates as the message is credible one where as if negative marks more importance as positive then it is misinformation.

a. *Term Frequency and Inverse Document Frequency* (TF×IDF)

Term Frequency (TF) indicates the number of times a keyword or term occurs in document and Inverse Document Frequency (IDF) is through dividing the number of documents in whole collection by the number of documents containing the term.

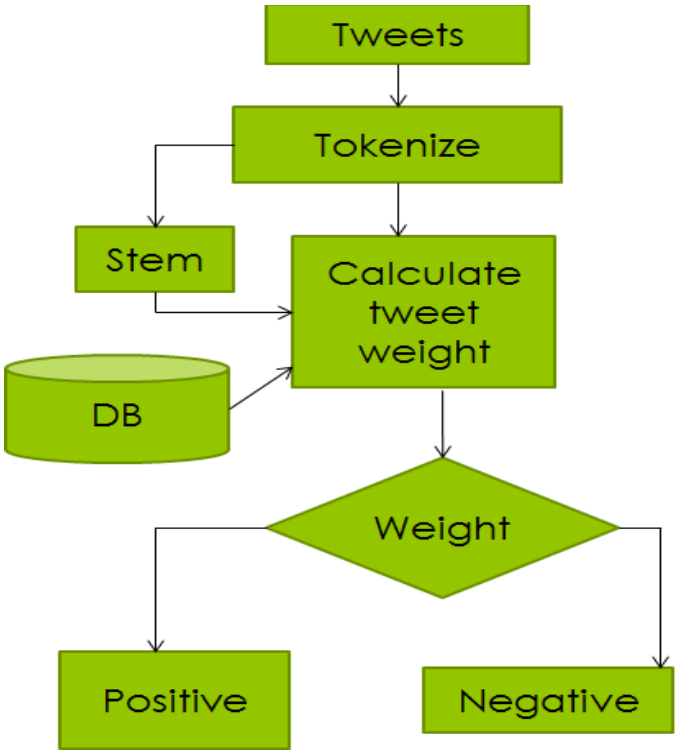


Figure 3.2 Sentiment analysis

$$\boldsymbol{TF}_{p,w_i} = \frac{TF'_{p,w_i}}{\sqrt{\sum_{w_i \epsilon D} (TF'_{p,w_i})^2}} \quad (1)$$

where TF_{p,w_i} denotes the number of times the word w_i appears in the document p, where TF'_{p,w_i} = 1+ ln TF_{p,w_i}

$$\mathbf{IDF}_{w_i} = \frac{IDF'_{w_i}}{\sqrt{\sum_{w_i \epsilon W_{Dx}} (IDF'_{w_i})^2}} \quad (2)$$

where IDF_{w_i} denotes the inverse document frequency of the word $w_{i,}$, where IDF'_{w_i} = ln (1+U/U_{wi}) and N denotes the total number of documents.

With help of TF IDF we calculate relevance score which is as follows

$$\textbf{Relevance Score S} = \sum_{w i \epsilon D} \boldsymbol{TF}_{p,w_i} \times \mathbf{IDF}_{w_i} \quad (3)$$

b. *Similarity*

We make use of cosine similarity in order to measure together with relevance score to provide accurate ranking. A ranking function is used to provide a relevance scores of matching files to a given search request. This is denoted in equation 4, the similarity function is given.

$$\text{Sim}(q,d_u) = \frac{\sum_{j=1}^{U} Sp,j \; .Sq,j}{sqrt(\sum_{j=1}^{U}(Sp,j)\text{^}2)\; Sqrt(\sum_{j=1}^{U}(Sq,j)\text{^}2)} \quad (4)$$

Above equation S_p denotes the relevance score of document p and S_q denotes the relevance score of query.

c. *Support vector machine*

Support vector machine is supervised machine learning algorithm for classification problems that are based on concept of decision planes that defines decision boundaries. It is based on structural risk minimization principle from computational learning theory. It measures the complexity of hypothesis based on margin with which they separate the data and not number of features. It maps the training data into a high-dimensional feature space in which we can construct a separate hyperplane maximizing the margin or distance from hyperplane to nearest training data points. Steps for SVM are denoted in figure 3.3.

Step 1.Normalize the data $A = x - \mu/\sigma$

Where $\mu = \sum x/n$

$\sigma = \sqrt{1/n\text{-}1 \times \sum(xi - \mu)}$

Step 2. Compute Augmented Matrix [A –e]

Step 3. Compute $H = D\,[A\text{-}e]$ and $H^T H$

Step 4. Compute $U = V \times [I - H[I/V + H^T H]^{-1} H^T] \times e$

Step 5.Compute $w = A^T DU$ and $\gamma = -e^T DU$

Step 6. Compute $w^T x - \gamma$

Step 7.Compare the sign $(w^T x - \gamma)$ with Input class label.

Figure 3.3 Steps for SVM

IV. IMPLEMENTATION AND RESULTS

A. Package and Technologies Used

We made use of Support vector machine algorithm in C# programming language with help of other APIs and packages. We have collected dataset from twitter and they are classified on basis as verified tweets or unverified tweets. Verified tweets are mainly from the any places that Twitter has verified and unverified tweets are usually from any user tweets.

B. Results

We collected many news related to misinformation and found few examples that how misinformations are detected with manual as well as automatic way. In manual way it is really very tedious to detect as the message is true or not. We consider an example as “In boathouse of Kerala beef was provided” this was detected as rumor manually by verified tweet of News channel “The Economic Times” in figure 4.1

Figure 4.1 Manual detection of rumor

Our goal is detected here on basis of number of tweets that are provided and also considers more weightage if the person who makes the tweets as well-known personality with good posting. Only their tweets have been taken up as main consideration and we have provided them for various factor process as denoted in architecture as preprocessing, semantic analysis , sentiment analysis and then their classification with help of algorithm. Our main idea concentrates in text so we make use ranking to provide the text their correct order and then with help of this ranking we would further classify them. We have provided graphical representation for providing an accuracy in output which is provided along with precision and recall.

In below figure 4.2 Boston tragedy was because of wrong information passed in social media. The story had begun when one student has become bomber. There was two blast taken place but the someone has provided suspects name unofficially and this news became viral in social media. Without knowing the truth everyone begun to retweet the same data believing as truth.

This resulted to best name as “Boston Bombing” and journalist began to flash the name of suspected bomber as the one responsible for this blast. After few hours this was founded as wrong news and apologies were made that real bomber were found and then their name was flashed which had affected many person.

Our output has been denoted with help of screenshot and graph. We have denoted the preprocessing of data by the user in figure 4.3.In this screenshot we are able to provide any type of text for their split up and then the other process begins.

Our graph is denoted in figure 4.4,4.5 and denotes the relationship between the various components in axis as shown below and their relationship helps to classify an accurate value for detection of rumor or credible message.

Figure 4.2 Boston Tragedy

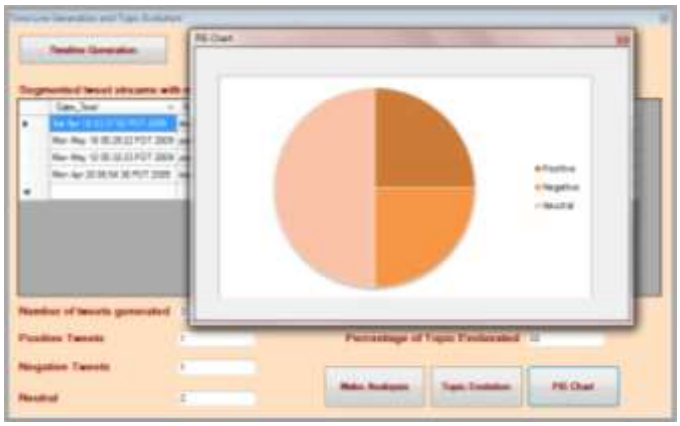


Figure 4.3 User Query Preprocessing and final output for rumor detection in real time with proposed architecture.

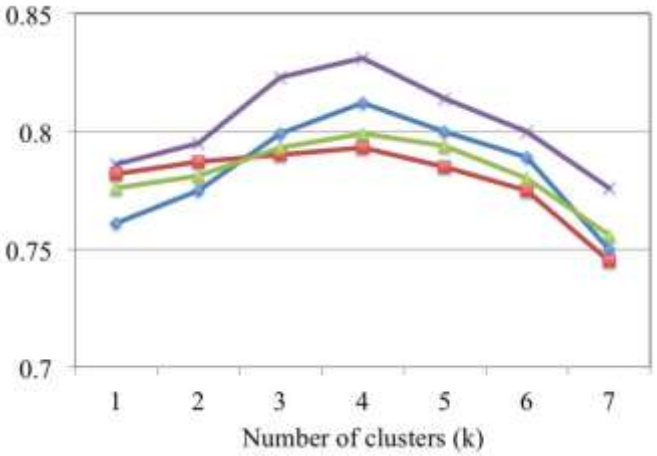


Figure 4.4 Number of dataset in clusters

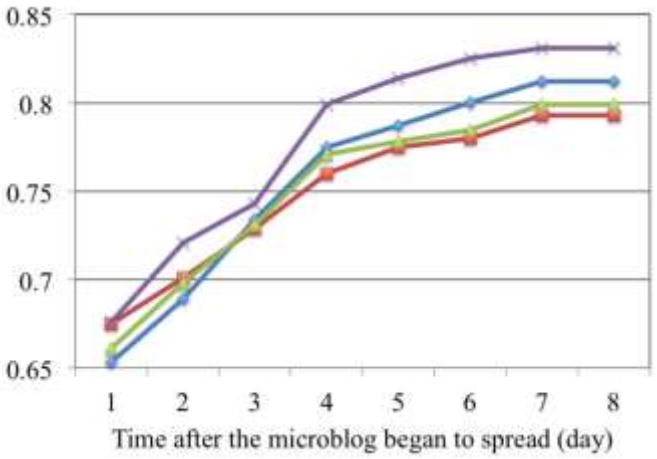
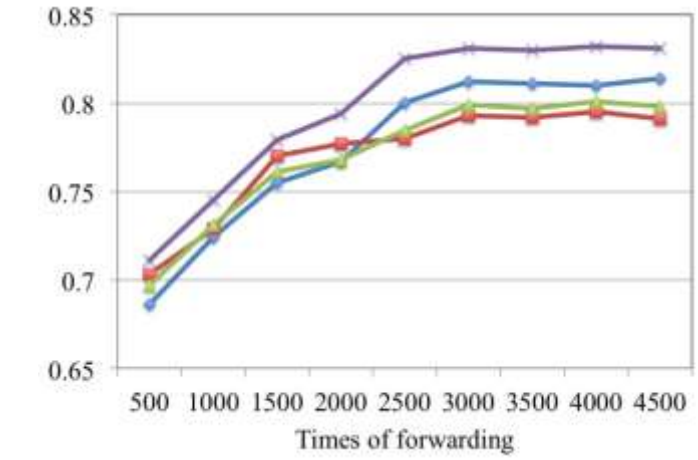
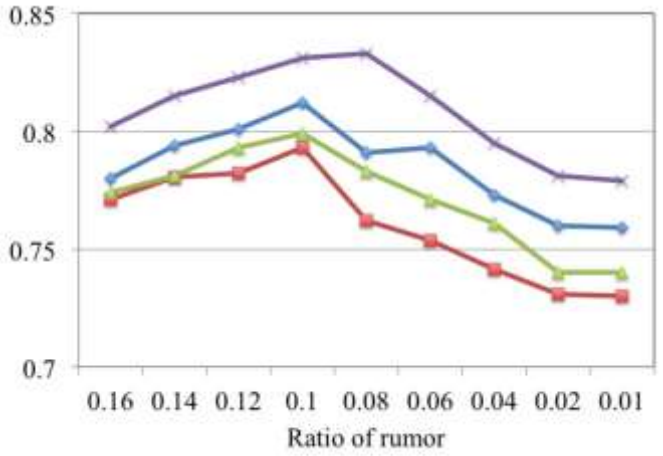


Figure 4.5 Graphs for accuracy , precision , recall





V . CONCLUSION

In this work, detection of rumor can be proved accurately by classifying the tweets with help as verified and unverified and then on basis of some renowned personality. Social media has become the new mechanism for interaction and information flow that requires variety of mechanisms. This work had been made using SVM classification for providing an accurate misinformation from credible message.

VI. FUTURE WORK

There can be many extensions for this research and many different techniques may be used. Along with text nowadays it is important to identify the images also. There may be chances or many probability as the images tweeted or posted in any platform may be true image or fake image. This is the main thing to be recognized and to be solved which is a tedious job.

REFERENCES

[1] Ahmer arif el. At. "How Information Snowballs: Exploring the Role of Exposure in Online Rumor Propagation" ,2016

[2] C. Budak, D. Agrawal, and A. E. Abbadi, "Detecting misinformation in online social networks using cognitive psychology," Human-centric Computing and Information Sciences. 2014, pp. 4-14.

[3] C. Castillo, M. Mendoza, and B. Poblete. "Information credibility on twitter." In Proceedings of the 20th international conference on World wide web, pages 675–684.ACM, 2011.

[4]Dayani, Raveena, et al."Rumor: Detecting Misinformation in Twitter.", 2016

[5] Dewan, Prateek, and Ponnurangam Kumaraguru. "Towards automatic real time identification of malicious posts on Facebook." Privacy,Security and Trust (PST),201513th Annual Conference on. IEEE, 2015.

[6] T. Takahashi and N. Igata, "Rumor detection on twitter," in *Proc. 6th Int. Conf. Soft Comput. Intell. Syst. (SCIS), 13th Int. Symp. Adv. Intell. Syst. (ISIS)*, Kobe, Japan, Nov. 2012,pp. 452–457.

[7] Victoria L.Rubin , Western University"Deception Detection and Rumor Debunking for Social Media" , 2017

[8] Vosoughi, Soroush , Automatic detection and verification of rumor on Twitter. Diss . Massachusetts Institute of technology, 2015.

[9] Zhoa,Zhe,Paul Resnick and Qiaozhu Mei "Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts" ,2015.

[10] Li Zeng, kate starbird, Emma S Spiro."*Rumors at the Speed of Light? Modeling the Rate of Rumor Transmission during Crisis*",2016.