International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

# Classification based Credibility Analysis on Twitter Data

**Parvathi R, Aravind J**

***Abstract:*** ***Twitter is an important place to get access to breaking news and information. So, it is necessary to check the trustworthiness of tweets. Credibility is used to assess the quality of being believable or worthy of trust. Credibility analysis refers to attempt to an ascertain truthfulness in short lie detection. In this work, credibility of the twitter data can be assessed using centrality measures. First the tweets are preprocessed using the preprocessing techniques. The preprocessing techniques on tweets: a stop word removal, stemming, pos tagging etc. are used to improve the performance. Preprocessed twitter data can be used to identify the tweet and author features. Then the centrality measures are applied to the preprocessed dataset. The proposed centrality measures used in this work are Betweeness centrality, Eigenvector centrality, Degree centrality and Closeness centrality. The centrality measures are used to find out the trust between the users and it will be given as input to classifiers. The classifiers like Naïve Bayesian, Support Vector Machine and K Nearest Neighbor are used to classify the tweets based on credibility.***

***Index Terms:*** ***Credibility, centrality, Truthfulness, Classifiers, and Proximity.***

## I. INTRODUCTION

In recent years, Social Networks not only act as a platform for connecting people and sharing thoughts but also provide a way to share and spread information. With the availability of enormous amount of information, People consider them as the primary source of news. Now a days, many social media platforms are available to share opinions, feelings and sentiments. Twitter is most widely used information sharing tool where people can share their opinions, thoughts and feelings via tweets. Twitter users create their profile and share their thoughts, feelings and opinions through tweets via the profile. The maximum length of the tweet can be 140 characters. Each tweet has two main components: the tweet content and user who published the tweet. With the flexibility of anyone can share anything over the platform, Social Media is more prone to the spread of irrelevant and misleading information. Detection of spam or false information on social media is a major research problem. The credibility and quality of information plays a crucial role during high impact events.

Manuscript published on 30 September 2019
* Correspondence Author
**Parvathi R***, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India.
**Aravind J**, School of Computing Science and Engineering, Vellore Institute of Technology, Chennai, India



The organization of the paper is as follows: Introduction is provided in Section I, Literature Survey is mentioned in Section II, Section III gives the Proposed Method, Section IV discusses the Tools and Technologies, Section V gives the Results and Discussion, Section VI gives the Conclusion and Future Work

## II. RELATED WORKS

Castillo et al. [1] examined methods of assessing credibility via microblog postings analysis. He used features extracted from topics of microblog posts to assess the credibility of tweets. Pal and Scott et al. [2] framed an approach which is used to show how name value bias affects the author's judgements. It has been shown by finding the correlation between name values and number of followers. Morris et al. [3] explained how users understand credibility of tweets. He has conducted a survey which showed features used by users to assess credibility. Kang et al. [4] proposed a model to do topic specific credibility on Twitter. He evaluated computational models such as social model, content based model and hybrid model. He used seven-topic specific datasets from Twitter to evaluate computational models. The results showed that the social model outperformed the others in terms of predictive accuracy. Ikegami et al. [5] proposed a model to do credibility analysis of Twitter tweets based on topic and opinion classification. Latent Dirichlet Allocation (LDA) is used to identify the topics. Sentimental Analysis has been done on tweets to find out whether the tweets are positive, negative or neutral. Credibility of tweets can be assessed by computing the ratios of similar opinions to all opinions on a topic. Aditi et.al [6] framed a model to do credibility analysis and ranking of tweets during high impact events. Important contents have been identified from tweets using statistical techniques such as regression analysis to predict the credibility of the information in a tweet. Ratkiewicz et al. [7] computed a trust score for micro blogging updates related to high impact events. In their work, they discussed certain cases of abusive behavior by Twitter users.

## III. PROPOSED ALGORITHM

The architecture of proposed system is given in the Fig.1. The tweets observed during emergency events have the ready to use information from the victims or people near the victims. These data, if properly analyzed, can give tremendous useful information for the emergency responders.

The main challenge in the analysis is to know whether the tweet to be processed is credible or not.
The solution for this problem is proposed using neural networks. Some of the problems of existing methodologies are inability in covering all the aspects of a tweet and minimal analysis of the features. The proposed model works better than the existing models as the features corresponding to the author, tweet and the related community are incorporated in the model. The main advantage of the system lies in the learning stage where it uses neural network where the text is analyzed in many stages, requires less statistical training and the ability to detect complex relationships between variables. The proposed system overcomes the overfitting problem by incorporating a feature selection method to extract the most relevant features. The tweets are extracted from the Twitter API. Three kind of features are going to be constructed from the tweets namely Author, Tweet and Network Features. The important features that are useful for analysis are obtained using the Chi Square Feature Selection method. These features are considered for further feature construction process. The classifiers like Naïve Bayesian, Support Vector Machine and K Nearest Neighbor are used to classify the tweets based on credibility as shown in Figure 1.

## A. Tweets Extraction

One of the reasons twitter is attracted by the research community is the availability of means to extract tweets through Application Programming Interfaces (API). Twitter offers two kinds of API Streaming API and REST API enabling the extraction of streaming tweets and access to read and write twitter data respectively. In this credibility assessment, REST API is utilized for the training purpose and Streaming API for testing purpose.

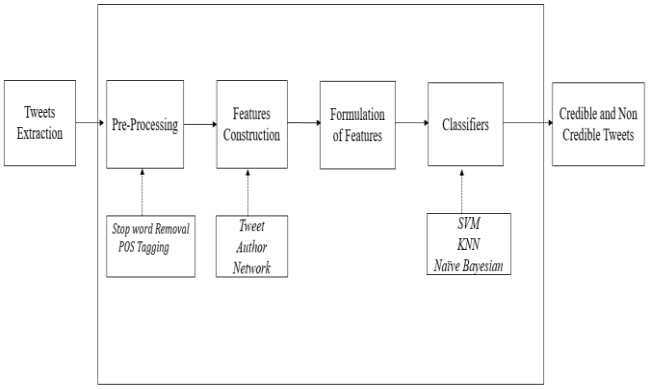



**Fig.1. System Architecture**

## B. Tweets Preprocessing

***Stop word Removal***: Removal of common words that are negligible in the analysis
***POS tagging***: Identify the Parts Of Speech tags for the words in the tweet. The POS tags carry the significance of the word in that sentence.
***Stemming:*** Identify the root of the words in the tweet.

## C. Features Construction

The following features are constructed from tweets.
***Tweet Features*** : Length of the tweet, number of words, number of unique characters, number of hashtags, number of retweets, tweet is a retweet, number of special symbols [$, !], tweet is a reply, Number of @-mentions, number of retweets, has URL, number of URLs, use of URL shortener service.
***Author Features*:** Registration age of the user, number of statuses, number of followers, number of friends, is a verified account, length of description, length of screen name, has URL, ratio of followers to followers.
***Network Features*:** Degree Centrality, Eigenvector Centrality, Betweeness Centrality, Closeness Centrality.

## D. Formulation of Tweet Features

***Vectorization of words***

The words in the tweets are converted to vector using the TFIDF (Term Frequency Inverse Document Frequency) score. The tf-idf weight is the product of its tf weight and its idf weight The score is calculated for all the words in the tweet with the following formula

$$W_{t,d} = (1+\log tf_{t,d}) * \log N / df_t \quad (1)$$

Where $tf_{t,d}$ is no. of occurrences of tweet word t in a tweet d
N is total number of tweets.
$df_t$ is document frequency of tweet word t

The value of TFIDF is

1. highest when term occurs many times within a small number of documents
2. lowest when the term occurs fewer times in a document, or occurs in many documents
3. lowest when the term occurs in virtually all documents.

## E. Formulation of Network Features

***Degree Centrality***

It measures the rank of node having more connections in terms of edges. It measures the total count of directed neighbors. It is classified as in degree and out degree.
In degree: It denotes the total number of incoming edges.
Out degree: It denotes the total number of outgoing edges.

The degree centrality is calculated by

$$\begin{aligned} C_d(v_i) &= d_i^{in} \quad (prestige), \\ C_d(v_i) &= d_i^{out} \quad (gregariousness), \\ C_d(v_i) &= d_i^{in} + d_i^{out}. \end{aligned} \quad (2)$$

***Betweeness Centrality***

Betweenness Centrality refers to how much a node acts like a bridge to all other nodes. It measures the count of shortest paths from one node to another among all the nodes passing through that two corresponding nodes.

The betweeness centrality is calculated by

$$C_b(v_i) = \sum_{s \neq t \neq v_i} \frac{\sigma_{st}(v_i)}{\sigma_{st}}, \quad (3)$$

Where
$\sigma_{st}$ = Total number of shortest path from node s to t.
$\sigma_{st}$ (v)= Number of paths that pass through v.

***Eigenvector centrality***

Eigenvector centrality refers to the how much the node is prestigious in the network. The centrality of the user is considered to be a function of the corresponding neighbors' centralities.

The eigenvector centrality is calculated by

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t \quad (4)$$

Where
M (v) = Set of nodes of all the neighbors of v
λ=constant

***Closeness centrality***

Closeness Centrality refers to how close to others a node is.

The closeness centrality is calculated by

$$C(x) = \frac{N}{\sum_y d(y, x)}. \quad (5)$$

Where d (y, x) is distance between x and y

## F. Feature Selection

Chi Square tests whether the occurrence of a specific feature and the occurrence of a specific class are independent. The aim is to select the features, of which the occurrence is highly dependent on the occurrence of the class. The higher value of the score, the more likelihood the feature is correlated with the class, thus it should be selected for model training.

$$\chi^2(D,t,c) = \sum_{e_t \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_t e_c} - E_{e_t e_c})^2}{E_{e_t e_c}} \quad (6)$$

Where
NN is the observed frequency in and EE the expected frequency
$e_t e_t$ takes the value 1 if the document contains term tt and 0 otherwise
$e_c e_c$ takes the value 1 if the document is in class cc and 0 otherwise

After calculating the Chi Square scores for all the features, the features are ranked and the top ranked features are chosen for training. For each feature (term), a corresponding high χ2χ2 score indicates that the null hypothesis H0of independence (meaning the document class has no influence over the term's frequency) should be rejected and the occurrence of the term and class are dependent.

## G. Classification

The classifiers like Naïve Bayesian, Support Vector Machine and K Nearest Neighbor are used to classify the tweets based on credibility.

***Support Vector Machine***

Linear SVM has been implemented to classify tweets based on credibility. Training data is used to model the SVM classifier and is used for classify the test data. SVM hyperplane is used to segregate the classes as Credible or Non-Credible news in the given training data, i.e., given the labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes the test data as Credible or Non-Credible.
The notation used to define a hyperplane is

$$f(x) = \beta_0 + \beta^T x,$$

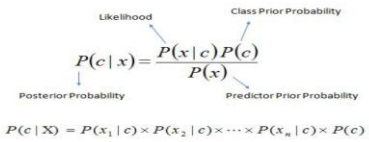
where $\beta$ is known as the *weight vector* and $\beta_0$ as the *bias*. (7)

***K Nearest Neighbor***

KNN can be used for both classification and regression predictive problems. For each data points we take the voting among its nearest neighbors. K denotes the number of neighbors used for voting. Euclidean distance is used to compare and find the nearest neighbors. The data point is assigned to the class which has got more votes from the neighbor nodes. The data will be classified as credible if the data is surrounded by credible data points or else classified as non-credible.

***Naïve Bayesian Classifier***

Naive Bayes is the classification technique used to derive the classification model for extracted data. Naive Bayes by default uses Gaussian distribution internally, as it classifies data based on every input feature (Retweets, Favorites, New_Feature).
The Gaussian distribution is as follows




(8)

# IV. TOOLS AND TECHNOLOGIES

## A. Dataset

In this work, the credibility analysis on twitter data is carried out. The dataset used is the collection of the real time tweets downloaded from twitter by using the tweet ids. 80 % of data is used for as training data to model the classifiers and 20 % of data is used as testing data

## B. Tools and Languages

Python language is used to analyze the twitter data. Python is widely used high level programming language. Preprocessing is done by using NLTK. It is more popular library for natural language processing and it is used to calculate the centrality measures.

## V. RESULTS AND DISCUSSION

The Figure 2 shows the network for the provided dataset with nodes and the corresponding relationships as edges. The below network contains 180 nodes and 182 edges.

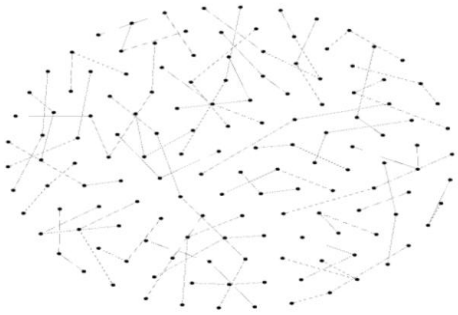


**Fig 2: Network based on the Dataset**

***Centrality Measures***

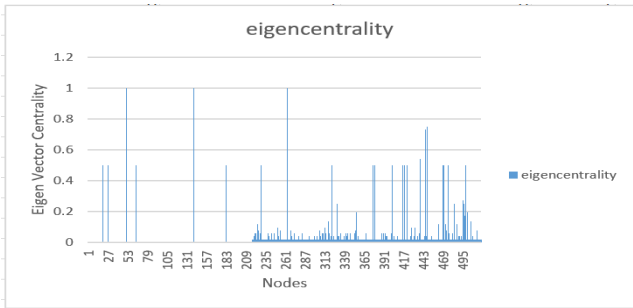



**Fig 3: Eigen Vector Centrality**

From the graph Figure 3, it is inferenced that some nodes have high eigen vector centrality value and betweenness centrality values. It indicates that those nodes are more important in the network to establish the communication between the other nodes.

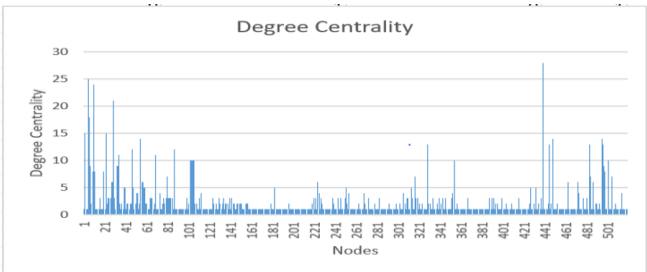



**Fig 4: Degree Centrality**

From the graph Figure 4, it is inferenced that almost all nodes have good degree values. It indicates that almost all nodes are connected to each other in a network.

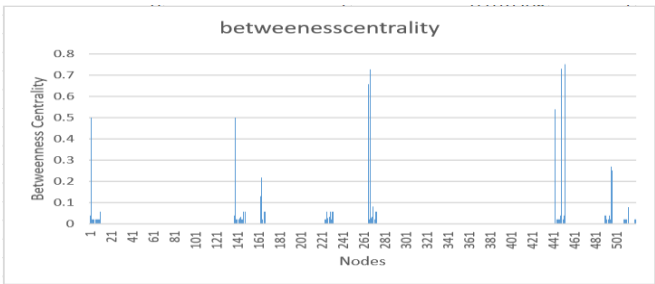



**Fig 5: Betweeness Centrality**

From the graph Figure 5, it is inferenced that some nodes acts as a bridge between all other nodes in order to establish the connection between them. It indicates that those nodes are more important nodes in the network.

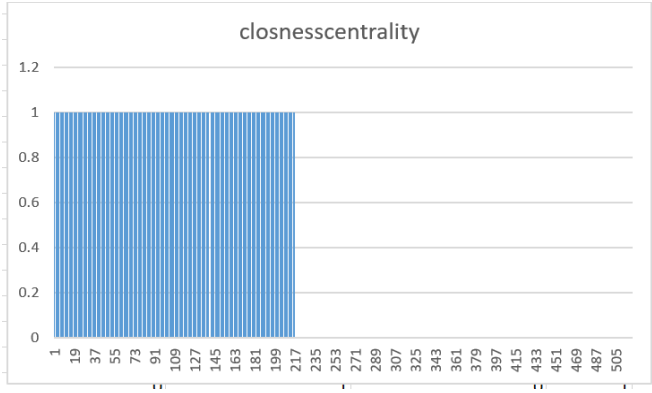



**Fig6: Closeness Centrality**

From the graph Figure 6, it is inferenced that set of nodes are closer to each other and they formed a community.

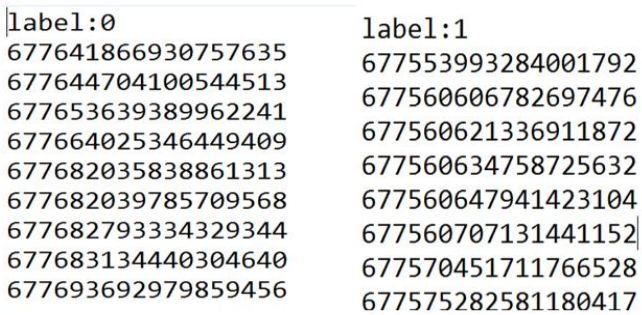



**Fig 7: Classified tweets based on credibility**

As in Figure 7 the tweets id's with label "0" as the credible tweets whereas the tweets with label "1" are not trustworthy or they are the rumors.

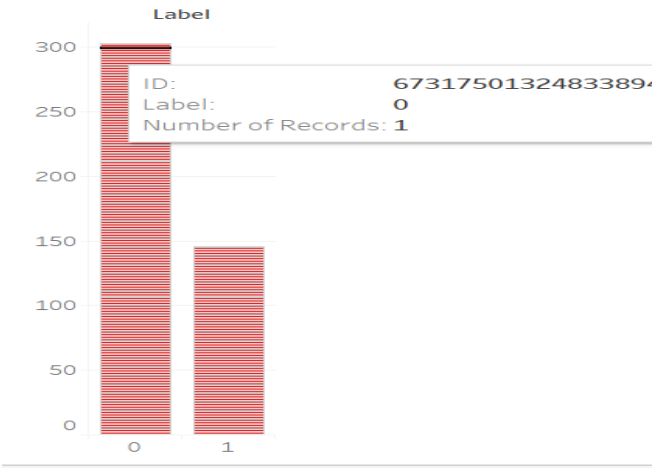



**Fig 8: Classified tweets visualization based on labels.**

As in Figure 8 the tweets ids are visualized based on the labels. The record in fig denotes the corresponding tweet id, label and the number of records associated with it. The visualization is only done for 300 tweets.

***Support Vector Machine***

**Table 1: SVM Classifier Confusion Matrix**

| | Credible | Non Credible |
|---|---|---|
| Credible | 894 | 105 |
| Non Credible | 856 | 6457 |

Table 1 shows the confusion matrix of SVM Classifier with accuracy. Accuracy (SVM Classifier): 88 %

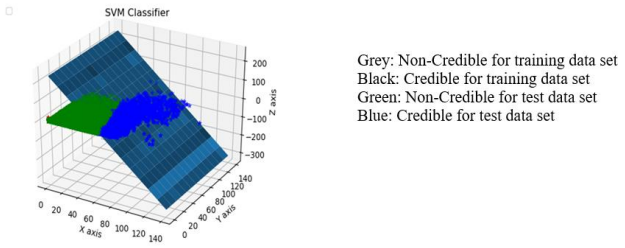



**Fig 9: Classified tweets based on Credibility using SVM**

Figure 9 explains about the classification of tweets based on credibility using SVM Classifier.

***K Nearest Neighbor***

**Table 2: KNN Classifier Confusion Matrix**

| | Credible | Non Credible |
|---|---|---|
| Credible | 894 | 105 |
| Non Credible | 675 | 6638 |

Table 2 shows the confusion matrix of KNN Classifier with accuracy. Accuracy (SVM Classifier): 91 %

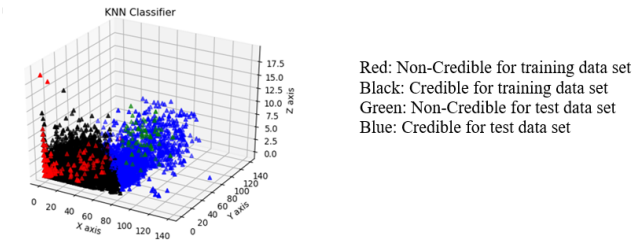



**Fig 10: Classified tweets based on Credibility using KNN**

**Figure 10 explains about the classification of tweets based on credibility using KNN Classifier**

***Naïve Bayesian Classifier***

**Table 3: Naïve Bayesian Classifier Confusion Matrix**

| | Credible | Non Credible |
|---|---|---|
| Credible | 932 | 67 |
| Non Credible | 1589 | 5724 |

Table 3 shows the confusion matrix of Naïve Bayesian Classifier with accuracy. Accuracy (Naïve Bayesian Classifier): 80 %

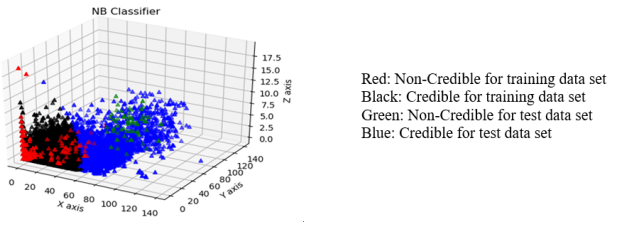



**Fig 11: Classified tweets based on Credibility using Naïve *Bayesian Classifier***

Figure 11 explains about the classification of tweets based on credibility using Naïve Bayesian Classifier and Figure 12 shows the comparison of Accuracy obtained from the SVM,KNN, and Naïve Bayesian Classifiers/

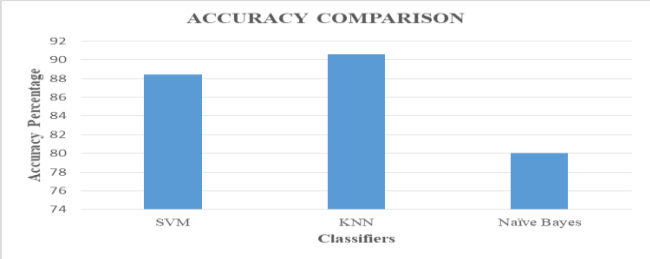



**Figure 12: Comparison of Accuracy of different classifiers**

## VI. CONCLUSION AND FUTURE WORK

In this work, the proposed model assess the trustworthiness of a tweet effectively using the centrality measures that incorporates the properties of the tweets. The proposed model can lessen the misfortune in expectation and increment the precision of the performance. The framework conveys an order on the given tweet. The features of the model are classified based on the centrality measures using graph theory and the credibility of the tweet is classified as Credible or Not Credible. Because of the size and quick advancement of micro blogs, for example, Twitter, it is exceedingly testing to completely comprehend the connections between highlight use and really trustworthy data. A resulting examination will focus on uniting farsighted limit of different features with both allotment what's more, particular usages to get a more through and through learning of the boggling collaboration's that occur in the Twitter space, and to offer comprehension to future legitimacy desire models.

## REFERENCES

1. C. Castillo, M. Mendoza, and B. Poblete, "*Information credibility on twitter*," presented at the Proceedings of the 20th international conference on World Wide Web, Hyderabad, India, 2011.
2. Pal, A. and Counts, *S. What's in a @name? How Name Value Biases Judgment of Microblog Authors*. in Proc. ICWSM, AAAI (2011).
3. M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "*Tweeting is believing? Understanding microblog credibility perceptions*," in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 2012, pp. 441-450.
4. B. Kang, J. O'Donovan, and T. Höllerer, "*Modeling topic specific credibility on twitter*," in Proceedings of the 2012 ACM international conference on Intelligent User Interfaces, 2012, pp. 179-188.
5. Y. Ikegami, K. Kawai, Y. Namihira, and S. Tsuruta, "*Topic and Opinion Classification Based Information Credibility Analysis on Twitter*," in Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, 2013, pp. 4676-4681.
6. Aditi Gupta and P. Kumaraguru, "*Credibility ranking of tweets during high impact events*," in Proceedings of the 1st Workshop on Privacy and Security in Online Social Media, 2012, p. 2
7. J. Ratkiewicz, M. Conover, M. Meiss, B. Gon_calves, S. Patil, A. Flammini, and F. Menczer. Truthy: "*Mapping the spread of AstroTurf in microblog streams*", WWW '11.
8. J. R. Nurse, I. Agrafiotis, M. Goldsmith, S. Creese, and K. Lamberts, "*Two sides of the coin: measuring and communicating the trustworthiness of online information*," Journal of Trust Management, vol. 1, pp. 1-20, 2014.

9. K. K. Kumar and G. Geethakumari, "*Detecting misinformation in online social networks using cognitive psychology*," Human-centric Computing and Information Sciences, vol. 4, pp. 1-22, 2014
10. D. Saez-Trumper, "*Fake Tweet Buster: A Webtool to Identify Users Promoting Fake News on Twitter*," 2014
11. S. Y. Rieh, M. R. Morris, M. J. Metzger, H. Francke, and G. Y. Jeon, "*Credibility Perceptions of Content Contributors and Consumers in Social Media*," 2014.
12. M. Al-Qurishi, R. Aldrees, M. AlRubaian, M. Al-Rakhami, S. M. M. Rahman, and A. Alamri, "*A new model for classifying social media users according to their behaviours*," in Web Applications and Networking (WSWAN), 2015 2nd World Symposium on, 2015, pp. 1-5.
13. A. A. AlMansour, L. Brankovic, and C. S. Iliopoulos, "*Evaluation of credibility assessment for microblogging: models and future directions*," in Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business, 2014, p. 32.
14. J. O'Donovan, B. Kang, G. Meyer, T. Hollerer, and S. Adalii, "*Credibility in context: An analysis of feature distributions in twitter*," in Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom), 2012, pp. 293- 301.
15. K. Lorek, J. Suehiro-Wicinski, M. l Jankowski-Lorek, and A. Gupta, "*Automated Credibility Assessment on Twitter*," Computer Science, vol. 16, pp. 157-168, 2015.
16. K. R. Canini, B. Suh, and P. L. Pirolli, "*Finding Credible Information Sources in Social Networks Based on Content and Social Structure*," in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, 2011, pp. 1-8.
17. Westerman, D., Spence, P.R., and Van Der Heide, B.: "*A social network as information: The effect of system generated reports of connectedness on credibility on Twitter*"', Computers in Human Behaviour, 2012, 28, (1), pp. 199-206.
18. Metaxas, Panagiotis Takas, Samantha Finn, and Eni Mustafaraj. "*Using Twitter Trails. Com to Investigate Rumour Propagation*." Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing. ACM, 2015


## AUTHORS PROFILE

Parvathi R completed her doctoral degree from Anna University, Chennai, India, by contributing his ideas to the field of spatial data mining. She has a teaching experience of over 20 years in the field of computer applications. Her research interests include Data Mining, Recommendation Systems and Social Network Analysis. She has authored articles in big data analytics for renowned publications.

Aravind J is pursuing M.Tech Computer Science and Engineering at Vellore Institute of Technology, Chennai. He completed his B.Tech in Information Technology at Madras Institute of Technology, Anna University. His research interests include Data analytics, Recommendation Systems and Social Mining.