

# A Framework for Finding Outspread News Pattern on Diverse Dataset using Time-Series

Manish Sharma, Bhasker Pant, Vijay Singh

*Abstract: Expressing feeling or opinion is an inherent property of the individual and Now a day's social media becomes an integral part of everyone's life. It is a great medium to analyze the feeling of mass, but sometimes it flows the false feeling in the form of fake news or contents posted on social media. These fake content affects the people in the form of sentiments or companies in the business loss/profit, because most of the people make opinions based on what they read on social media. In fact, fake news or false information can create the damage among the individual, so it should be identified as early as possible. The interest in finding the pattern of fake news has been growing very rapidly in the last few years. In this article we proposed a comprehensive pattern analysis of viral contents, real or fake news on twitter using time series analysis. The proposed technique is simple but effective for detecting and analysis fake contents on the social networks. Experiments results shows that our proposed technique outperformed for differentiating real vs fake news on twitter. Finally, we identify and discuss future direction.*

**Keywords:** Fake news, pattern analysis, time series analysis, sentiment analysis.

## I. INTRODUCTION

Twitter has been accepted as one of the most frequently used micro-blogging service. It is known for providing most up-to-date real-time information, but sometimes user spread contents or stories without knowing their truthfulness. The concept of fake news exists before the emergence of computer or internet technologies. With the increasing penetration of smartphone among the youth moreover in all age groups, social networks become the most powerful tool for communication. With the growing demand of social media, the twitter or the other social networking sites becomes the ideal place for spreading the fake news or fake contents, political sentiments, communal sentiments, rumors etc. First time fake news gains popularity in the year of 2016, US presidential election [1]. Thereafter in every major election throughout the world, social networks used as a tool for polarizing the voters, because voters can be easily influenced by the politician's fake sentiments, promises and claims [2]. Some studies show that dogmatic individuals and religious fundamentalist are more likely believe in fake

news.[3]

### A. Importance of finding fake new pattern

Over the last few years, as the social networks gains exponential growth throughout the world, people try to solve their hidden agendas or unknowingly involved in spreading the fake news or contents on the social network websites. By spreading fake news, it can destroy the faith and belief against the individual or organization, resulting personal or business loss [4].

### B. Contribution and motivation of this article

As "fake news" become the most discussed jargon in the computing world, and most of the major software companies are trying to find the solution of this issue, because false information destroying the real feeling of mass against the people, product organization, political parties etc. Day by day peoples are losing faith in the contents posted on social media. The motivation of this article can be outlined as finding the pattern of non-real news spreading on twitter as well as comparison with real news spreading, which will be addressing for finding fake news as early as possible and improving the existing fake news detection framework. Recently, lots of paper published considering fake news or false information detection as an emerging area of research. In fact, the number of fake news contents on social networks is directly proportional to the preference of the individual [5].

## II. RELATED WORK

**A.** This section focuses on few important related works for finding the fake news and spreading pattern of fake news and real news. This section focuses on few important related works for finding the fake news and spreading pattern of fake news and real news. In the beginning of the fake news analysis process scientist try to identify fake news with help of some information characteristics like velocity, veracity and volume of the flowing information through social media like twitter and Facebook [8]. So, this process helps to detect rumors faster than other sources, like print media. After that to more accurate concept of Machinelearning was introduced. ML is able to detect the text pattern, microblogs, topic through new tools and guidelines [9]. A system was designed to detect astroturfing campaigns in 2010[9] which help to detect reliability of the news [10].

It has been noticed that it is not sufficient, so a model has been developed known as Hoaxy<sup>TM</sup> to crosscheck the source of information [10]. Some other methodology also developed parallelly to explicitly manage the information on social network for provide the uniform information framework across the network [11]. Heuristic function isobgging modeled to face this problem by using an algorithm which study specific URL on the network of news sites [12]. In November 2016 Facebook formed a task force to analyze the fake news [13][14].

**Revised Manuscript Received on January 05, 2020.**

\* Correspondence Author

**Manish Sharma\***, Computer Science and Engineering ,Graphic Era deemed to be university, Dehradun Uttarakhand , India . Email: manishsharma.cse@geu.ac.in

**Bhasker Pant**, Computer Science and Engineering , Graphic Era deemed to be university, Dehradun Uttarakhand , India . Email: pantbhaskar2@gmail.com

**Vijay Singh**, Computer Science and Engineering , Graphic Era deemed to be university, Dehradun Uttarakhand , India .. Email: vijaysingh\_agra@hotmail.com

**B.** In time series analysis a benchmark model is the autoregressive integrated moving average (ARIMA). Time series use statistical property of time series dataset as know Box Jenkins framework[16] for building models. ARIMA also implement so many exponential models for smoothing[17]. Limitation with is it is pre assumed linear model therefore nonlinear can't be processed by the ARIMA.

**III. METHOD**

**A. Data Collection**

The main issue in this area is relevant and quality dataset collection. With the help of diverse news data set from social network twitter, python news APT, RSS feeds of news web application, we collected 10 fakes news tweets and 10 viral real news tweets. Data have collected based on popular hashtags, popular keywords in particular same time frame. Fake news tweets, RSS feeds are mostly found in sort life span in count and retweet, but count is high. But real news tweets and feeds count grows slowly and retain for long time of span. Python news API collect news from globally. Popular hashtags-based tweets have collected from all over the world for spreading pattern analysis.

**B. Time series approach for spreading pattern analysis**

To building an effective spreading pattern analysis system first we analyse characteristic of non-real and real news in terms of Volume, Velocity with reference to time series. A time series collection of serial datapoint plotted on a graph in some predefine and equal distributed time order. Time series use for two purpose, a) understand the force of progressive data like volume and velocity of data generation b) used for forecasting, monitoring and prediction like stock market analysis, sales forecasting, census analysis, workload projection etc.

**C. Modeling for time series analysis**

A time series is defined as in equation 1

$$y_t = f(t) \tag{1}$$

Here,  $y_t$  is the variable at time  $t$ . If the variable at the various time period  $t_1, t_2, t_3, \dots, t_n$ . Then the time series is:  $t_1, t_2, t_3, \dots, t_n$  and  $y_t: y_{t1}, y_{t2}, y_{t3}, \dots, y_{tn}$ .

**D. Time series models**

▪ **Additive Model:** -If  $y_t$  is the time series value at time  $t$ .  $T_t$  is trend value,  $S_t$  seasonal value,  $C_t$  is cyclic value, and  $R_t$  is at time  $t$  random fluctuations of frequency respectively. As stated by to the Additive Model, a time series can be indicate as in equation 2

$$y_t = T_t + S_t + C_t + R_t \tag{2}$$

In This model all four  $T_t, S_t, C_t$  and  $R_t$  variables of the time series act separately.

▪ **Multiplicative Model:** -In this model take it that the differentelement in a time series work proportionately to each other. According to this model as in equation 3

$$y_t = T_t \times S_t \times C_t \times R_t \tag{3}$$

▪ **Mixed Models:** -Different assumptions lead to different combinations of additive and multiplicative models as

$$y_t = T_t + S_t + C_t R_t \tag{4}$$

The time series analysis can also be done using the model as as in equation 4,5,6.

$$y_t = T_t + S_t \times C_t \times R_t \tag{5}$$

$$\text{or } y_t = T_t \times C_t + S_t \times R_t \tag{6}$$

**IV. IMPLIMENTATION**

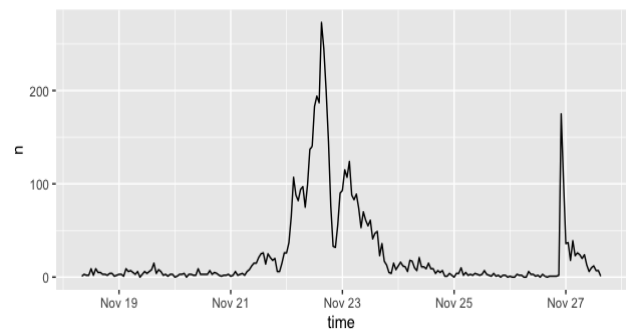
**A. Twitter Data Processing**

- **Twitter Dataset:** - Twitter data set collection of two type dataset one group of data for viral fake news and one for viral real news. We have collected 90 field for any tweet by default.
- **Data Cleaning:** - After filtering the tweets we reduce the attribute in two only one is time attribute name as created\_at, and second is text. Remove all empty text field from all dataset.
- **Dataset frequency count:** - In second phase convert raw data into time series data. In this dataset, we calculate number of tweets in per day and hours. We break it into two type dataset a) tweets frequency per day b) tweets frequency per hours for fake news and same process perform for the all viral real news. Finally, we have time series data to perform analysis.

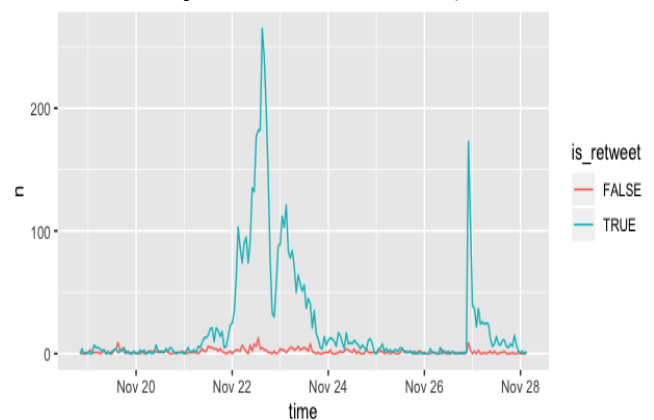
**B. Time series exploratory analysis for real and fake viral news**

Time series exploratory analysis is the process of analysis the Relationship between time and count of tweets and extract meaning full statistics. So that for analyzing first plot the graphs of tweets. There is natural temporal ordering, this make time series special.

**Case Study1:** - Morphed image of soccer stars go viral with 'Pray for Gaza' message. x axis time and y axis count of tweets as shown Figure 1 and Figure 2.

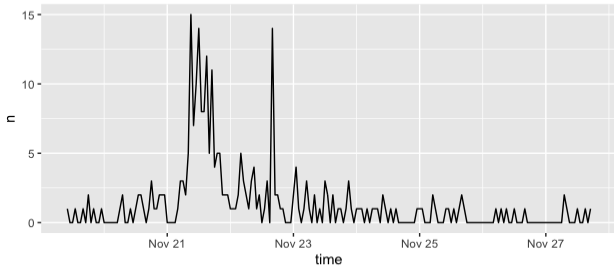


**Fig 1 'Pray for Gaza' message with morphed image . x axis time and y axis count of of tweets, time in hours.**

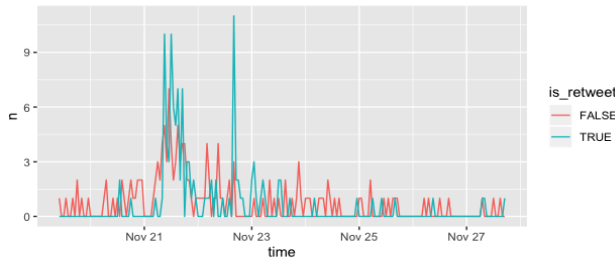


**Fig 2 'Pray for Gaza' message with morphed image . x axis time and y axis count of retweet.**

**Case Study2:** - Fake newsrestaurant getserving licence for Human Flesh in NY USAas shown Figure 3 and Figure 4.

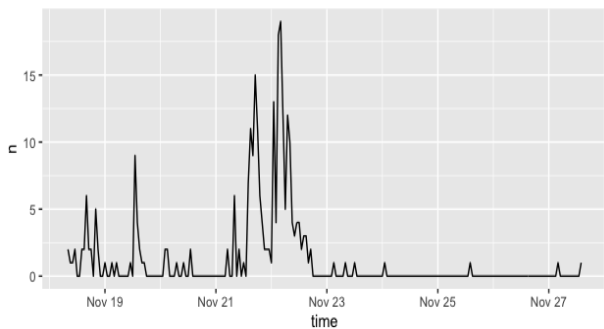


**Fig 3** Fake newsrestaurant getserving licence for Human Flesh in NY USA, tweets.

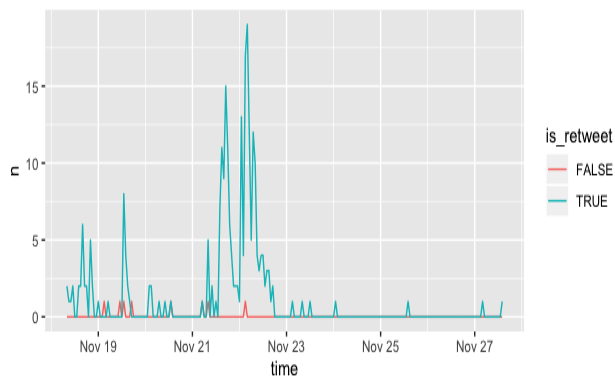


**Fig 4** Fake newsrestaurant getserving licence for Human Flesh in NY USA, tweets, retweets

**Case Study3:** - Original idols of Lord Ram, Lakshman and Sita from Ayodhyaas shown Figure 5 and Figure 6.

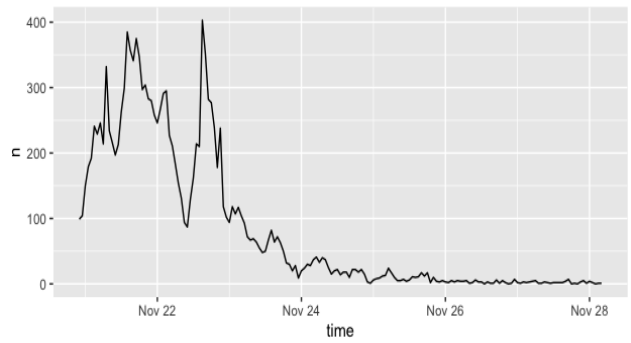


**Fig 5** Fake news on Original idols of Ram, from Ayodhya?

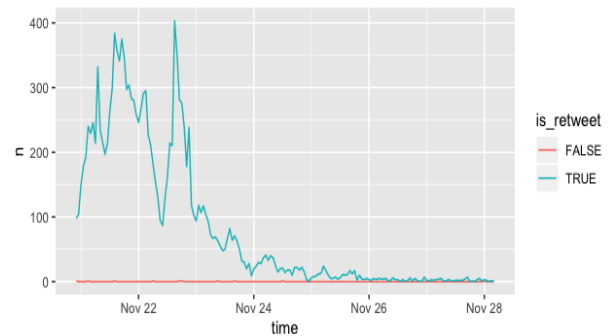


**Fig 6** Fake news on Original idols of Ram, from Ayodhya?

**Case Study4:** - STEM is hard for everyone-grades don't mean you're not good enough to do it/Real Viral as shown Figure 7 and Figure 8.

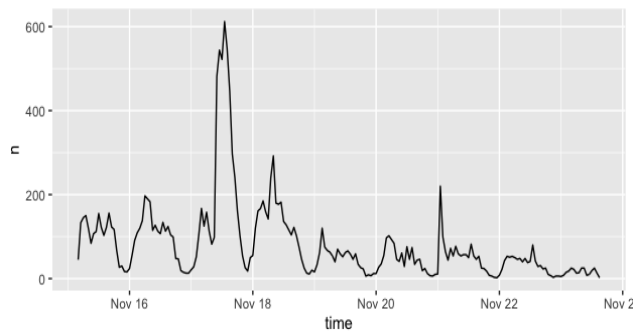


**Fig 7** Google CEO statement Real news STEM is hard for everyone—grades don't mean you're not good enough to do it

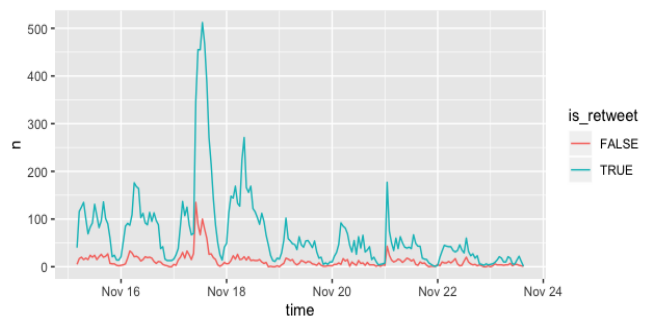


**Fig 8** retweet time series Google CEO statement Real news STEM is hard for everyone—grades don't mean you're not good enough to do it

**Case Study5:** - AyodhyaVerdict: Supreme Court verdict on Ayodhya disputes as shown Figure 9 and Figure 10.

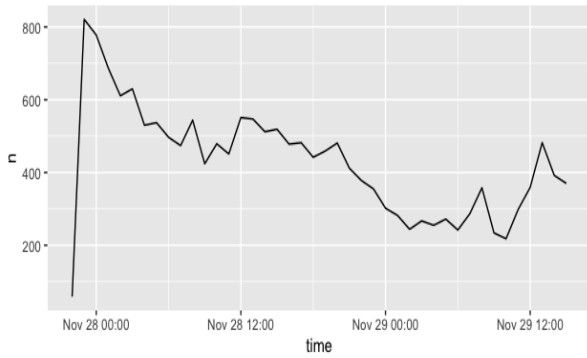


**Fig 9** Tweets on real news Supreme Court Ayodhya verdict.

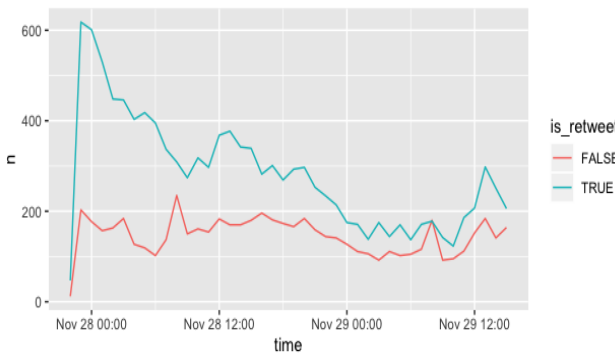


**Fig 10.** reTweets on real news Supreme CourtAyodhya verdict

**Case Study6:** - Tesla CyberTruck launch event reactions as shown Figure 11 and Figure 12.



**Fig 11** tweets on real viral event of Tesla CyberTruck launch event reactions

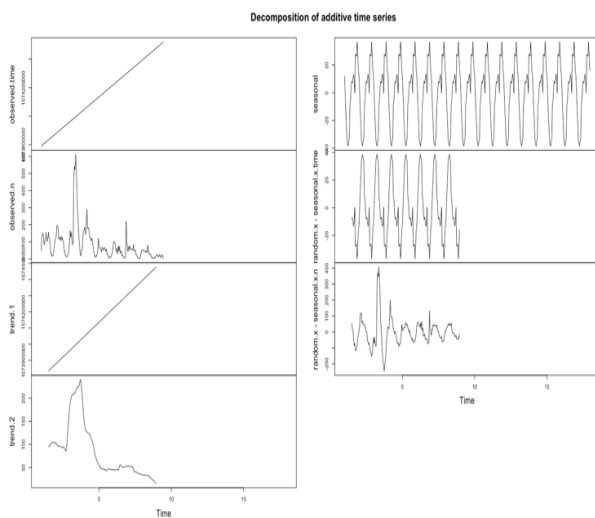


**Fig 12.**retweets on real viral event of Tesla CyberTruck launch event reactions

**V. RESULT ANALYSIS AND FINDING**

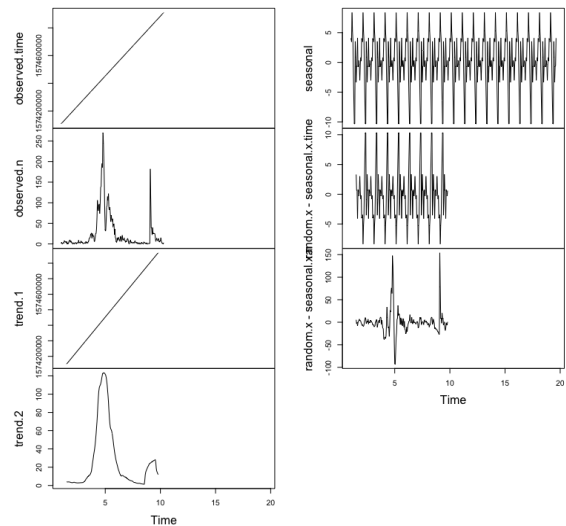
As per case study 1,2,3 are fake news and 4,5, 6 real news. We perform (ARIMA) estimation on time series dataset. Time series data generate four type data patterns: -

- a) The actual data plot (observed),
- b) The overall Movement of data points that is upward or downward (Trend),
- c) Any hourly data points pattern (Seasonal),
- d) Data points unexplainable part (Random).



**Fig. 13.**ARIMA patterns observed, seasonal, trend, random of Ayodhya Verdict finding Viral real news

**Decomposition of additive time series**



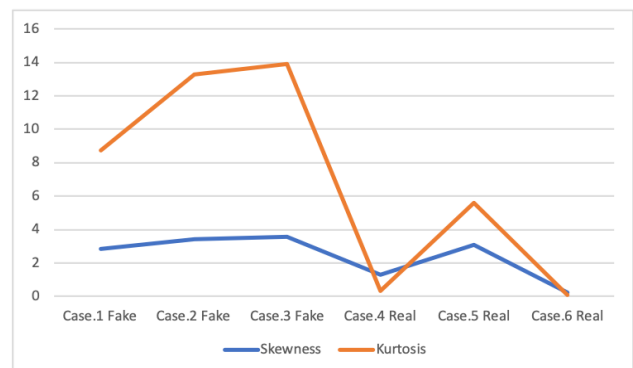
**Fig. 14.** ARIMA patterns observed, seasonal, trend, random of pray for gaza message.

Skewness is calculated for symmetry analysis of values as shown Figure 15. Skewness defines the symmetry of spreading news. And Kurtosis is calculated whether data are normal distribution. It means data with high Kurtosis high value tend to abnormal distribution of pattern and data with low Kurtosis mean data is normal in distribution. As shown in Table 1.

**Table- I: Comparison of real news vs fake news**

Cases	News Type	Skewness	Kurtosis
C1	Fake	2.84	8.74
C2	Fake	3.44	13.29
C3	Fake	3.58	13.89
C4	Real	1.28	.34
C5	Real	5.07	5.60
C6	Real	.25	.10

Observing all four data pattern graphs of ARIMA and all-time series tweets and retweet pattern as shown Figure 13 and Figure 14. There are two types of clear findings from this time series analysis that fake news trend data points pattern is drop quickly but the real viral news graphs velocity sustain or we can say velocity of tweets maintain for long time compare the velocity of fake news.



**Fig 15.** Spreading Pattern Analysis of real news vs. fake



## VI. CONCLUSION

Now a days it is difficult for social media users to identify news reliability because amount of news is increasing day by day. In this study, a framework is developed to detect diverse dataset news, fake news in social media by using Time series pattern. This process has been tested on ten different original twitter data set, five fake news and five real news diverse dataset we collect. We perform timeseries analysis process and calculate different pattern metrics skewness, kurtosis within all dataset. According to the obtained result of skewness and kurtosis of time series news pattern seem that value of both parameter goes down for real news as compare to fake news. In future, this analysis may be improved by exploring time series pattern like retweet pattern of news and extract more feature of dataset.



**Dr. Bhasker Pant**, is an associate professor and Dean Research of Department of Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun. His research interest are Data Mining and Bioinformatics. He works on Data mining algorithms, applications of bioinformatics, research methodology, and more. He may be reached at [pantbhaskar2@gmail.com](mailto:pantbhaskar2@gmail.com).



**Dr. Vijay Singh** is an assistant professor in the Department of Computer Science and Engineering at Graphic Era Deemed to be University, Dehradun. His research interests are Recommender system, sentiment analysis and Data Science. He may be reached at [vijaysingh\\_agra@hotmail.com](mailto:vijaysingh_agra@hotmail.com).

## REFERENCES

1. Horne, B. D., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. ArXiv preprints.
2. Riedel, B., Augenstein, I., Spithourakis, G. P., & Riedel, S. (2017). A simple but tough-to-beat baseline for the fake news challenge stance detection task. ArXiv e-prints.
3. Zhang, G. Peter. "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing* 50 (2003): 159-175.
4. Ahmed, H. (2017). Detecting opinion spam and fake news using n-gram analysis and semantic similarity Ph.D. thesis.
5. Allport, G., & Postman, L. (1965). *The psychology of rumour*. Russell & Russell.
6. Veracity assessment of online data 2019 Decision Support Systems.
7. The current state of fake news challenges and opp 2017 Procedia Computer Sc(13-20)
8. Vosoughi, S., Mohsenvand, M.N., and Roy, D., Rumour Gauge: Predicting the Veracity of Rumours on Twitter. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2017. 11(4): p. 50.
9. Castillo, C., Mendoza, M., and Poblete, B. Information credibility on twitter. in *Proceedings of the 20th international conference on World wide web*. 2011. ACM.
10. Ratkiewicz, J., et al., Detecting and Tracking Political Abuse in Social Media. *ICWSM*, 2011. 11: p. 297-304.
11. Shao, C., et al. Hoaxy: A platform for tracking online misinformation. in *Proceedings of the 25th International Conference Companion on World Wide Web*. 2016. International World Wide Web Conferences Steering Committee.
12. Huberman, B.A., Romero, D.M., and Wu, F., Social networks that matter: Twitter under the microscope. 2008.
13. Gomez Rodriguez, M., Leskovec, J., and Krause, A. Inferring networks of diffusion and influence. in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2010. ACM.
14. BBC, N. How Facebook is starting to tackle fake news in your news feed. 2017; Available from: <http://www.bbc.co.uk/newsbeat/article/38827101/how-facebook-is-starting-to-tackle-fake-news-in-your-news-feed>.
15. Itkowitz, C. Fake news on Facebook is a real problem. These college students came up with a fix in 36 hours. 2016; Available from: [https://www.washingtonpost.com/news/inspired-life/wp/2016/11/18/fake-news-on-facebook-is-a-real-problem-these-college-students-came-up-with-a-fix/?utm\\_term=.f9be631551fd](https://www.washingtonpost.com/news/inspired-life/wp/2016/11/18/fake-news-on-facebook-is-a-real-problem-these-college-students-came-up-with-a-fix/?utm_term=.f9be631551fd).
16. G.E.P. Box, G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA, 1970.
17. E.D. McKenzie, General exponential smoothing and the equivalent ARMA process, *J. Forecasting* 3 (1984) 333-344.

## AUTHORS PROFILE



**Mr. Manish Sharma** is an associate professor in the Department of Computer Science and Engineering at Graphic Era Deemed to be University, Dehradun. His research interests are Recommender System, sentiment analysis and Data mining. He may be reached at [manishsharma.cse@geu.ac.in](mailto:manishsharma.cse@geu.ac.in).

