IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 9, No. 4, December 2020, pp. 726~733
ISSN: 2252-8938, DOI: 10.11591/ijai.v9.i4.pp726-733

A proposed system for opinion mining using machine learning, NLP and classifiers

Poonam Tanwar[1], Priyanka Rai[2]
[1]Department of Computer Science & Engineering, Manav Rachna International Institute of Research & Studies, Faridabad, Haryana, India
[2]Ramdev collage of science and technology, Nagpur, India

Article Info

Article history:

Received May 12, 2020
Revised Oct 14, 2020
Accepted Oct 29, 2020

Keywords:

Classification
Machine learning
Natural language processing
Opinion mining

ABSTRACT

In today's life consumer reviews are the part of everyday life. User read the reviews before purchase, or stores it for finding the best product through comparison of the product review. From customers view point the reviews play vital role to make a decision regarding an online purchase as well as spammers to write the fake reviews which can increase or defame the reputation of any product. Spammers are using these platforms illegally for financial benefits/incentives are involved in writing fake reviews and they are trying to achieve their motive in terms of financial or to defeat the competitor which causes an explosive growth of sentiment/opinion spamming of writing forged/fake reviews. The present studies and research are used to analyse and categorize the opinion spamming into three different detection targets opinion spam, spammers, and to find the collusive opinion spammer groups so that false opinions can be avoided. Opinion spamming further divided into three different types based on textual and linguistic, behavioral, and relational features. The motivation behind this work is to study the dynamics of spam diffusion and extract the latent features that fuel the diffusion process. The user-based features and content-based features have been used for the categorization of spam/non-spam content. The contributions of this work are building the datasetwhich assists as the ground-truth for classifying/analyzing the variation of fraud/genuine and non-spam/spam information diffusion and to analyze the effects of topics over the diffusibility of non-spam and spam evidences/information. The paper, carried out an in-depth analysis of Twitter Spam diffusion.

Corresponding Author:

Poonam Tanwar
Associate Prof. Department of Computer Science & Engineering
Manav Rachna International Institute of Research & Studies, Faridabad
Haryana, India,121002
Email: poonam.tanwar@rediffmail.com

1. INTRODUCTION

Now a days, the World Wide Web has become radical, and it change the way of thinking, communication and share users opinion globally. Opinion is expressed through tweet, comments, reviews, news, discussion forum and social networking sites. In Reviews people write own experience about the product and services related to product and the company on the e-commerce site. Before making any purchase the customer going through all reviews written by the experienced customers and on the bases of that reviews the buyer/customer will take decision about the purchase. These reviews are the basic unit for any purchase to the customer or any business. However, the authenticity of these opinions is very important

and also received attention by the customer or the reviewers or opinions expressed by individuals. The customer's reviews play vital role as they contain the valuable information about the product/merchandise/artifact and the company. So the purchase increasingly depends on the reviews. When most of the reviews are positive then the customer maygoing to buy the product, while in the negative reviews give the negative feedback about the product and change the view of customer toward the product. The dependency between the customers and the reviews paved the spammer for financial gain and write the fake reviews. The sole aim of the spammer to downgrade or defame the reputation of the product or the company. These reviews are named the spam or fraud/fake reviews. In this research opinion spam detection were studied on three perspective [1]:

– Identification of fake online reviews,
– Identification of persons involved in writing fake reviews
– Identifying the network of opinion spammers/false reviewers

E-commerce review websites Yelp.com have all the range of the product and services, and it also containing reviews about product and services [1-4]. It have more than 60 million reviews restaurants, hotels, barbers, mechanics, and other services. By analyzing the big set of data of yelp.com [2-4]. The authors of the website provided a basic categorization of data: fraud/fake reviews category, category of untruthful reviews, and the reviews on particular brand and negativereviews category and the solution to detect them.

Opinion spam

Opinion are the basic unit of any reviews, post, comment, and tweets. Opinion spam contains the irrelevant or spontaneous opinion about the product and services. The motive of our research is opinion spam-detection in efficient way, and filter the effective and accurate opinion -spam and identify the original reviews.

Categorization of opinion spam

Opinion spam is categorized into false or untruthful opinions. The spam is further is divided into four category. Opinion/sentiments can be categories as spam opinion (email spam, web spam, social spam and opinion spam) are shown in Table 1 [1, 5] as explained by Ajay *et.al.*

Online reviews structure

Online opinions is articulated in various forms like comment, post, and status, tweets and reviews. In this research paper, we discussed about the opinions which usually expressed about the product or any company. These reviews are posted on a variety of e-commerce sites. A product review contains the number of information about product, customer and company. These reviews are posted on the e-commerce sites which can be further referred by user can use them as a reference. The product review is fabricated into various sections based on review components shown in Figure 1 defined by Ajay *et.al*. [1] with the product's identity number (PID), the user's unique identity number (UID), the number of helpfulness votes received to review i.e. positive comments, the review rating the 5-star rating scale varying from 1 to 5, 1 for lowest rating and 5 for highest, the review Text, time and date of review, the title of review.

Table 1. Categorization of opinion spam [1]

SPAM				
Email Spam	Web Spam	Opinion Spam		Social Spam
		Deceptive opinion [1, 6]	Disruptive opinion [1, 6]	
		Hyper Spam	Advertisement Spam	
		Defaming Spam	Announcement Spam	
			Random Text Spam	

Figure 1. Reviews components [1]

Problem defination

We are witnessing an unprecedented growth of various social media applications like Facebook, Twitter, and YouTube. They allow the Internet users worldwide to produce, share and consume content. A prominent feature of these sites is the relationship formation among users. These relationships serve as a channel to share the information to others. Twitter has become the main foundation of information. Apart from that providing the valuable information, OSNs also have power of influencing the people's perception. During 2016 US Presidential elections, it is found that Twitter plays a key role in campaigning. Social media also provide the services to health care professionals in engaging with the public, counseling patients, and consulting the colleagues regarding patients. Rising the popularity of OSN It not only attracts legitimate users

but also thespammers. While spammers are the users who misuse the OSNs for the illicit purposes, i.e., driving the attention to unrelated products or services, lure others to click on malicious links. With all the benefits of OSNs, concerns have been raised by the scientists regarding the power of these networks in creating chaos, manipulating believes, opinion and behavior of the people. There is a proliferation of spamming over all online communication mediums i.e., web spam, email spam, and review spam. In web spam, people boost the rank of their web pages by manipulating the page content or through link farming. Nowadays, reviews of online products have become an important part of the buying process for people. Review spam is used for presenting a false image of the products. Due to large number of users at online social networking sites, spammers have extended their targets to OSNs. They can easily access the data from these sites through the provided application programming interfaces (APIs). Twitter defines the spam as bulk or aggressive activity that attempts to drive traffic or attention to unrelated accounts, products or ser-vices. Spam detection systems examine the content as well the behavior of the users to tackle spam problem. 89% of spam accounts have fewer than 10 followers and 17% of spam users exploit hashtags to make their tweets visible in search and trending topics. It is analogous to web spamming where popular and trending search keywords are hijacked for spamming. The activity of posting the duplicate content is recorded by using the near duplicate detection techniques.

In this research we optimized and identified the fake reviews and separate those from the original Data Set, process presented in Figure 2. The users, products, and reviews all are grouped into classes and the classes are user (object), classes (honest and fraud), product (object), classes (good, bad), and finally the review (object), classes (original, fake). Based on that a product is (good, bad) using latent dirichlet allocation (LDA), a topic modelling technique.

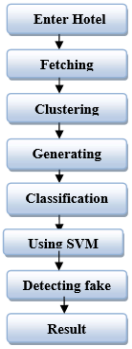

Figure 2. Extraction of fake reviews

Types of opinion-spam detection targets

Opinion spam have number of categories for example spamming in posts, spamming in comments, spamming in status, spamming in blogs, spamming g in reviews etc. so the collection of the reviews and categorization of the reviews is very difficult. Example: suppose we are discussing online about any topic on the discussion forums, and the opinion spammer can divert the discussion towards any other planned direction, and on micro blogging site, the spammers target particular post or blogs, and news websites. Spammer can also play a role to post the spamming comments on any brand and services. Here we will discuss the Deceptive opinion spam in online product or reviews. The opinion spam detection has three goals i.e. deceptive opinion/sentiments spam, deceptive opinion/sentiments spammers, and collusive opinion/sentiments spammers' sets, example of spam twits shown in Figure 2.

Opinion spam detection-

In opinion spam detection the deceptive reviews written by the hoax persons whose intentions are suspicious, we will analysis them. Further partitioned them into low quality review and high-quality reviews. A low-quality review are written by the genuine user do not have the enough experience to write a good review. While high-quality reviews are written by the spammer. Factor through which we analysis the spam review other then quality, the reviews are classified into exact duplicate, near/close duplicate, fractional/partial duplicate, and unique categories.

Opinion spammer's detection

The next task to detect the opinion spammers who redistribution the malicious reviews to mislead the readers. Types of review spammers:

- Identification of Product review spammer
- Identification of Store- review spammer

Identification of product/artifact review spammer

In product review spammer detection the spammer can write the review about the product like car, mobile movie, or restaurant, hotels or any online services. Product review contains the features about the product along with the product star ratings. Detecting spammer is very challenging task. Example suppose a reviewer can writes multiple reviews for the same product or on the same brand, with the star rating of the product and review also contain the repeating text, this kind of reviewers are called spammer.

There are two types of spammer based on their behaviors,

- Target-based
- Deviation based

The authors keep count on every reviewer by combining their review's pattern and then divide the reviews into spammer and non spammer reviews. Spammer use different id or name to spread their reviews in the network.

Identification of store review spammer

In store review spammer the spammer intend to eulogize or to criticize the target object, the reputation of any object. This they of spammer review the information provided in review was false. This type of reviewer target the whole store/company rather than the products of the store/company like Bigbazaar.com, sabhyata.com etc. The reliability of reviews and the reviewers are given below:

- The reliability of review's are depends on the store: it contain the information about the store. And the reliability of reviews is identify by the store reputation, means the store is really doing well or mislead the customer. And the number of positive reviews about the store from the other reviewer.
- The reliability of reviews are depends on the reviewer: A reviewer's credibility depends on the type of reviews, he/she has posted and the connection between reviewer and product via reviews [1, 6].

The effectiveness of process/algorithm depends on the following three sets of process features [6].

- Linguistic textual features
- Behavioral features,
- Relational features.

Linguistic and textual features

Opinion spammerhas their different writing pattern then the non spammer and they copy that pattern into writing the different reviews. Linguistic and textual features convert the reviewers' text into the vector form.

Subsections of the textual features are:

- N-gram features
- POS tags and LIWC features
- Semantic and stylistic features
- Behavioral features
- Relational features

Litrature Review

Benevenuto *et.al*. in 2008 applied machine learning (ML) technique to recognize video spammers on YouTube by using both user-related and video-related features [7-9]. Later, they presented a ML oriented system to identified spammers (Twitter) in 2010 [10]. Whereas Kuak *et.al*. worked on Spam detection incorporated fixed thresholds [6] without ML methods. They used tweet context based features and account-based features to differentiate spammers from standard users. Zhou *et.al*. proposed a Bayesian classifier based approach to detect spammers on Twitter [11]. The popularity of (OSN) online social networks is growing rapidly among all age group globally not only for professional and healthy networking, research interest but also to carry our certain criminal and malicious activities. Traditional spammers in email have transferred to OSNs due to its huge user base and easy-suspicious [2, 4, 6, 12], as initial information are available in public domain to perform certain homework by spammer and to target an individual with more conviction. In order to minimize and tackle the cybercrime and spamming activities efficiently, researchers have proposed a number of significant methods/works in a short time period. More recent works are focusing on the early detection of spam so as to quickly mitigate threats and to prevent any suspicious activity and cybercrime. Text based detection techniques are very capable as they only extract information from tweets to

process further [13-14]. In this proposed work we find an optimized method to identified the fake reviews and separate them from the original Data Set. The users, products, and reviews all are grouped into classes and the classes are user [15-18] (object), classes (honest and fraud), product (object), classes (good, bad), and finally the review (object), classes (original, fake). Based on that a product is (good, bad) using latent dirichlet allocation (LDA) [19-21], a topic modelling technique [22-23].

2. PROPOSED SYSTEM/ARCHTECTURE

Implementation steps are as follow:

Step 1: Reviews need to be extracted followed by text pre-processing, shown in Figure 3, to filter the reviews like to remove stop word, tag, double codes etc, to extract the exact These reviews are suspicious or fake reviews [6].

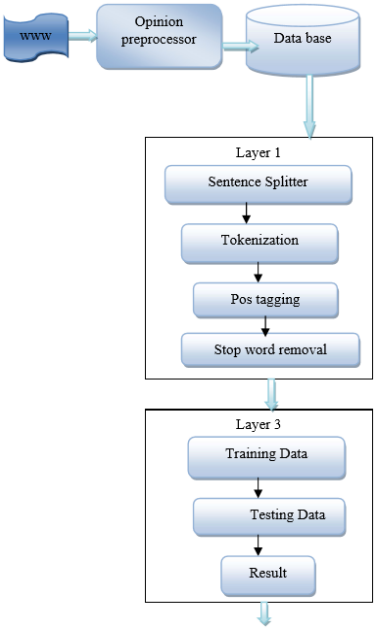

Figure 3. Architecture of proposed opinion analysis system

Step 2: The clean reviews dataset are processed further to classify the data.

- Preprocessed training dataset is now passed to the classifier and a trained model is created. Trained model is then tested and validated by using classifier.
- The available data set is split into training and testing sets in the ratio of 60:40.
- On the training set, the labeled data of varying sizes (from 50 to 2000) was created, the remaining data, we removed the labels and considered it to be the unlabeled data set, apply tweet pooling techniquesand LDA. Then trained the classifier individually on these sets of labeled data and tested it on the test set noting the accuracy [22-23].

Step 3: Will work on the different frequencies like unigram, bigram and the review length for our data model and frequency of attributes.

Step 4: Hybrid of naïve byes classifier (NBC) and support vector machines (SVM) will be used to training data. Spam detection analysis is performed using support vector machine (SVM) classifier, and Naïve Bayes classifier is used. To validate classifier efficiency 10-fold validation process is used.

Step 5: When the NBC and SVM are used for unigram frequency and bigram frequency and for the review length, and they are used to generate the accuracy of the reviews, and detect them whether they are fake or not.

Step 6: The trained naïve byes classifier (NBC) and support vector machines (SVM) classifiers gives us the accuracy of the tested data [24-25].

The main motivation behind this work is to study the dynamics of spam diffusion and extract the latent features that fuel the diffusion process. So far, user-based features (i.e., number of followers, number

of followees) and content-based features (i.e., number of links, number of hash tags) have been used for the categorization of spam/non-spam content. Spammers tend to have less number of followers and more number of followees. Butthere is a variation in the behavior of social media users, some follows back their followers while others not. So, by following a large number of users there is a probability of increasing the number of followers. It means, users based features do not always work well, users can get more number of followers by following those users who likely follow back their followers. This research is helpful in combating the diffusion of spam information and protecting the society from getting panic. It also prevents the people from getting misleading information regarding some services or products. Organization can rely more on the feed-back regarding their products and maintain a good relation with customers, by taking a timely action.

3. RESULT ANALYSIS

This research is helpful in combating the diffusion of spam information and protecting the society from getting panic. It also prevents the people from getting misleading information regarding some services or products. To understand the diffusion dynamics of spam/non-spam information. Collected 10,000 tweets from Twitter. Twitter data fetched on the basis of REST API, after fetching the labeling of data based on the Hspam14 and categorized them as spam and ham tweet. Topic modeling Techniques is applied on the data. The goal of thisstep is to extract the informative and non-redundant features from the preprocessed corpus. It facilitates the machine learning models in generating a statement of prediction. LDA technique has been applied to extract the topics from the tweet. Tweet pooling techniques is applied to overcome the character restriction in twitter. Twitter can have the functionality of "re tweeting" that has been used as a measure of diffusibility and divide the collection of spam and non-spam tweets into 9 categories. Table 2 shows the result of classifiers (SVM and Naïve Bayes).

Table 2. Classifier accuracy

Classifier	Naïve Bayes	SVM
Correctly classify instance	738	1069
Incorrectly classify instance	238	558
Accuracy of correctly classify instance	75%	65.70%
Accuracy of incorrectly classify instance	24%	34.29%

Figure 4 represents that spammers target the social media users by using terms like follow, free, #teamfollowback, #instantfollowback, #followngain, win, #500aday, #follow2be-followed, etc. These terms are claimed to provide certain benefits to the users i.e., 500 bucks a day (#500aday), some free service (free), more followers (#follow2be-followed), etc. As people click these hashtags to get the claimed benefits, they get re-directed to some unrelated accounts or products. In order to prevent people from facing these spammy information, these terms need to be addressed properly during designing the rules for information diffusion at social media platform i.e., Twitter or Facebook. Figure 5 shows that non-spam tweets are free from such terms that claim to provide free services or more followers.

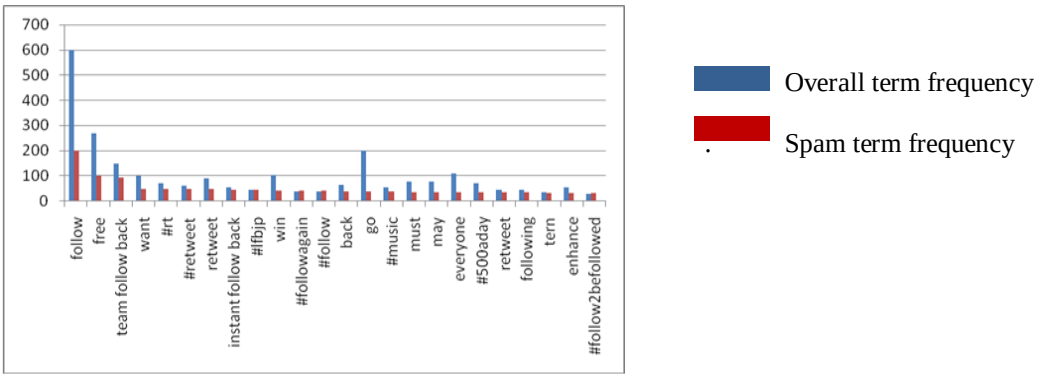

Figure 4. Most relevant term of spam tweet

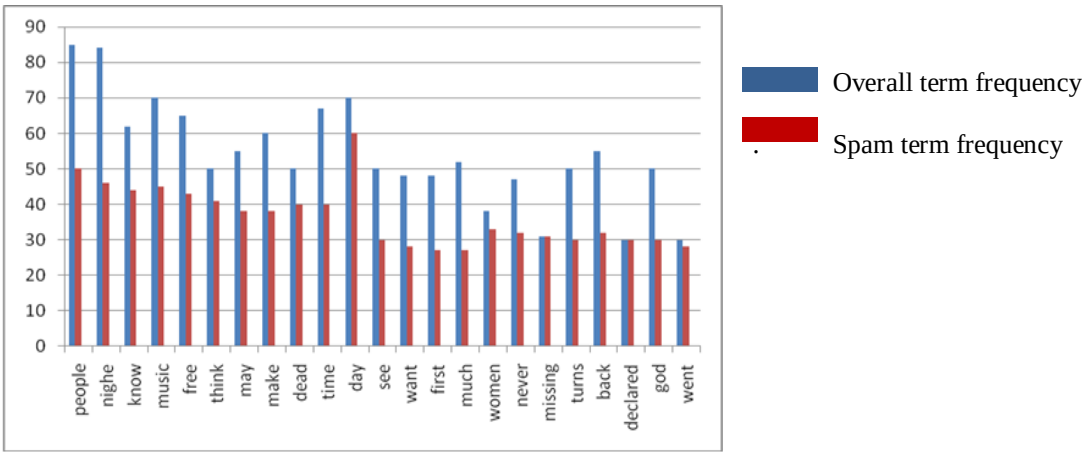

Figure 5. Most relevant term of ham tweet

4. CONCLUSION

Because of limited number of resources/features available fake reviews detection is a major challenge in today life and even all existing approaches/methods suffers in terms of accuracy. In some cases opinion spam detection is based on both linguistic and behavioral features whereas in many cases linguistic and behavioral features are act as an independent entity. The performance of the existing deceptive opinion spam system can be improved by adding the set of optimal features and the hybrid features in our model for training and for the better performance and prediction. In this research work we analysis the pattern of how spam information get diffused in social media, and spamming strategies, and also extract the latent topic from the corpus. However, the research is never going to be ended. There is lots of research area and future works can be considered. Twitter deals with more than 600 million tweets every day. It is very difficult task to identify the spam and ham, Thus LDA approach to address “Spam Diffusion" problem has been applied. The LDA component learns from the detected tweets. When LDA techniques are applied directly on the messages posted at the microblogging sites and return those topics that are hardly informative and tough to interpret.

REFERENCES

[1] Ajay Rastogi and Monica Malhotra, Article in Journal of Information & Knowledge Management, September 2017.
[2] Nitin Jindal and Bing Liu., "Opinion Spam and Analysis" by ACM-2008.
[3] Lim, E., Nguyen, V., Jindal, N., Liu, B. Lauw, H. 2010. Detecting product review spammers using rating behavior. CIKM.
[4] Mukherjee, A., Liu, B. and Glance, N. 2012. Spotting fake reviewer groups in consumer reviews. WWW.
[5] Poonam *et.al*,” A Tour towards the Various Knowledge Representation Techniques for Cognitive Hybrid Sentence Modeling and Analyzer, *International Journal of Informatics and Communication Technology (IJ-ICT)*, Vol.7, No.3, pp. 124-134, December 2018.
[6] Tanwar Poonam, Priyanka, “Spam Diffusion in Social Networking Media using Latent Dirichlet Allocation”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Vol.8, No.12, October, 2019.
[7] Al-Najada, H and X Zhu (2014). iSRD: Spam review detection with imbalanced data distributions. *In Proceedings of the 15th IEEE International Conference on InformationReuse and Integration (IRI)*, pp. 553-560. New York, IEEE.
[8] Banerjee, S and AYK Chua (2014). Understanding the process of writing fake online reviews. *In Proceedings of the 9th International Conference on Digital Information Management (ICDIM)*, pp. 68-73. New York, IEEE.
[9] S. Ghosh, B. Viswanath, F. Kooti, N. K. Sharma, G. Korlam, F. Benevenuto, N.Ganguly, and K. P. Gummadi. “Understanding and combating link farming in the twitter social network”.
[10] Allahbakhsh, M and A Ignjatovic (2015). An iterative method for calculating robust ratingscores. *IEEE Transactions on Parallel and Distributed Systems*, 26(2), 340-350.
[11] Y. Zhou, Z. Wang, W. Zhou, X. Jiang, Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets, *in: Proc. 19th Annu. Netw. Distrib. Syst. Secur. Symp.*, San Diego, California, USA, 2012. http://www.csd.uoc.gr/~hy558/papers/mal_apps.pdf.
[12] A. Rastogi and M. Mehrotra, September 14, 2017 2:10:00pm WSPC/188-JIKM 1750036 ISSN: 0219-6492.
[13] Chen, YR and HH Chen (2015). Opinion spam detection in web forum: A real case study. *In Proceedings of the 24th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee*, pp. 173-183.
[14] Xie, S., Wang, G., Lin, S., and Yu, P.S. 2012. Review spam detection via temporal pattern discovery. KDD.

[15] Distributional Footprints of Deceptive Product Reviews by Feng, S., Xing, L., Gogar, A., and Choi, Y. 2012aICWSM. Akers, RL (2011).
[16] Akoglu, L, R Chandy and C Faloutsos (2013). Opinion fraud detection in online reviews by network e®ects. *In Proceedings of the 7th AAAI International Conference on Weblogs andSocial Media (ICWSM'13)*, pp. 2-11. Palo Alto, CA: AAAI.
[17] Algur, SP, AP Patil, PS Hiremath and S Shivashan (2010). Conceptual level similaritymeasure based review spam detection. *In Proceedings of the 2010 International Conferenceon Signal and Image Processing (ICSIP)*, pp. 416-423. New York: IEEE.
[18] Banerjee, S, AYK Chua and JJ Kim (2015). Using supervised learning to classify authenticand fake online reviews. *In Proceedings of the 9th International Conference on UbiquitousInformation Management and Communication*, p. 88. New York: ACM.
[19] Chakraborty, M, S Pal, R Pramanik and CR Chowdary (2016). Recent developments in socialspam detection and combating techniques: A survey. *Information Processing and Management*, 52(6), 1053–1073.
[20] Chen, C, K Wu, V Srinivasan and X Zhang (2015). A comprehensive analysis of detection of online paid posters. *In Recommendation and Search in Social Networks*, pp. 101-118. Berlin: Springer.
[21] Chengzhang, J and DK Kang (2015). Detecting spamming stores by analyzing their suspicious behaviors. *In Proceedings of the 2015 17th International Conference on Advanced Communication Technology (ICACT)*, pp. 502–507. New York: IEEE.
[22] H. Achrekar, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu. "Online social networks u trend tracker": A novel sensory approach to predict u trends. In J. Gabriel, J. Schier, S. Hu_el, E. Conchon, C. Correia, A. Fred, and H. Gamboa, editors, Biomedical Engineering Systems and Technologies, volume 357.
[23] Castillo, C., Mendoza, M., & Poblete, B. (2011, March). "Information credibility on twitter". *In Proceedings of the 20th international conference on World Wide Web*, (pp. 675-684). ACM.
[24] Vedanshu Sharma *et.al*.," Live Twitter Sentiment Analysis", *Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020*, May 2020, Available at SSRN: https://ssrn.com/abstract=3609792 or http://dx.doi.org/10.2139/ssrn.3609792.
[25] Tanwar *et.al*., A Tour towards Sentiments Analysis using Data Mining, *International Journal of Engineering Research & Technology*, Volume 05, Issue 12, December 2016.

BIOGRAPHIES OF AUTHORS

Dr. Poonam Tanwar has 17 years of Teaching Experience working as associate prof. in Manav Rachna International Institute of Research & Studies, Faridabad. She has filled 4 patents. She was Guest Editor for Special issue of "Advancement in machine learning (ML) and Knowledge Mining (KM)" for International Journal Of Recent Patents in Engineering (UAE).She has organized various Science & Technology awareness program for rural development.. Beside this she has even published more than 40 research papers in various International Journals and Conferences. She is Technical program committee member for various International Conferences.

Ms. Priyanks Rai, did M.Tech from Manav Rachna International Institute of Research & Studies in 2018. Currently working as Assistant professor Ramdev collage of science and technology, Nagpur