International Journal of Electrical and Computer Engineering (IJECE)
Vol. 9, No. 1, February 2019, pp. 562~569
ISSN: 2088-8708, DOI: 10.11591/ijece.v9i1.pp562-569

Integrated approach to detect spam in social media networks using hybrid features

K. Subba Reddy, E. Srinivasa Reddy
Department of Computer Science and Engineering, Acharya Nagarjuna University, India

Article Info

Article history:

Received Jan 30, 2018
Revised Aug 27, 2018
Accepted Sep 11, 2018

Keywords:

Classification
Malicious links
Online social network
Spammer
Twitter

ABSTRACT

Online social networking sites are becoming more popular amongst Internet users. The Internet users spend some amount of time on popular social networking sites like Facebook, Twitter and LinkedIn etc. Online social networks are considered to be much useful tool to the society used by Internet lovers to communicate and transmit information. These social networking platforms are useful to share information, opinions and ideas, make new friends, and create new friend groups. Social networking sites provide large amount of technical information to the users. This large amount of information in social networking sites attracts cyber criminals to misuse these sites information. These users create their own accounts and spread vulnerable information to the genuine users. This information may be advertising some product, send some malicious links etc to disturb the natural users on social sites. Spammer detection is a major problem now days in social networking sites. Previous spam detection techniques use different set of features to classify spam and non spam users. In this paper we proposed a hybrid approach which uses content based and user based features for identification of spam on Twitter network. In this hybrid approach we used decision tree induction algorithm and Bayesian network algorithm to construct a classification model. We have analysed the proposed technique on twitter dataset. Our analysis shows that our proposed methodology is better than some other existing techniques.

Corresponding Author:

K. Subba Reddy,
Department of Computer Science and Engineering,
Acharya Nagarjuna University,
Guntur, India.
Email: kurapatisr80@gmail.com

1. INTRODUCTION

Social networks are growing rapidly with networks like Facebook and Twitter. These networks [illegible] nication [1]. Social media are primarily internet based tools for sharing and discussing information. In social media people can share comments, personal details, photos, videos and establish relationships with other users [2]. Increase in popularity of internet and social network sites, it is very easy to gather large amount of information. Due to this large amount of data present in social network sites it attracts malicious users [3]. The spammers use their own strategies to attract genuine users, spread misinformation and propaganda. Large scale of spam messages are produced by spammers and spread these messages into social networks. Spam messages mislead the users and occupy the band width. To detect these spam messages various spam detection classification algorithms are used like decision tree induction algorithm, KNN algorithm, Naive bayes algorithm, Neural network algorithm and SVM algorithms [4]. Spam detection methodologies are analyzed by various researchers. Various techniques and frameworks are developed for spam detection [5]. Various spam detection methodologies used KNN and decision tree induction algorithms, but KNN algorithm is better than Decision tree induction algorithm for finding spam messages in social networks [6].

In this paper we proposed an integrated methodology to detect spam messages. In this work, the twitter dataset is used. Thereafter the dataset is pre-processed to obtain normalized set of features depending on which the activities of spammers were studied. The key features extracted were content based and user based features. After obtaining these features, features are combined to improve the spam detection accuracy. In order to improve detection of spam messages, a integrated proposed approach was devised which combines the advantages of the two classification algorithms ie Decision tree algorithm, Naive bayes algorithm. The improvement in spam detection is measured on the basis of accuracy parameters. The results obtained shows that integrated approach that combine algorithms outperforms other classification approaches in terms of overall accuracy. In section 2, several research works are described as related work. Proposed method is presented in section 3. Experimental analysis is presented in section 4. The paper is concluded in section 5 with future enhancements.

2. RELATED WORK

The issue of spam detection in social networking sites is very critical task. The researchers are interested to do their research work on these areas. Many researchers have concentrated to find efficient methods to identify spam. This section summarizes the major contribution of various researchers on spam detection in social networks. Benjamin Markines et. al [7] proposed a spam detection methodology with various supervised learning algorithms. These algorithms are evaluated on six different features such as TagSpam, TagBlur, DomFp, Numads, Plagiarism, ValidLinks. Kyumin Lee et. al [8] describes spam detection approach with honeypots and SVM machine learning algorithm. Xin Jin et. al [9] used GAD clustering algorithm to detect spammers in social networks. This approach deal with scalability and real time spam detection challenges. Xueying et. al [10] classify the social dataset messages into spam messages with using ELM algorithm.

This classification model is developed with various features such as proportion spammer messages with original messages, messages containing URL’s and life time of account. Hongyu et. al [11] filter the spam messages over social network sites. Faraz Ahmed et. al [12] classify the spam profiles in online social networks with Markov clustering. In this approach a weighted graph is used. From this weighted graph find active friends, page likes and shared URLs features. Cheng Cao et. al [13] classify the spam with behavioural analysis of the users. To analyse the behaviour of network users use click based features and post based features. Saini Jacob Soman et. al [14] describes detecting malicious tweets in social network with user based features, location based features, content based features and text based features.

With these features a classification model is proposed by SVM classifier and ELM classifier algorithm. Proposed ELM based spam detection approach performs better spam detection rate compare to SVM. Kaiyu Wang et. al [5] proposed a methodology to detect spam with combining of network features and textual features. With these features a spam detection model is constructed by SVM machine learning algorithm. They measured overall accuracy of model is increased up to 29%. Fabricio Benevenuto et. al [15] classify user profiles into spammers or non spammers based on content based features and user based features. To classify the users they have used support vector machine algorithm. Sajid Yousuf Bhat et. al [16] describes a methodology to detect spam users in social networks by ensemble learning methods. To train these learning algorithms facebook data is used. In this methodology network structure based features are used. Hailu Xu et. al [17] analyzed different features to detect spam in various social network sites such as facebook, twitter.

Arushi Gupta et. al [18] propose a mechanism to detect spammers in twitter network. In their work used tweet level features, user level features, URL’s, spam word features. They have used integrated approach to develop a model with Naive Bayes classifier, clustering and decision trees. Xianghan Zheng et. al [19] studies a methodology to detect spammers in social network. They have used content based features and user based features along with SVM machine learning algorithm to construct a spammer detection model. Zahra Mashayekhi et. al [20] analyzed content based features and non content based features to detect spam in E-mail messages. They have combined decision tree algorithm and Neural Network algorithm to develop a model. They have implemented this model on Lingaspam data. Anjali Sharma et. al [4] analysed various spam detection techniques.

They have studied various origin based spam detection techniques such as Blacklists filters, white lists filters, Realtime Blackhole list filters and content based spam detection techniques such as Rule based filters, Bayesian filters, Support vector machines and Artificial Neural Network algorithms. Saumya Goyal et. al [6] classify the spam messages in twitter network. They have used decision tree and KNN algorithm to construct a classification model. Chen Lin et. al [21] describes a spam detection procedure with Extreme Machine Learning (ELM) algorithm. They have analysed content based features, user based features and social interactivity features. Prabhjot Kaur et. al [22] did the survey on various spam detection techniques. They have done the survey on various user based spam detection techniques, content based spam detection techniques, hybrid based techniques and relation based techniques. Bhagyashri Toke et. al [23] describes a spam detection methodology to detect spam in facebook dataset.

They have used combined approach of Naive bayes algorithm and Rule based algorithm. Bhagyasri Toke et. al [24] analyse the integrated approach to detect spam in social network sites. Ala M et. al [25] describes a spam detection approach to detect spam in social networks. In this approach they have used various features like content features, user based features. They have used various feature selection methods such information gain, relief methods. Malik Mateen et. al [26] describes spam detection approach to detect spam messages in twitter data. Aziah khamis et. al[27] analysed various data mining algorithms to extract for islanding detection in power distribution. Alhamza et. al [28] studied various unsupervised classifiers to classify network flow. Sachin Kamley et. al [29] analysed various machine learning techniques such as Decision tree, Neural Network, Support Vector Machines, Genetic algorithms and Bayesian Networks for performance forecasting of Share Market. Sharvil Shah et. al. [30] studied various classifier algorithms for sentimental analysis in Twitter data. Our proposed approach used hybrid features i.e., combining of content based features, user based features and graph based features.

3. PROPOSED METHOD

Most of the previous research has been done in detecting spam messages and identifying spammer profiles. The previous research papers use different spam message detection and spammer profile detection methodologies. Every methodology uses its own dataset and features for data classification. Spam detection approaches used various kinds of features such as user based features, content based features and graph based features [19], [26]. Each feature set has its own advantages and disadvantages. Based on these features we proposed a methodology that uses combination of user based features and content based features. We use these features to construct a classification model that classify the messages into non spam and spam messages. In our approach we proposed an integrated classification model that uses decision tree algorithm and Naive bayes classifier. An overview of the complete process of spam detection is shown in diagram in Figure 1, each of process steps are explained in this section

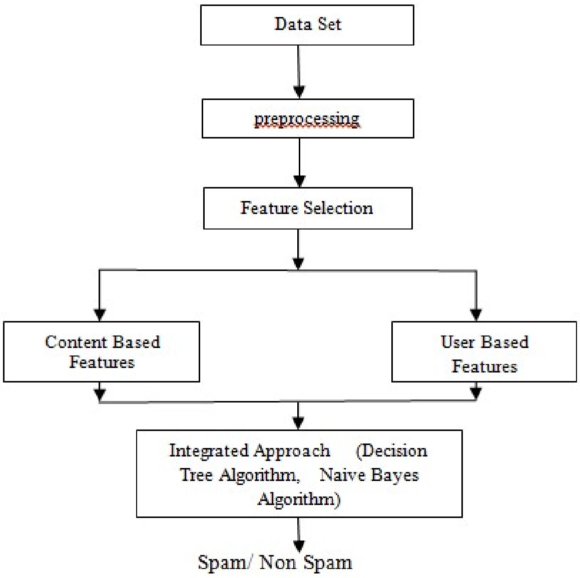

Figure 1. Proposed spam detection approach

A. Dataset description: To evaluate our proposed classification methodology use twitter dataset. There is no publicly available dataset due to twitter policy. So we had to build one such labelled dataset or use the dataset that are used in previous spam detection approaches. Most of the previous researchers have used their own dataset. The researchers extract their required dataset from twitter by using API. Here we used manually collected dataset from twitter for our experimental purpose. The dataset consist of 100 users and 900 tweets.

B. Pre-processing: In this step, all continuous features were converted into discrete features.

C. Feature Selection: From all the dataset features, process relevant features are selected. In this proposed methodology content based features and user based features are selected. These features are integrated to improve the spam detection accuracy.

1. User Based Features: User based features are used to describe the behaviour of users in twitter [16]. These features are based on user relationships and properties of user accounts in twitter dataset. Generally in social media networks users can develop their own social networks with other users. In social network one user follows other users and allows other users to follow him. Spammers want to follow many profiles to spread misinformation to them, so they try to follow large number of users to spread misinformation. Generally we consider, the number of users following is more than number of users following him, such user account is considered as spam account. Here we are using different user based features to construct a model. User features are related to user accounts and the features are extracted from user accounts. The various user based features used in our approach are:
   a) Number of Followers: This feature specifies the number of other users in network follow your account tweets. Generally followers define the popularity of someone profile. Generally spammers have less popularity and have less number of followers.
   b) Number of Following: This feature specifies the number of other user accounts you follow. In twitter if you follow someone means you will see their tweets in your timeline. Twitter network knows to whom you follow and who is following you.
   c) Age of Account: This feature specifies when the account has been created.
   d) Follower to Following Ratio: This is the ratio of followers to following ratio in network for any user account. Generally ff ratio is less for normal users and this ratio is high for spammers.

$$FF\ ratio = \frac{\text{Number of following}}{\text{Number of followers}}$$

   e) Reputation: This is the ratio between number of followers to sum of following and followers

$$Reputation = \frac{\text{Followers}}{Followers + Following}$$

2. Content Based Features: These features are related to tweets posted by user. Generally normal users can't post duplicate content but spammers post lot of duplicate tweets. Content based features are based on messages that users write. The content based features are important to detect spam messages. Spammers are malicious users, who spread large amount of misinformation to the network users [16], [17]. The misinformation contains advertisements about their product and malicious links. The various content based features are used in our approach are:
   a) Number of Tweets: Total number of tweets posted by user after creating his account.
   b) Hashtag Ratio: This is the ratio between the tweets containing hashtags to total tweets posted and those tweets containing unique hashtag.

$$Total\ Hashtags = \frac{\text{Duplicate HashTags}}{\text{Unique Hashtags } \times \text{Tweet Count}}$$

   c) *URL's Ratio*: This is the ratio between duplicate URL's to number of distinct URL's in tweets and sum of tweets.

$$Total\ URLs = \frac{\text{Hash duplicate URLs}}{\text{Hash Unique URLs } \times \text{Tweet Count}}$$

   d) Mentions Ratio: Twitter account users are identified by @username. @username can be written anywhere in the tweet. Spammers misuse this feature to send spam messages to the genuine users in network. Generally the user messages contain large number of mention and reply tags then user is consider as spam user.

$$@Tweets = \frac{\text{Tweets Containing@}}{\text{Total Number of Tweets}}$$

   e) Tweet Frequency: Generally the tweet frequency of spammers is greater than genuine twitter user.

f) Spam words: we use specific spam words and count their occurrence in tweets of users. The spammers use this spam words and spread misinformation to the users.

D. Integrated approach: In our proposed approach we use supervised machine learning algorithms. These algorithms are first trained on the labelled data set to develop classification models. These models are applied on unlabelled data to predict which data as spam data and which data as non spam data. In our proposed approach we integrate decision tree induction algorithm and Naive Bayes classification algorithm to improve spam detection accuracy. In our methodology first decision tree algorithm classifies the dataset as spam or non spam. To improve the classification accuracy of spam detection, categorized spam records of decision tree is given as input to the Naive bayes algorithm. Naive bayes algorithm further classifies any misclassified messages into spam or non spam. In this way categorized non spam messages of decision tree are also given as input to the Naive bayes algorithm to classify the any misclassified messages.

1. Decision Tree Induction: The decision tree is one of the known classification algo rithms used in machine learning to guide the decision making process [28]. Many researchers used [6], [20] this classification algorithm to detect spam messages. The decision tree has three types of nodes. The root node has no incoming edges and zero or more outgoing edges. An internal node has exactly one incoming edge and two or more outgoing edges. The leaf or terminal node has exactly one incoming edge and no outgoing edges. Decision tree induction algorithms must provide a method for expressing an attribute test condition for different types of attributes like binary, nominal, ordinal and continuous attributes. There are many measures that can be used to determine the best way to split the records. These measures are defined in terms of the class distribution the records before and after splitting. The measures developed for selecting the best split are often based on the degree of impurity of the child nodes. Different impurity measures are

Entropy (t) = $-\sum_{i=0}^{c-1} p(i|t)\log_2 p(i|t)$

Gini (t) = $1-\sum_{i=0}^{c-1}[p(i|t)]^2$

Classification error (t) = $1-\max_i[p(i|t)]$

P (i|t) denote the fraction of records belonging to class i at a given node t.

Different decision tree induction algorithms ID3, C4.5, CART, J48

2. Naive Bayes Classifier- This is one of the best machine learning algorithm for spam classification [17], [26]. To classify the message as a spam or non spam can be generalized by probability theory. The spam messages contain the specific words. The relationship between the attribute set and class variable within dataset is non-deterministic. To resolve this problem Bayes theorem introduces a statistical principle for combining prior knowledge of the classes with new evidence gathered from given data. Let X and Y be a pair of random variables. The joint probability, P (X=x, Y=y), gives the probability that variable X will take on the value x and variable Y will take on the value y. The conditional probability is the probability that a random variable will take on a particular value given that the outcome of another random variable is known. The conditional probability p (X=x| Y=y), gives the probability that the variable Y will take on value y, given that the variable X is observed to have the value x. Based on joint and conditional probabilities.

Bayes theorem, $P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$

Y is the event that a given tweet belongs to a given class. X is the d dimensional feature vector corresponding to the tweet. The Naive bayes model makes the independence assumption that the attributes are all independent.

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^{d}(Xi|Y)}{P(X)}$$

The demonstrator is same P(X) for spam and non spam classes. So we can discard for classification.

P(Spam) $\prod_{i=1}^{d}(Xi|Spam)$, and P(Non Spam)$\prod_{i=1}^{d}(Xi|NonSpam)$

Classify the given messages with higher probability. The given training dataset was used to determine the conditional probabilities.

4. EXPERIMENTS AND EVALUATION

For classification task, we trained and tested integrated approach and compared with other machine learning algorithms. In order to evaluate the model, evaluation metrics (precision, recall and F measure) are used. The evaluation metrics are, Precision is the ratio of number of instances correctly classified to the total number of instances and is expressed as

$$precision = \frac{\mathrm{TP}}{\mathrm{TP + FP}}$$

Recall is the ratio of number of instances correctly classified to the total number of predicted instances and is expressed as

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP{+}F}}$$

F measure is the harmonic mean between precision and recall and is expressed as

$$F\ measur = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

Accuracy is the total number of correctly classified instances of both classes over the total number of all instances in the dataset.

$$Accuracy = \frac{\mathrm{TP{+}T}}{\mathrm{TP{+}FN{+}FP{+}T}}$$

We have evaluated the results using two most classifiers, i.e., Decision tree induction and Naive bayes classifiers. Figure 2 shows that results of decision tree and Naive bayes classifiers based on user based features. Naive bayes classifier has closer recall, precision and F-measure as compared to decision tree induction classifier. Figure 3 shows the results of classifiers based on content based features. Decision tree has the highest recall, precision and F-measure compare to Naive bayes algorithm. But Naive bayes algorithm has almost same precision value compared to decision tree. Figure 4 shows the precision, recall and F-measure of classifiers based on combining of user based features and content based features. Compare to individual performance of classifiers based on user based features and content based features hybrid based feature classifiers are outperformed. To improve the further performance of classification model we integrate the decision tree induction classifier and Naive bayes classifier. Figure 5 shows that performance of integrated model based on hybrid features.

We have analysed our results with the earlier works [15], [17], [18], [26], [31]. The earlier works used user based features, content based features and hybrid features. We observe an amount of improvement in the results of our approach from the features used in [15]. The top 20 word features are used [17] for spam detection. Compared to earlier classifications, our approach has highest recall, precision and F measure.

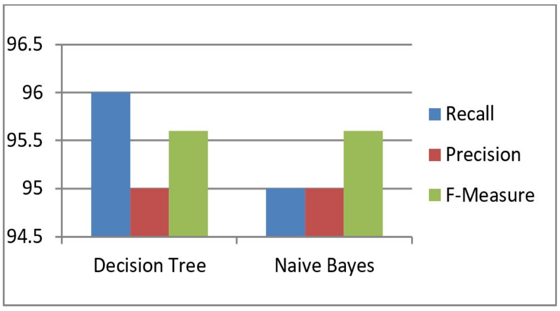

Figure 2. Classification results using user based features

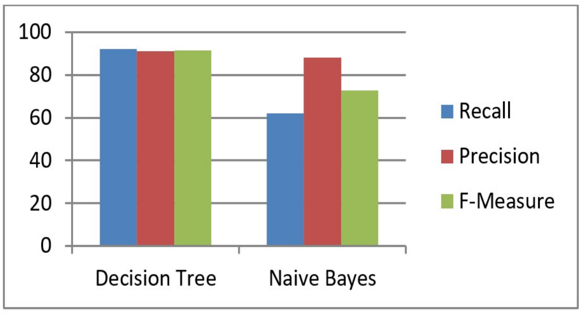

Figure 3. Classification results using content based features

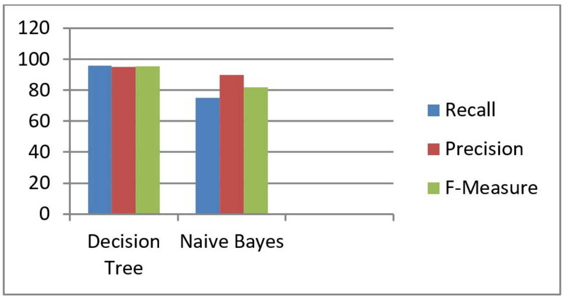

Figure 4. Classification results using hybrid features

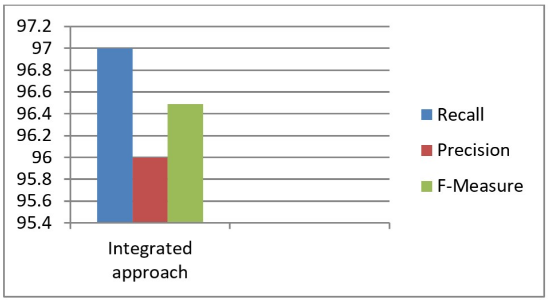

Figure 5. Classification results of proposed approach

In Table 1 we describe the comparison of our integrated approach with other classification models. Compare to other classifications our hybrid features integrated model out performs with high recall, precision and F measure.

Table 1. Comparison between Integrated Approach and Other Classifiers

Classifier	Features	Recall	Precision	F measure
Decision Tree	User based	0.96	0.95	0.9560
Naive bayes	User based	0.96	0.95	0.9560
Decision Tree	Content based	0.92	0.91	0.9149
Naive bayes	Content based	0.62	0.88	0.7274
Decision Tree	Integrated features (user based, content based features)	0.96	0.95	0.9549
Naive bayes	Integrated features (user based, content based features)	0.75	0.90	0.8181
Integrated Classifier (Decision Tree and Naive bayes)	Integrated features (user based, content based features)	0.97	0.96	0.9649

5. CONCLUSION

In this paper we used the hybrid set of features for detecting spam messages on social networking sites. We proposed an integrated spam detection technique to detect spam messages in twitter dataset. In this integrated approach we used decision tree induction classifier, Naive bayes classifier and hybrid features, such as combining of user based features and content based features. Our approach shows performance with high recall, precision and F-measure. In future we will extend our approach to other new set of features and also do our experiments on other social media networks with large amount of dataset.

REFERENCES

[1] “Facebook” https://en.wikipedia.org/wiki/Facebook#Impact
[2] “Twitter” https://en.wikipedia.org/wiki/Twitter#Statistics
[3] “Spam Types” https://en.wikipedia.org/wiki/Social_spam#Types
[4] Anjali Sharma, Manisha, Dr.Manisha, Dr.Rekha Jain, “A Survey on Spam Detection Techniques,” *International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, issue 12,* December 2014.
[5] Kaiyu Wang, Yumei Wang, Hongqiao Li, Yilin Xiong and Xinyu Zhang, “A New Approach for Detecting Spam Microblogs Based on Text and Users Social Network Features,” *National college Innovation program,* Beijing University, china
[6] Saumya Goyal, R.K.sharma, Shabnam Parveen, “Spam Detection using KNN and Decision Tree Mechanism in Social Network,” *Fourth International Conference on Parallel, Distributed and Grid Computing(PDGC), IEEE*, 2016.
[7] Benjamin Markines, Ciro Cattuto and Filippo Menczer, “Social Spam Detection,” Airweb’09 spain, ACM 978-1-60558-438-6
[8] Kyumin Lee, James Caverlee and Steve Webb, “Uncovering Social Spammers: Social honeypots + Machine learning, *SIGIR, july-10, Switzerland, ACM* 978-1-60558-896-4/10/07
[9] Xin Jin, Cindy Xide Lin, Jiebo Luo and Jiawei Han, “SocialSpamGuard: A Data mining based Spam Detection System for Social Media Networks,” *37th international conference on very large data bases, Washington*, 215 8097/11/08, 2011.
[10] Xueying Zhang and Xianghan Zheng, “A Novel Method for Spammer Detection in Social Networks,” *IEEE,* 2015.
[11] Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia and Alok Choudhary, “Towards Online Spam Filtering in Social Networks,” *cucis.ece.northwestern.edu/publications/pdf/GaoChe12.pdf*

[12] Faraz Ahmed and Muhammad Abulaish, "An MCL-Based Approach for Spam Profile Detection in Online Social Networks," *IEEE*, DOI 10.1109/Trustcom.2012.83, 2015.

[13] Cheng Cao and James Caverlee, "Detecting Spam URLs in Social Media via Behavioral Analysis,"*springer*, ECIR 2015, LNCS 9022, pp. 703-714, 2015.

[14] Saini Jacob Soman and Dr. S. Murugappan, "Detecting Malicious Tweets in Trending Topics using Clustering and Classification," *International Conference on Recent Trends in Information Technology, IEEE,* 2014.

[15] Fabricio Benevenuto, Gabriel Magno, Tiago Rodrigues and Virgilio Almedia, "Detecting Spammers on Twitter," *CEAS* 2010-*seventh annual collaboration, electronic messaging, anti-abuse and spam conference july*-2010.

[16] Sajid Yousuf Bhat, Muhammad Abulaish, Abdulrahman A. Mirza, "Spammer Classification using Ensemble Methods over Structural Social Network Features," *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent agent Technologies*, 2014, DOI 10.1109/WI_IAT.2014.133

[17] Hailu Xu, Weiqinh Sun, ahmad Javid, "Efficient Spam Detection across Online Social Networks," *IEEE*-2015

[18] Arushi Gupta, Rishabh Kaushal, "Improving Spam Detection in Online Social Networks," *IEEE*-2015

[19] Xianghan zheng, Zhipeng Zeng, Zheyi Chen, Yuanlong Yu,Chunming Rong, "Detecting Spammers on Social Networks," *Elsevier, NeuroComputing* 159 ,27-34, 2015.

[20] Zahra Mashayekhi, Ali HarounAbadi, "A Hybrid Approach for Detection Based on Decision tree Algorithm and Neural Network," *International Journal of Mechatronics, Electrical and Computer Technology,* 2017

[21] Chen Liu, Genying Wang, "Analysis and Detection of Spam accounts in Social Networks," *2nd IEEE International Conference on Computer and Communications, IEEE*, 2016.

[22] Prabhjot Kaur, Anubha Singhal, Jasleen Kaur, "Spam Detection on Twitter: A Survey," *IEEE,* 2016.

[23] Bhagyashri Toke, Dinesh Puri, "Spam Detection in Online Social Networks using Integrated Approach,'' *International Journal of Innovating Research in computer nd communication Engineering, vol. 4, issue 12,*

[24] Bhagyashri Toke, Dinesh Puri, "Review on Spam Detection in OSN using Integrated Approach," *International Research Journal of Engineering and Technology,* Volume:3, Issue:5,May-2016.

[25] Ala M. AI-Zoubi, Ja far Alqatawna, Hossam Faris, "Spam Profile Detection in Social Networks Based on Public Features," *8th International Conference on Information and Communication Systems, IEEE,* 2017.

[26] Malik Mateen, Mahammad Aleem, Mihammad Azhar Iqbal, Muhammad Arshad Islam, "A Hybrid Approach for Spam Detection for Twitter," *proceedings of 2017 14th International Bhurban Conference on Applied Sciences & Technology, IEEE,* 2017.

[27] Aziah Khamis, Yan Xu, and Azah Mohamed, "Comparative Study in Determining Features Extraction for Islanding Detection using Data Mining Technique: Correlation and Coefficient Analysis," *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 7, No. 3, pp. 1112 – 1124 ISSN: 2088-8708, June 2017

[28] Alhamza Munther, Rozmie Razif, Mosleh AbuAlhaj, Mohammed Anbar, Shahrul Nizam, "A Preliminary Performance Evaluation of K-means, KNN and EM Unsupervised Machine Learning Methods for Network Flow Classification"*, International Journal of Electrical and Computer Engineering (IJECE)* Vol. 6, No. 2, pp. 778~784 ISSN: 2088-8708, April 2016.

[29] Sachin Kamley, Shailesh Jaloree, R. S. Thakur, "Performance Forecasting of Share Market using Machine Learning Techniques: A Review," *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 6, No. 6, pp. 3196~3204 ISSN: 2088-8708, December 2016, DOI: 10.11591/ijece.v6i6.13323

[30] Sharvil Shah*, K Kumar**, Ra. K. Saravanaguru**, "Sentimental Analysis of Twitter Data Using Classifier Algorithms," *International Journal of Electrical and Computer Engineering (IJECE)* Vol. 6, No. 1, pp. 357~366 ISSN: 2088-8708, February 2016, DOI: 10.11591/ijece.v6i1.8982

[31] K Subba Reddy, E Srinivasa Reddy, "An Efficient Methodology to Detect Spam In Social Networking Sites," *International Journal of Computer Science and Information Security (IJCSIS),* Vol 15 No. 7, July 2017

BIOGRAPHIES OF AUTHORS

Mr K Subba Reddy is PhD student in Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, India. He has published papers in international conferences and journals. His area of interest is Big Data, Data mining and machine learning.

Dr E Srinivasa Reddy, PhD., is currently serving as principal in University College of Engineering and Technology and also serving as Head of Department in Computer Science and Engineering, Acharya Nagarjuna University, Guntur, India. He has more than 24 years teaching experience. He is guiding PhD to 8 scholars and 15 has completed his PhD. Dissertations and contributed 45 articles in conferences and 120 papers in Research Journals. His area of interest is Image Processing and Data mining. He may be contacted at: esreddy67@gmail.com