Spam Detection in Social Networking Sites using Artificial Intelligence Technique

Amit Pratap Singh, Maitreyee Dutta

Abstract: ***Social networks provide a way for users to remain in contact with their friends. The increasing popularity of social networks allows social site users to gather large amounts of individual information about their friends. Among numerous sites, Twitter is the fastest growing website. Its popularity has also attracted many spammers to use large amounts of spam to penetrate legitimate users' accounts. In this research work, the Spam detection system in social sites" is designed to detect the spammer by using a machine learning approach. Initially, data is collected from H-Spam14 site and then different pre-processing schemes such as to convert data into lowercase; stop word removal will be applied. After this, the data enters into the feature extraction phase, in which tokenization process is used to divide the entire sentence into a group of words and hence extract the best features from the raw data. To select an appropriate value of extracted feature set, Artificial Bee Colony (ABC) has been applied as an optimization algorithm to determine the optimal feature sets from spam as well as non-spam data. Then, the classification process has been performed using Artificial Neural network (ANN) to distinguish the spam and non-spam data. At the end of the process, performance metrics and comparison will be performed between proposed and existing work to validate the proposed work. The proposed spam detection system can obtained higher accuracy precision, recall and F-measure compared to the existing classifiers such as naïve Bayes and Support vector machine (SVM).***

Keywords: ***Spam detection, Twitter, ABC and ANN.***

I. INTRODUCTION

A number of online social networking sites like Facebook, LinkedIn, and Twitter authorized the social user to meet new people, keep in touch with friends, build professional contacts, and many more [9] .According to the report, Twitter is the fastest growing social networking site among all social networks. Twitter delivers micro blogging services to social users who can send messages to users called tweets. An individual tweet is limited to 140 characters and tweets can be included only in text and HTTP connections. The exchange of tweet allows friends and colleagues to communicate and stay in touch [1]. MICRO-BLOGGING services attracted not only legal users but also spammers. Spam is becoming a growing problem in online social networks like Twitter. Grier et al. have reported that 0.13% of the spam messages are posted on Twitter and is twice compared to the email spam. As Twitter becomes an attractive platform for the spammers as the click rate on Twitter has increased day by day. An increase in spam measures has adversely affected many of the user experience and user behaviour, such as reviewing and recommending [2]. To combat the growing threats of spammers, Twitter uses several ways to outline spam. The spam on Twitter can be reported simply by clicking on the report as a spam website link. The posted report is examined by the Twitter team and the user account will be hanged if it is found as spam. Another way to submit spam on Twitter is to send a tweet to the "@spam @username" format where @username spam accounts are mentioned. However, this service is also exploited by a spammer. Some Twitter apps also permit users to write spam [3].

The management team of the Twitter site put efforts to lock or close the suspicious accounts as well as separate malicious tweets from genuine or useful tweets. But sometimes the legitimate user post complains that their twitter account had been mistakenly stopped by Twitter's managing efforts towards closing the spam accounts[4]. All these usual methods depend on the experience of the user to determine the spam manually. But, nowadays, we need some automatic tools for the detection and close up spammers account. In addition, we need more accurate but effective spam detection methods to avoid dissatisfaction with legitimate users[5].

The main aim of this paper is to provide an automatic detection system for the identification and removal of spam from social networking sites. The main contribution of this paper is as follows:

- To develop a new pre-processing technique based on the corpus method.
- To design a novel fitness function for ABC optimization technique
- To classify the spam and non-spam data, ANN-based training mechanism is designed.
- We compare the efficiency of the designed model with the existing classifiers named as Naïve Bayes and SVM.

II. BACKGROUND AND RELATED WORK

2.1 The Twitter

Twitter is the most fashionable online micro-blogging and social networking website. Just like the other online social media websites, Twitter has low published barriers and permit users to post content in the structure of tweet. Unlike YouTube, Twitter allows users to create post text and attach multimedia content like images, URLs, and videos as outside entities[6]. The snapshot of the Twitter website along with different activities that are performed on the website

Revised Manuscript Received on May 22, 2019.

Amit Pratap Singh, Lecturer,Department of Information Technology, RKGITM, Ghaziabad, India, Email: amitit1991@gmail.com

Dr. Maitreyee Dutta, Professor & Head, Information Management and Coordination Unit, NITTTR, Chandigarh(U.T.), India, Email: d_maitreyee@yahoo.co.in

Retrieval Number: H10070688S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

is depicted in figure 1. A shot of a Twitter program and various activities that could be completed on the website is shown in figure 1. The text on Twitter posts is known as tweets with 140 characters and so is called microposts[7].

Figure 1: Snapshot of Twitter[16]

2.2 Related work

Cao et al. (8, 2015) presented a scheme utilized for identifying spam URLs in social sites which have been used to protect users from links that are related with malware and other low-quality suspicious text. The behaviour has been analyzed using two different schemes (i) initially, study the links posted by publicon Twitter; (ii) secondly is how these links are accessed by the user. **Jain et al. (9, 2018)** proposed a CNN (Convolution neural network) as a classification technique to detect spam in the social network. A semantic layer has been added and hence the proposed model is known as SCNN (Semantic convolution neural network). Word2Vec has been utilized for training SCNN and hence obtained semantic enrich words. 94.40 % of accuracy has been obtained compared to the Twitter dataset. **Ezpeleta, et al (10, 2018)** proposed a classifier "Bayesian spam filtering" for detecting the spam. To analyze the outcome of the designed algorithm the experiment has been carried out on two different datasets such as "Youtube Comments dataset" and Youtube spam collection dataset. **Dwyer et al. (11, 2007)** proposed a Bayesian classification approach to differentiate between genuine and suspicious behaviour of the message. The performance for precision, F-measure has been obtained and concluded that Bayesian classifier works well among other existing algorithms. **Dutta et al. (12, 2018)** developed an attribute selection approach by using the principle of rough set theory. The experiment has been performed on five various spam classification datasets and compares the results with the proposed method. Attribute selection is the main process in machine learning with data mining **Ala'M et al. (13, 2018)** presented a hybrid machine learning algorithm that comprises of SVM as a classifier and whale as an optimization algorithm. These algorithms have been utilized for recognizing spammers in OSNs. **Aslan et al. (14, 2018)** presented an automatic detection system for providing security to Twitter users against spammers. In this paper, the author used three classification techniques named as SVM, random forests and decision tree. From the experiments, it has been observed that when random forest used with decision tree algorithm the overall accuracy up to 95% have been obtained. In the case when the random forest is used with behavioural feature the accuracy of the detection system increases and become 97.877%.

III. PROPOSED SPAM DETECTION MODEL

The proposed spam detection consists of four modules named as input, processing, classification and evaluation as depicted in figure 2. It is assumed that ANN is trained as per the text data features.

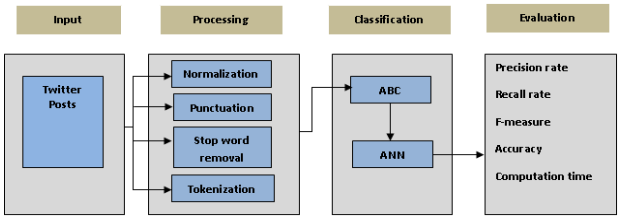

Figure 2: Designed Model

The designed model consists of different phases such as input, processing, classification and evaluation as depicted in figure 2. The twitter posts are applied as input data to the model. As the twitter posts consist of raw data, therefore, it is necessary to remove the unwanted text and receive the tweets that include the words that are essential for the classification process. In pre-processing number of processes are carried out (i) Normalization is used to convert the input tweets in small letters (ii) in punctuation process, the marks like as comma, full stop, brackets and so on that are used to separate sentence are removed, (iii) Stop word removal process, the stop words such as *is, am, are, there, here, that, those, which and so on* that are very common in all the sentence are removed. (iv) in the last step of pre-processing, tokenization is used to find features. Tokenization process breaks the sentence into individual words. It also used to determine the weight of the string with respect to the alphabets. The token words are applied as an input to the ABC algorithm, to obtain optimized features of the alphabet. The fitness function is defined on the basis of which the alphabets are optimized. The value of the features which is greater than the fitness function is applied as an input to the classification algorithm; otherwise, the remaining value is ignored. ANN is trained as per the optimized features of the spam data and during the testing process; the text is compared and analyzed on the basis of the performance parameters as illustrated under the evaluation block.

3.1 ABC

In the proposed work, ABC algorithm is utilized for discovering the optimal feature sets from the spam and non-spam files. The features of the spam and non spam text will be optimized using ABC algorithm by designing an appropriate fitness function. Using fitness function a threshold value will be selected on the basis of which the text has been considered as spam and non- spam. To optimize the extracted feature sets, ABC algorithm will be used and the unwanted feature sets are eliminated. The algorithm is written below:

ABC Algorithm:

Retrieval Number: H10070688S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Input: Feature of Spam & Non-spam data in terms of token value and Fitness Function of ABC
Output: Optimized token value

Step 1: Initialize the ABC with their operating functions – Bee Size
– Employee Bee
– Onlooker Eggs
– Scout Eggs
Step 2: Define fitness function of ABC for optimal feature selection

$$Fitness\_function = \begin{cases} True; & if\ Bee_{Employee} > Bee_{Onlooker} \\ False; & otherwise \end{cases}$$

Step 3: Calculate size of Feature in terms of row and columns (R, C)
Step 4: Start feature selection
Step 5: **For** i=1 → R
Step 6: **For** j→1 **to** C
Step 7: $Bee_{Employee}$= Current Bee (i, j)
Step 8: $Bee_{Onlooker}$ = Respective Bee (i, j)
Step 9: Fit_function =Call Fit_function ($Bee_{Employee}$, $Bee_{Onlooker}$)
Step 10: Fitdata=CS (Fit_function, Feature of Text)
Step 11: **End**
Step 12: **End**
Step 13: **Returns:** Fit data as an optimized token value of text file

3.2 ANN

ANN (artificial neural network) is used to differentiate the spammer and the genuine user. ANNs are computer programs with biological inspirations designed to mimic the way in which human information processes the information via brain. ANN collects the awareness by noticing patterns and data relationships and learning (or trained) during experience, not from programming. The ANN is developed from varied units, processing elements or artificial neurons related to weights, which form the neural formation and are arranged in layers. The ANN algorithm for spam detection is written below:

ANN Algorithm:

Input: Feature of Spam & Non-spam data in terms of token value and Fitness Function of ABC
Output: Optimized token value

Step 1: Initialize the ABC with their operating functions – Bee Size
– Employee Bee
– Onlooker Eggs
– Scout Eggs
Step 2: Define fitness function of ABC for optimal feature selection

$$Fitness\_function = \begin{cases} True; & if\ Bee_{Employee} > Bee_{Onlooker} \\ False; & otherwise \end{cases}$$

Step 3: Calculate size of Feature in terms of row and columns (R, C)
Step 4: Start feature selection
Step 5: **For** i=1 → R
Step 6: **For** j→1 **to** C
Step 7: $Bee_{Employee}$= Current Bee (i, j)
Step 8: $Bee_{Onlooker}$ = Respective Bee (i, j)
Step 9: Fit_function =Call Fit_function ($Bee_{Employee}$, $Bee_{Onlooker}$)
Step 10: Fitdata=CS (Fit_function, Feature of Text)
Step 11: **End**
Step 12: **End**
Step 13: **Returns:** Fit data as an optimized token value of text file

3.3 Naïve Bayes

It is a machine learning approach, that works on Baye's theorem In an easy point of view, Naive Bayes believes that the existence of a particular attribute in one class depends on the existence of another attribute[15]. The Naive Bayes model is very easy to install and is mainly useful for large and complex data sets. In addition to simplicity, it is even better than the high class comprehensive classification methods. For some probability models, naive Bayes classifiers can receive very productive training within the controlled learning setting. In many practical applications, the parameter estimation for Naive Bayes models uses the maximum probability method; In other words, the Bayesian can work with the naive Bayes model without accepting the possibility or using Bayesian methods. Despite their extreme simplicity of their loyal design and appearance, the naive Bayes classifiers worked well in a number of complicated real world situations. In, this research the optimized text obtained from ABC algorithm are used to trained Naivy Bays and this approach is used to classify spam in the text.

3.4 SVM

It is a supervised learning approach, which is used to resolve the regression as well as classification problem. In this approach, the data is plotted in the n-dimensional space and each data comprises of feature value along with their coordinate. The hyperplane has been determined to differentiate normal and spam text. In this research, SVM is used to separate the category of normal text from the spam text.

IV. RESULTS AND SIMULATION

The experiment has been performed on Twitter dataset taken from Hspam 14. HSpam14 comprises of more than 14 million tweets. The data has been collected using the trending theme on Hashtags.org, 2019. In this research, we have used the last couple of day tweets. Almost all tweets in HSpam14 are marked as spam and ham, and the remaining small parts are marked as unknown because their labels cannot be determined even with manual inspection.

Table 1: Computed parameters

Test samples	Error (%)	Execution Time (s)	Precision	Recall	F-measure	Accuracy	Classification result
1	1.03	0.096	0.6869	0.9815	0.9841	98.96	Spam
2	0.75	0.024	0.9863	09817	0.984	99.24	Spam
3	0.79	0.021	0.9865	0.980	0.983	99.20	Spam
4	0.945	0.017	0.9862	0.978	0.982	99.05	Non-Spam
5	0.94	0.017	0.9867	0.982	0.9843	99.05	Spam
6	0.73	0.023	0.9868	0.979	0.9831	99.26	Spam
7	0.081	0.011	0.9869	0.980	0.983	99.26	Spam

Retrieval Number: H10070688S319/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

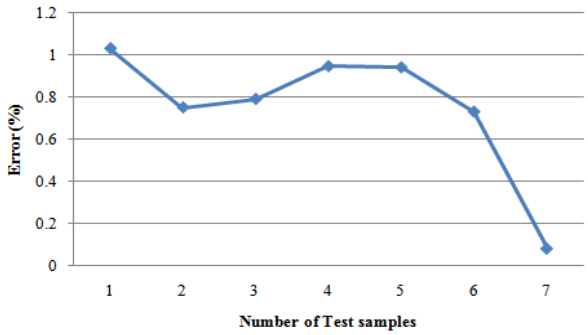

Figure 3: Error(%) vs Number of test samples

The error occurred during the classification process of spam is illustrated in figure 3. The average of error observed for the seven different test data sample is about 0.752 (%), which is very small means that the designed spam model classify spam and no-spam text with high accuracy.

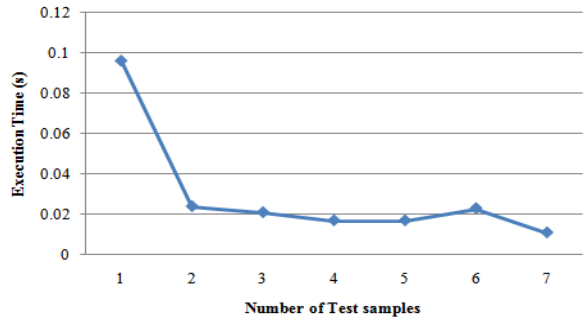

Figure 4: Execution Time (s) vs Number of test samples

Execution time is defined as the total time required by the designed model starting from the testing process until the classification and evaluation process. The average execution time determined for seven test samples is approximately 0.029 s, which means that the system is fast enough and detects spam immediately.

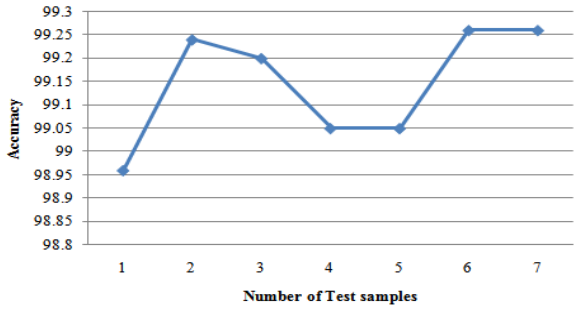

Figure 5: Accuracy v/s Number of test samples

The accuracy of the designed system is shown in figure 5; the average accuracy of about 99.14% has been detected. This means that the model adaptively learns the new spam activities and maintains high accuracy for spam detection in a tweet post.

Table 2: Comparison of proposed work with existing work

		Proposed Work				Naïve Bayes			SVM			
Test samples	**Accuracy**	**Precision**	**Recall**	**F-measure**	**Accuracy**	**Precision**	**Recall**	**F-measure**	**Accuracy**	**Precision**	**Recall**	**F-measure**
1	98.96	0.99	0.98	0.98	93.25	0.95	0.96	0.95	90.25	0.94	0.94	0.94
2	99.24	0.98	0.98	0.98	94.75	0.97	0.96	0.96	91.73	0.96	0.95	0.95
3	99.20	0.98	0.98	0.98	94.12	0.93	0.96	0.94	91.85	0.94	0.92	0.92
4	99.05	0.98	0.97	0.98	94.86	0.94	0.94	0.94	91.75	0.92	0.91	0.91
5	99.05	0.98	0.98	0.98	93.89	0.97	0.93	0.94	91.68	0.95	0.96	0.95
6	99.26	0.98	0.97	0.98	94.12	0.92	0.92	0.92	91.75	0.91	0.95	0.92
7	99.26	0.98	0.98	0.98	94.86	0.97	0.94	0.95	90.57	0.96	0.97	0.96

Retrieval Number: H10070688S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

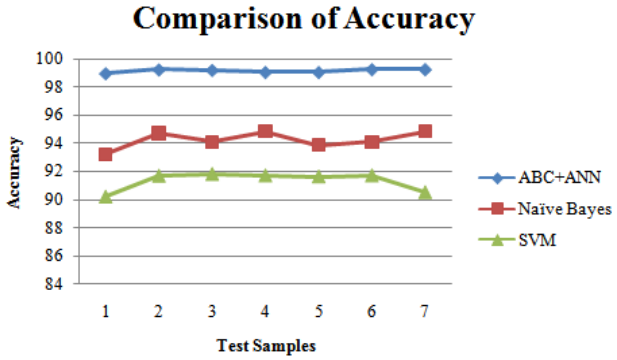

Figure 6: Comparison of Accuracy

Figure 6 represents the accuracy graph designed for proposed algorithm (ABC with ANN) along with two other existing classification algorithms named as Naïve Bayes and SVM. From the graph it is clear that when the machine learning scheme is trained with the optimized features the accuracy of the system is higher compared to the individual classifiers. The average accuracy of proposed work, Naïve Bayes and SVM are 99.14, 94.26 and 91.36 respectively. Thus there is an increase of 5.18 % from Naïve Bayes and 8.52 % from SVM.

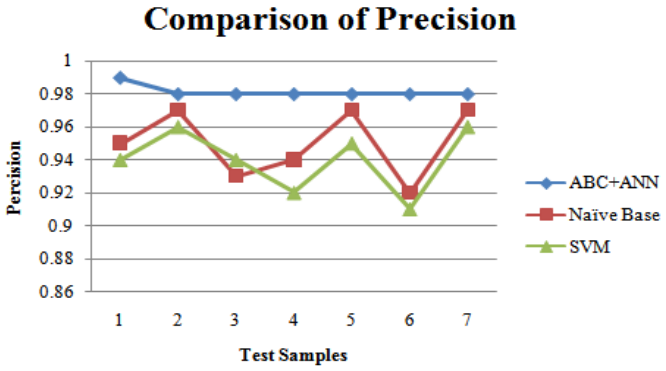

Figure 7: Comparison of Precision

The comparison of precision examined for the proposed model has been compared with the existing classifiers Naïve Bayes and SVM. From the graph, it has been clear that the detection of identifying spam using the ABC algorithm in hybridization with ANN perform better compared to the Naïve Bayes and SVM. The average value of precision measured for the proposed, Naïve Bayes and SVM are 0.98 and 0.95, and 0.94 respectively. Thus, there is an enhancement of 3.16% in the precision rate while ANN with ABC algorithm compared to naivey Bayes and of 4.26% compared to SVM approach during the classification of spam in the designed model.

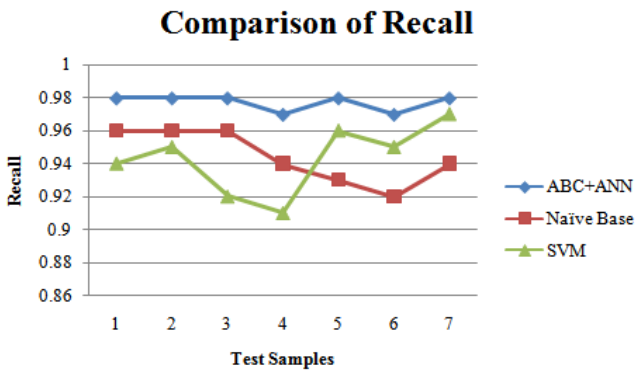

Figure 8: Comparison of Recall

The recall rate computed by Naïve Bayes, SVM and the proposed work is depicted in figure 8. The red and the blue and the green colour line represent the recall rate observed for seven number of test samples determined for Naïve Bayes, proposed and SVM respectively The average value of recall measured for these three classification algorithms (ABC+ANN, Naïve Bayes and SVM) are 0.977, 0.944 and 0.942 respectively. It is observed that the recall rate of the proposed work has been increased by 3.5% from naivey bayes and 3.72 % from SVM approach.

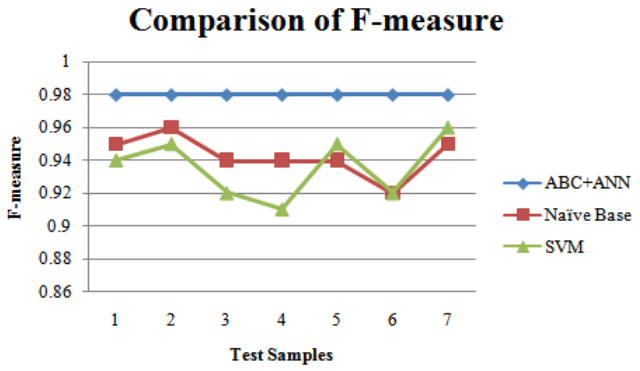

Figure 9: Comparison of F-measure

F-measure is the harmonic means of precision and recall. The values measured for the F-measure parameter for proposed along with Naïve Bayes and SVM is depicted in figure 9. The average value of F-measure determined for the proposed, Naïve Bayes and SVM are 0.98 and 0.94, 0.93 respectively. It is observed that the F-measure of the proposed work has been increased by 4.26% from naïve Bayes and 5.38% from SVM respectively.

V. CONCLUSION

In this research, we focused on the detection of social spam preferably in the Twitter micro-blogging site. ANN approach with ABC has been applied to differentiate between the spam and the non –spam tweets. According to twitter's spam policy, features of the text data are extracted using tokenization mechanism. The experimental results demonstrated the effectiveness of the proposed work in terms of computed parameters such as error, execution time, accuracy, precision, recall and F-measure. From the experiment, it has been

Retrieval Number: H10070688S319/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

observed that pre-processing, optimization with classification technique increased the accuracy of the spam detection system. The accuracy of the proposed system to detect spam in the Twitter site of about 99.14% has been achieved. At last, the comparison between proposed technique and existing classification algorithms (Naïve Bayes and SVM) has been provided. From the experiment, it has been observed that the proposed scheme (ABC with ANN) has perform well compared to individual classifiers (Naïve Bayes and SVM).
In future, the work can be extended by using other social media dataset and identifying the spam.

REFERENCES

1. Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010, October). @ spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security* (pp. 27-37). ACM.
2. Wang, A. H. (2010, June). Detecting spam bots in online social networking sites: a machine learning approach. In *IFIP Annual Conference on Data and Applications Security and Privacy* (pp. 335-342). Springer, Berlin, Heidelberg.
3. Wu, T., Liu, S., Zhang, J., & Xiang, Y. (2017, January). Twitter spam detection based on deep learning. In *Proceedings of the Australasian Computer Science Week Multiconference* (p. 3). ACM.
4. Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015). Detecting spammers on social networks. *Neurocomputing*, *159*, 27-34.
5. Alsaleh, M., Alarifi, A., Al-Salman, A. M., Alfayez, M., & Almuhaysin, A. (2014, December). Tsd: Detecting sybil accounts in twitter. In *2014 13th International Conference on Machine Learning and Applications* (pp. 463-469). IEEE.
6. Verma, M., & Sofat, S. (2014). Techniques to detect spammers in twitter-a survey. *International Journal of Computer Applications*, *85*(10).
7. Wang, D., Irani, D., & Pu, C. (2011, September). A social-spam detection framework. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference* (pp. 46-54). ACM.
8. Cao, C., & Caverlee, J. (2015, March). Detecting spam urls in social media via behavioral analysis. In *European Conference on Information Retrieval* Springer, Cham, pp. 703-714.
9. Jain, G., Sharma, M., & Agarwal, B. (2018). Spam detection on social media using semantic convolutional neural network. *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, *8*(1), 12-26.
10. Ezpeleta, E., Iturbe, M., Garitano, I., de Mendizabal, I. V., & Zurutuza, U. (2018, June). A Mood Analysis on Youtube Comments and a Method for Improved Social Spam Detection. In *International Conference on Hybrid Artificial Intelligence Systems* Springer, Cham, pp. 514-525.
11. Dwyer, C., Hiltz, S., & Passerini, K. (2007). Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *AMCIS 2007 proceedings*, 339.
12. Dutta, S., Ghatak, S., Dey, R., Das, A. K., & Ghosh, S. (2018). Attribute selection for improving spam classification in online social networks: a rough set theory-based approach. *Social Network Analysis and Mining*, *8*(1), 7.
13. Ala'M, A. Z., Faris, H., & Hassonah, M. A. (2018). Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts. *Knowledge-Based Systems*, *153*, 91-104.
14. Aslan, Ç. B., Sağlam, R. B., & Li, S. (2018). Automatic Detection of Cyber Security Related Accounts on Online Social Networks: Twitter as an example.
15. Sedhai, S., & Sun, A. (2018). Semi-supervised spam detection in the Twitter stream. *IEEE Transactions on Computational Social Systems*, *5*(1), 169-175.
16. http://shodhganga.inflibnet.ac.in/bitstream/10603/183741/7/07_chapter%201.pdf

BIOGRAPHIES OF AUTHORS

Amit Pratap Singh received his Bachelor's degree in Information Technology from GBTU, and currently perusing Master's in computer science and engineering from NITTTR, Chandigarh.His current research interest include data mining.

Dr. Maitreyee Dutta received her Bachelor's degree in Electronics and Communication Engineering from Guwahati University and Master's in Electronics and Communication Engineering from Panjab University, Chandigarh. She did her Ph.D. degree in Engineering and Technology from Panjab University, Chandigarh. Her current research interests include Digital Signal Processing, Advanced Computer Architecture, Data Warehousing and Mining, Image Processing.

Retrieval Number: H10070688S319/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication