International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11s
DOI: https://doi.org/10.17762/ijritcc.v11i11s.8086
***Article Received:** 21 June 2023 **Revised:** 18 August 2023 **Accepted:** 02 September 2023*

Enhanced Spam Detection System for Twitter Social Networking Platform

Partibha Yadav[1], Ajay Kumar[2], Shivani[3], Reena Hooda[4], Sudhir[5], Pooja[6]
[1]Department of Mathematics, Indira Gandhi University Meerpur, Rewari, Haryana, India
E-mail: pratibha1007@gmail.com
[2]Department of Computer Science, Indira Gandhi University Meerpur, Rewari, Haryana, India
E-mail: yajay2@gmail.com
[3]Department of Computer Science, Indira Gandhi University Meerpur, Rewari, Haryana, India
E-mail: shivanigupta646@gmail.com
[4]Department of Computer Science, Indira Gandhi University Meerpur, Rewari, Haryana, India
E-mail: reenah2013@gmail.com
[5]Department of Computer Science, PGT, Board of School Education, Rewari, Haryana, India
E-mail: sudhir.yadav984@gmail.com
[6]Department of Computer Science, Indira Gandhi University Meerpur, Rewari, Haryana, India
E-mail: poojay220@gmail.com

Abstract: Twitter social site is one of the most popular Online Social Networking Site (OSN) used by popular people such as Ministers, businessman, large companies, actors to share their information. In this site, around 500 million of tweets are posted monthly by the total 313 million Twitter active users. The widespread of Twitter has drawn the interest of spammers. These malicious actors exploit the platform for various nefarious purposes, including monitoring authentic users, disseminating harmful software, and promoting their agendas through URLs embedded in tweets. They engage in tactics like secret following and unfollowing legitimate users, all with the intent of gathering sensitive information.To resolve this problem, a secure spam detection based on machine learning approach is designed. The designed used stop word removal, word to vector model to refined and dimensionally reduced the data. To enhance the quality of the data Cosine similarity is also been applied to measure the similarity score among the tweets and based upon that Artificial Neural Network (ANN) is trained. Later on, it is used to test the efficiency by examining the performance parameters in terms of precision, recall and F-measure. Also, the comparative analysis has been performed to present the efficiency of the work. The average precision, recall and F measure of proposed spam detection model of 0.9252, 0.6107 and 0.734 are obtained.

Keywords-Spam, Twitter, Cosine Similarity, Artificial neural Network.

I. INTRODUCTION

In the present time, the use of Online Social Networking sites tremendously increases and the users posts their ideals and share their feelings around the world [1].Especially notable is twitter’s ability to allure users due to its straightforward nature as a social network. It grants you the means to stay informed about your areas of interest, whether they involve communities, celebrities, or even individuals who aren’t widely known but are familiar to you.It provides free services to access social networking sites and can send or receive messages maximum of 140 characters. Using OSN, the user can interconnect to other user and hence share their information’s. From a survey it has been analysed that there are around 42 million of new users registered in Twitter [2].

As shown in Figure 1, the growth of users increases linearly from the year 2010 to2020 and also predicted the same for the upcoming three years (2021 to 2023). Unfortunately, due to the increase in the Twitter popularity, spammers can post a lot of malicious information that includes suspicious URLs along with multiple duplicate posts to attack normal users and hence steal their personal information[3].Spammers also use offensive terms on popular Twitter topics many times, and these topics appear on Twitter's homepage many times. These activitiesenforced Twitter to momentarily disable popular topics and eliminateinvasive terms [4]. Twitter tried numerous methods to combat spam, including adding a "spam report" feature to its service and hence secure Twitter accounts from suspicious users. However, legitimate Twitter users voice their grievances regarding the platform’s anti-spam measure[5]. In a recent development, Twitter acknowledged that it had inadvertently suspended its own account while attempting to remove spam content.

In this research article, the spam bot behaviour is studied. The aim of this research is to apply artificial intelligence techniques to distinguished between normal and spam bots [6].

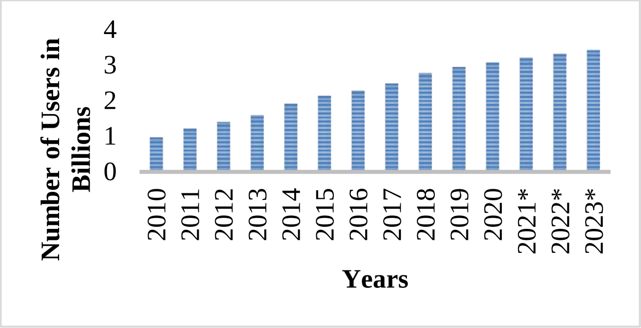


Figure 1: Growth of Twitter User https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

The organization of the paper is as follows: in section 2, state-of art is discussed. A novel designed approach along with the used methods are discussed in section 3 that helps to detect spam from normal tweets, Section 4 discussed the simulation results. Conclusion is presented in section 5 followed by the references.

II. RELATED WORK

The serious spam problem on Twitter has attractedresearchers to design a secure OSN. Few of researchers have studied the characteristics of spam and then based on the analysed features of spam, researchers have proposed some important work to detect Twitter spam. Therefore, we discuss previous state-of-artsrelated to the detection of spam on Twitter.

To detect Twitter spam, many researchers **Yang et al. (2013)** worked on it. But, most of the researchers have utilized machine learning technique to distinguished spam bot and normal. Initially, the researchers such as **Chen et al. (2015)[8]** have utilized the content features like as life of user account, count of followers, ratio of URL and the tweet length to separate normal users from spammers. It has been concluded that the above defined features can be extracted easily. Also, the researchers, **Wang (2010)[9] and Stringhini (2010)[10]**have used social graph to extract features and hence to evade feature fabrication.**Song et al. (2011)** have distinguished spam bot from normal one using distance between the tweeter transmitter and a receiver user**[11]**. In 2013, Yang et al. have presented an improved twitter spam detection system in which the features have been separated using social graph by utilizing clustering coefficient, link relation in both directions. However, gathering these features is a highly time-consuming process, resulting in the creation of a sizable graph.

Also, the collection of such features of tweets is impracticable as the incoming data is in the form of streams**[12]. Thomas et al. (2011)** have used URL in tweets to identify spam bot. The URL features like as token of domain, path token, domain name system (DNS)'s features along with information related to landing page have been used to identify spam bot**[13].Al-Janabi et al. (2017)**have used random forest as classification approach to distinguish spam bot by labelling the collected data as normal and spam and hence during testing classify that data based on their label. The performance shows precision and recall of 89 %and 92 % respectively**[14].**

III. PROPOSED WORK

The entire procedure performed to design a secure social network (Twitter) against spam bot using Artificial Neural network as a machine learning approach is illustrated in Figure 2. Before, classifying tweets as spam or normal, the training of ANN has been performed based on the collected data set “” ”. The dataset comprises a total of 200,000 tweets, each accompanied by its respective URL. An underlying presumption is that all tweets include URLs, with the intention of luring social media user towards malicious destinations like spam bots.After training process, the system is ready to detect uploaded tweets as spam or normal. In brief, we can say that the whole work has been performed in two phases that is (i) Training and (ii) Classification.

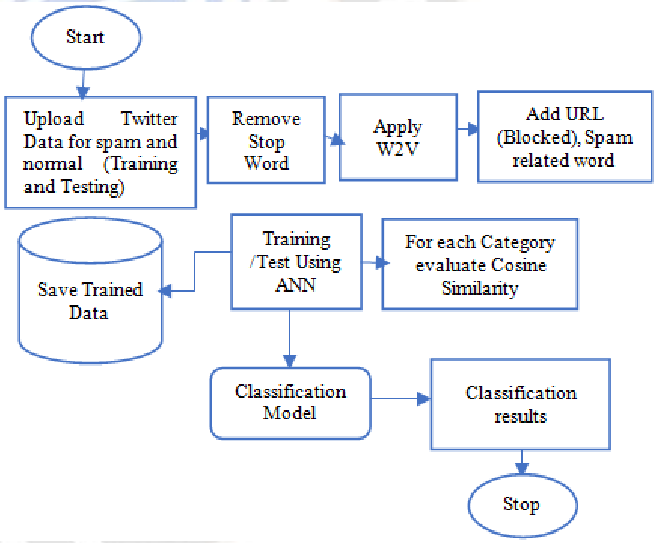

Start
Upload Twitter Data for spam and normal (Training and Testing)
Remove Stop Word
Apply W2V
Add URL (Blocked), Spam related word
Save Trained Data
Training /Test Using ANN
For each Category evaluate Cosine Similarity
Classification Model
Classification results
Stop

Figure 2: Design of a secure social network (Twitter) against spam bot using Artificial Neural network

A. *Stop Word Removal*

After uploading dataset, stop word (do not contains essential information) are removed from the dataset so that the data contains only the relevant information. This process helps to reduce data size as well as to enhance the quality of data, thus increasing training and then classification accuracy. Few words that are removed from the dataset as stop words are listed in Table 1.

DOI: https://doi.org/10.17762/ijritcc.v11i11s.8086
Article Received: *21 June 2023* ***Revised:*** *18 August 2023* ***Accepted:*** *02 September 2023*

Table 1: Stop words (https://gist.github.com/sebleier/554280)

other	some	such
being	have	has
having	do	Does
few	doing	A
more	The	but

B. *Word to Vector (W2V)*

It is one of the fastest ways to learn word representation in txt data. The word representation is performed by predicting surrounding word around the sentence/word using skip-gram scheme. Also, the word prediction can be performed by finding the sentence's central word using Continuous Bag of word architecture. Both techniques are simple and easy to learn. After applying W2V technique, the data is converted into weighted numbers and then URL based features is applied.

C. *URL based features*

URL is abbreviated as "Uniform Resource Locator**", which is the** World Wide Web (WWW) access link that includes protocols, host address, path along with dictionary as shown in Figure 3.

For information exchange among devices like as HTTP, FTP; URL used protocol along with the defined strategy. For the separation between the server and the web hosts, a domain name is used, which comprises of a registered domain name (2nd level domain) along with the top-level domain. Using this technique those URL that contains spammer and normal data is levelled and pass to the cosine similarity model.

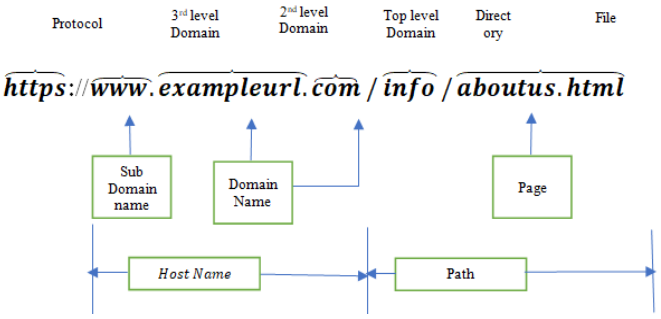


Figure 4: Example of a URL

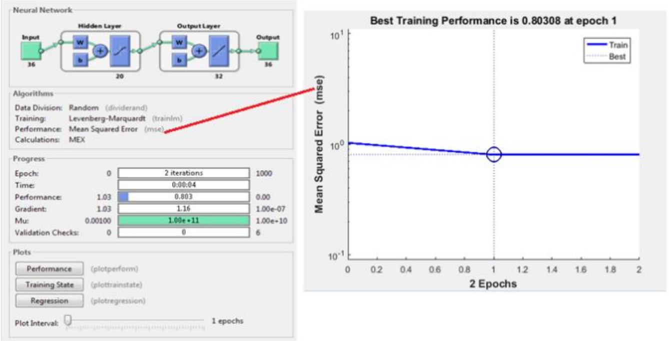
D. *Cosine Similarity*

Cosine similarity is applied to group words of tweets into clusters based on the calculated cosine similarity score. The formula used to calculate tweets similarity is written by equation (1).

$$Sim_{i,j} = \frac{W_i \times W_j}{\|W_i\| \|W_j\|} \qquad (1)$$

W_i and W_j are the weight generated by the W2V model for tweet 1 and tweet 2 respectively.

Based on the obtained similarity score, ANN is trained and later used for classification of spam bot and normal. The trained ANN structure along with MSE (mean square Error) is shown in Figure 5.



Algorithms
Data Division: Random (dividerand)
Training: Levenberg-Marquardt (trainlm)
Performance: Mean Squared Error (mse)
Calculations: MEX

Progress

Epoch:	0	2 iterations	1000
Time:		0:00:04	
Performance:	1.03	0.803	0.00
Gradient:	1.03	1.16	1.00e-07
Mu:	0.00100	1.00e+11	1.00e+10
Validation Checks:	0	0	6

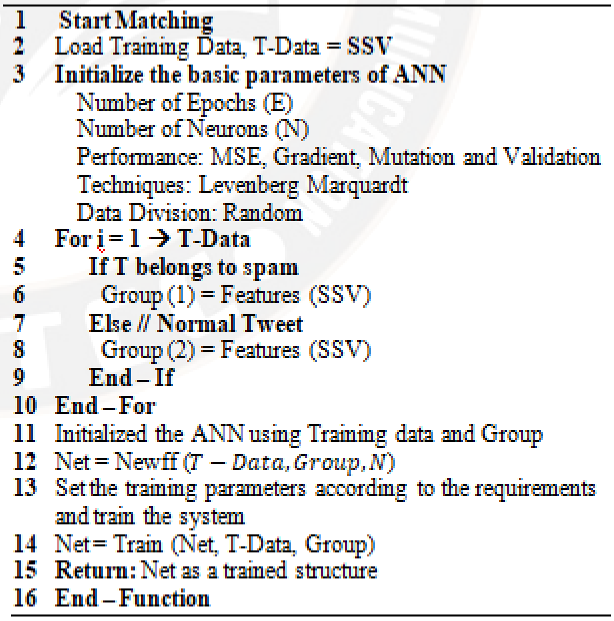
Plots
Performance (plotperform)
Training State (plottrainstate)
Regression (plotregression)
Plot Interval: 1 epochs

Figure 5: Training of ANN with MSE value

Figure 5 represents the MSE value along with the training of designed spam detection twitter system. The system has been designed with minimum error of about 0.80308 at 1 epoch.

Table 2: Input and output for ANN

Required Input:	SSV← Training Data as a similarity score value
	S← Target/Category of spam and non- spam data
	N← Total Number of 'Neurons'
Obtained Output:	Net ←'ANN' Trained structure



```
Start Matching
Load Training Data, T-Data = SSV
Initialize the basic parameters of ANN
    Number of Epochs (E)
    Number of Neurons (N)
    Performance: MSE, Gradient, Mutation and Validation
    Techniques: Levenberg Marquardt
    Data Division: Random
For i = 1 → T-Data
    If T belongs to spam
        Group (1) = Features (SSV)
    Else // Normal Tweet
        Group (2) = Features (SSV)
    End – If
End – For
Initialized the ANN using Training data and Group
Net = Newff ($T - Data, Group, N$)
Set the training parameters according to the requirements
and train the system
Net = Train (Net, T-Data, Group)
Return: Net as a trained structure
End – Function
```

At last, the performance of the designed system has been computed as described in section IV.

DOI: https://doi.org/10.17762/ijritcc.v11i11s.8086
***Article Received:** 21 June 2023 **Revised:** 18 August 2023 **Accepted:** 02 September 2023*

IV. PERFORMANCE EVALUATION

The performance of the designed Machine learning based spam detection model is evaluated using the following parameters:

$$Precision = \frac{T_p}{T_p + F_p} \qquad (2)$$

$$Recall = \frac{T_p}{T_p + F_n} \qquad (3)$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (4)$$

Here, T_p →Total tweets counted as spam and also predicted as nasty.

F_n → the number of tweets that are being predicted as real but are junk and contains nasty information.

F_p →The number of tweets are real but considered as spam

T_n → The number of predicted real tweets.

Table 3: Precision

Number of Uploaded Tweets	**Cosine Similarity**	**ANN**
100	0.752	0.943
200	0.746	0.940
300	0.728	0.936
400	0.708	0.922
500	0.697	0.917
600	0.682	0.911
700	0.638	0.908

The precision value analysed for the proposed machine learning based spam detection system using cosine similarity and ANN is shown in Figure 6. The orange and the blue bar represent the precision values for ANN and Cosine similarity respectively. The graph has been plotted for number of uploaded Tweets with respect to the precision values for two techniques. The rate of detecting spam and normal tweets using ANN approach is high this is because of the appropriate training as the data is passed through different process and hence ANN is trained using refined data obtained after applying stop word removal, W2V and then similar score is determined using Cosine similarity measure. The average precision value examined using cosine similarity and ANN with Cosine are 0.707 and 0.9252 respectively, which indicates that the uploaded test data has been identified as spam and normal with higher accuracy.

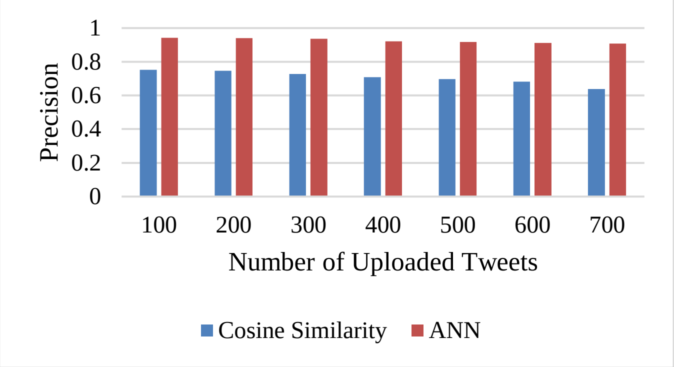


Figure 6: Precision

Recall is the fraction of the desired results that is spam and normal tweets in this case, which have been detected successfully. As in Figure, recall rate has been examined by uploading Tweets ranges from 100 to 700 in step of 100. Therefore, recall is calculated as the number of correctly detected spam/non-spam tweets divided by the total number of uploaded tweets that must be returned. Here, the recall value analysed for Cosine similarity and ANN are 0.7204 and 0.6107 respectively.

Table 4: Recall

Number of Uploaded Tweets	**Cosine Similarity**	**ANN**
100	0.725	0.652
200	0.735	0.642
300	0.741	0.625
400	0.723	0.618
500	0.716	0.609
600	0.705	0.571
700	0.698	0.558

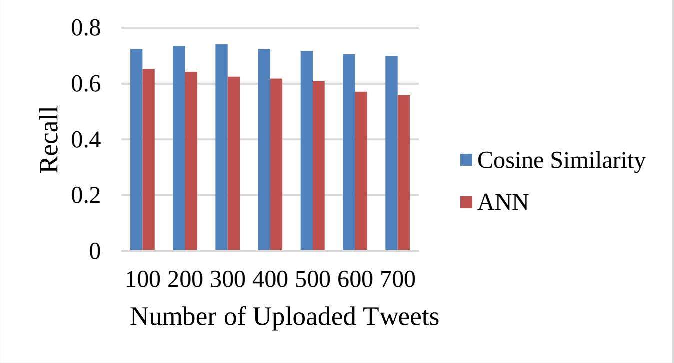


Figure 7: Recall

The F score represents the harmonic means of the precision and the recall values measured for the proposed work. The graph shows that for 100 Tweets F score is maximum and slightly decreases with the increase in the number of uploaded documents. The average F score value determined using cosine similarity and ANN are 0.7131 and 0.7347 respectively.

Table 5: F measure

Number of Uploaded Tweets	Cosine Similarity	ANN
100	0.738	0.770
200	0.740	0.762
300	0.734	0.749
400	0.715	0.739
500	0.706	0.731
600	0.693	0.701
700	0.666	0.691

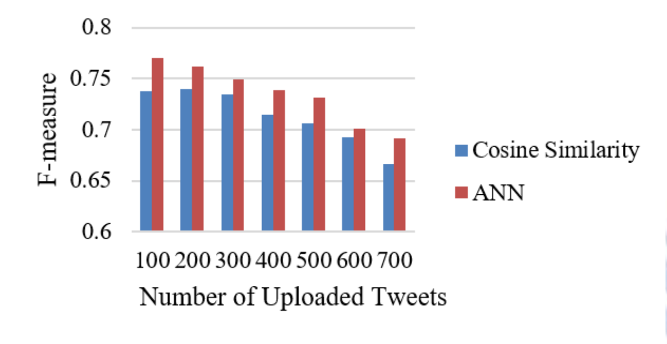


Figure 8: F-measure

The comparative analysis has been performed among the proposed work and the three different techniques used by researchers Al-Janabi et al. (2017) [14], Murugan et al. (2018) [15] and K Subba et al. (2019) [16]. Each researcher followed different techniques as depicted in Table 5. From the above graph shown in Figure 9, the precision value represented by the proposed work is better compared to other three techniques and the improvement of 3.96%, 6.34% and 1.67% in contrast to the Al-Janabi et al. (2017) [14], Murugan et al. (2018) [15] and K Subba et al. (2019) [16] work.

Table 6: Comparison parametric values

Techniques Used		Precision	Recall	F-measure
Proposed Work	Cosine Similarity with ANN	0.9252	0.6107	0.734
Al-Janabi et al. (2017) [14]	Random Forest	0.89	0.92	0.904
Murugan and Devi (2018) [15]	Decision tree, 'Particle Swarm Optimization (PSO) and Genetic algorithm (GA)'	0.87	0.15	0.84
K Subba and E. Srinivasa (2019) [16]	'principal component analysis' (PCA) and Decision Tree	0.91	0.93	0.919

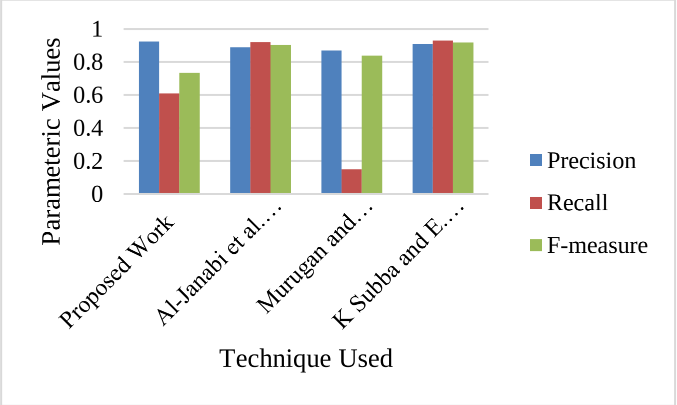


Figure 9: Comparative Analysis

V. CONCLUSION

In this research, the spam behaviour of a well known OSN site, Twitter has been studied by exploring the present issues faced by the researchers while designing a secure twitter system against spam. Here, a novel approach has been designed in which the Twitter data has been collected which is labelled and consisting of spam and normal Tweets. To reduce the dimension of the data, pre-processing using stop word removal and W2V approach has been applied. The, Cosine similarity has been applied to measure the similarity score and hence pass to the neural network for training. Using the above defined work, it has been concluded that the designed approach is simple and easy to design and also provide better precision, recall and F score compared to prior techniques used by [14], [15], and [16]. Also, an improvement of 3.96%, 6.34% and 1.67% has been examined in contrast to the [14], [15], and [16] work.

REFERENCES

[1] N. Kundu, and Y.K. Meena, "Detecting Suspicious Users in Social Networks Using Text Analysis," In Smart Systems and IoT: Innovations in Computing, Springer, Singapore, pp. 463-473, 2020.

[2] T. Satija, and N.Kar, "Detecting Malicious Twitter Bots Using Machine Learning," In International Conference on Computational Intelligence, Security and Internet of Things, Springer, Singapore, pp. 182-194, December 2019.

[3] A. T. Kabakus, and R.Kara, "TwitterSpamDetector: A Spam Detection Framework for Twitter," International Journal of Knowledge and Systems Science (IJKSS), vol.10, no.3, pp. 1-14, 2019.

[4] R. Paudel, P. Kandel, and W. Eberle, "Detecting Spam Tweets in Trending Topics Using Graph-Based Approach," In Proceedings of the Future Technologies Conference, Springer, Cham, pp. 526-546, October 2019.

[5] Chaudhary, D. S. . (2021). ECG Signal Analysis for Myocardial Disease Prediction by Classification with Feature Extraction Machine Learning Architectures. Research Journal of Computer Systems and Engineering, 2(1), 06:10. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/12.

[6] V. Mardi, A. Kini, V. M. Sukanya, and S. Rachana, "Text-Based Spam Tweets Detection Using Neural Networks," In Advances in Computing and Intelligent Systems, Springer, Singapore, pp. 401-408, 2020.

[7] H. Tajalizadeh, and R. Boostani, "A novel stream clustering framework for spam detection in twitter," IEEE Transactions on Computational Social Systems, vol. 6, no. 3, pp. 525-534, 2019.

[8] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," IEEE Transactions on Information Forensics and Security, vol. 8, no. 8, pp. 1280-1293, 2013.

[9] C. Chen, J. Zhang, Y. Xiang, and W. Zhou, "Asymmetric self-learning for tackling twitter spam drift," In 2015 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, pp. 208-213, April 2015.

[10] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, "Making the most of tweet-inherent features for social spam detection on Twitter," arXiv preprint arXiv, vol.1503, pp. 07405, 2015.

[11] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," In Proceedings of the 26th annual computer security applications conference, pp. 1-9, December 2010.

[12] J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender receiver relationship," in Proc. 14th Int. Conf. Recent Adv. Intrusion Detect., pp. 301–317, 2011.

[13] Lee, B.-K. . (2023). A Study on Image Quality Improvement for 3D Pagoda Restoration. International Journal of Intelligent Systems and Applications in Engineering, 11(4s), 150–156. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2582

[14] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving Twitter spammers," IEEE Trans. Inf. Forensics Sec., vol. 8, no. 8, pp. 1280–1293, Aug. 2013.

[15] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time url spam filtering service," In 2011 IEEE symposium on security and privacy, IEEE , pp. 447-462, May 2011.

[16] M. Al-Janabi, E. D. Quincey, and P. Andras, "Using supervised machine learning algorithms to detect suspicious URLs in online social networks," In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, pp. 1104-1111, July 2017.

[17] N. S. Murugan, and G. U. Devi, "Detecting streaming of Twitter spam using hybrid method," Wireless Personal Communications, vol. 103, no. 2, pp. 1353-1374, 2018.

[18] K. S. Reddy and E. S. Reddy, "Using Reduced Set of Features to Detect Spam in Twitter Data with Decision Tree and KNN Classifier Algorithms," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 9, pp. 6-12, 2019.

[19] A. Begum, and S. Badugu, "A Study of Malicious URL Detection Using Machine Learning and Heuristic Approaches," In Advances in Decision Sciences, Image Processing, Security and Computer Vision, Springer, Cham, pp. 587-597, 2020.

Dr. Partibha Yadav presently working as a faculty member in Department of Mathematics Indira Gandhi University, Meerpur, Rewari (India). She had graduated her Ph.D. in Computer Science from Indira Gandhi University, Meerpur, Rewari (India). She holds M.Tech. in Computer Engineering and have about 12 years of teaching experience. in reputed College/University. She has about a dozen of published research papers/attended conference.

Dr. Ajay Kumar is presently working a faculty member in Department of Computer Science & Engineering, Indira Gandhi University, Meerpur (Rewari). He earned his Master Degrees of MCA and M.Tech(CSE) from MDU Rohtak and also Doctorate in Computer Science & Engineering also he has UGC NET qualifed. He has 13 Years of teaching experience and served in multiple reputed Institution and University. His areas of Interests focused at IOT and Theory of Computation. He has successfully supervised more than 25 Major Projects of MCA at IGU Meerpur. He has been actively participating in various seminars, conference and also delivered many expert lectures. He has also published multiple research papers in national, international and reputed journals. He has received certification in Cloud Computing and Internet of Things from NPTEL.

Dr. Shivani is currently working as a faculty member in Department of Computer Science, Indira Gandhi University, Meerpur, Rewari. She earned her Doctorate degree in Computer Science from Glocal University, Saharanpur, U.P. in 2023 and completed her M.Tech in Computer Science from Vaish College of Engineering, Rohtak, Haryana in 2015. She has successfully supervised more than 20 Major Projects of MCA. She has two papers in her credit in International Journals. She has been actively participating in various seminars, conferences and workshops. She has research interest in neural network, data mining, data science, machine learning etc.

Dr. (Mrs.) Reena Hooda is presently working as Assistant Professor, Department of Computer Science & Engineering, Indira Gandhi University, Meerpur (Rewari). She earned her Master Degrees of MCA and MBA from MDU Rohtak and also Doctorate in Computer Science & Application from the MD University. She has also successfully completed the Data Science programs from Harvard University

DOI: https://doi.org/10.17762/ijritcc.v11i11s.8086
***Article Received:** 21 June 2023 **Revised:** 18 August 2023 **Accepted:** 02 September 2023*

(USA) and Machine Learning & Data Analysis programs from IBM. She has been into teaching since 14 Years and served many reputed Institutes. Her areas of Interests focused at Machine Learning, Data Science, and Data Warehousing & Mining. She has successfully supervised more than 50 Major Projects of MCA and currently guiding 4 Ph. D. scholars at IGU Meerpur. Dr. Reena has published more than thirty papers in the national and international journal and fourteen chapters in edited book.

Dr. Sudhir is currently working in Haryana Education Department as Lecturer of Computer Science. He obtained his engineering degree from Maharishi Dayanand University and he obtained his PhD in cloud computing and big data related research area. He has qualified both NET and GATE examination. He has helped many students in their projects. Apart from many prestigious institutions he taught as a faculty member in Indira Gandhi University, Meerpur, Rewari for about seven years. During his tenure, he made the student work on the latest technology like making live projects on Machine Learning and Internet of Things. Paved the way for children's projects on topics like Artificial Intelligence, IoT. He has received Elite category certification in Cloud Computing and Internet of Things from NPTEL.

Mrs. Pooja is currently working as a faculty member in Department of Computer Science, Indira Gandhi University, Meerpur, Rewari. She completed her M.Tech in Computer Science from Mody University Rajasthan in 2016. She has more than 6 Years of teaching experience and served in multiple reputed Institution and University. She has successfully supervised more than 20 Major Projects of B.Tech(CSE). She has also published multiple research papers in national, international and reputed journals. She has been actively participating in various seminars, conferences and workshops. She has research interest in image processing in AI etc.