Machine Learning based Optimization Scheme for Detection of Spam and Malware Propagation in Twitter

Savita Kumari Sheoran[1]
Associate Prof. in Computer Science and Engineering Dept
Indira Gandhi University, Meerpur
Rewari, India

Partibha Yadav[2]
Ph.D. Scholar in Computer Science and Engineering Dept
Indira Gandhi University, Meerpur
Rewari, India

***Abstract*—Social networking sites are new generation of web-services providing global community of users in an online environment. Twitter is one of such popular social networks having more than 152 million daily active users making a half billions of tweets per day. Owing to its immense popularity, the accounts of legitimate Twitter users are always at a risk from spammers and hackers. Spam and Malware are the most affecting threats reported on the Twitter platform. To preserve the privacy and ensure data safety for online Twitter community, it is necessary develop a framework to safeguard their accounts from such malicious attackers. Machine Learning is recently matured and widely used technique useful to prevent the propagation of such malicious activities in social media. However, the Machine Learning based techniques have yielded a promising result in filtering the undesired contents from the user tweets but its efficiency always remains restricted within the technological limits of the technique used. To devise a more efficient model to detect propagation of spam and malware in the Twitter, this research has proposed a Machine Learning based optimization scheme based on hybrid similarity (Cosine and Jaccard) measured in conjunction with Genetic Algorithm (GA) and Artificial Neural Network (ANN). The Cosine with Jaccard in hybridization has been applied on the Twitter dataset to identify the tweets containing spam and malware. In conjunction to it the GA has been used to enhance the training rate and reduce training error by automatically selecting the designed fitness function while the ANN was applied to classify malicious tweets from through voting rule. The simulation experiments were conducted to compute the precision rate, recall, F-measures. The results of Machine Learning based optimization scheme proposed in this research were compared with the existing state-of-arts techniques already available in this regime. The comparative study reveals that the model proposed in this research is faster and more precise then the existing models.**

Keywords*—Social networking sites, Twitter, spam, malware, Cosine similarity, Jaccard similarity, genetic algorithm, artificial neural network

I. INTRODUCTION

The last two decades have witnessed an unprecedented growth in social networking sites, where people share information with the other users through radio means without verifying their identity. Several social networking sites such as Twitter, Facebook, LinkedIn, Instagram and WhatsApp, etc. have emerged as a powerful tool to facilitate the users to share information in the form of audio, video, text and pictures. These social media platforms are governed by common instinct that all of them require creating an account using personal information and need access to data computing device before actual operation. The registered users can form a socio-digital network with the other users having similar interest and can share the contents of his choice in a manner prescribed the concerned website owner. Users use these sites for varied purposes including fun, entertainments, business, and advertisement etc. For instance, Twitter, a typical micro blogging social media platform, allows users to send message up to 140 characters, make comments, attach image and pdf documents. Apart from it blogs, PDF files, picture or videos and web page can be forwarded over the platform. On twitter the registered users enjoy unrestricted access to post, like, comments, reply and re-tweet while the unregistered users can only read the tweets. Twitter users are linked in the form of an exponential hierarchy where the user's tweets are available to followers in the form of public and protected tweets.

One of astonishing feature of Twitter which differentiates it from other social media platforms is that in Twitter the relationship between users and their followers is asymmetric while in other networks it follows a symmetric or cyclic pattern. In Twitter when a user gets followers the vice-versa is not always remains true and hence followers necessarily may not have to access all the tweets of their ancestral user [1]. The tweet post is twitter can be accessed with unrestricted right by immediate followers but not to the followers at a third level of followers in tree hierarchy. But the re-tweet from second level of users will be available to third level of users. The social media communities are more liberal on their community standards and generally groups are formed between those users, who are more active and share information frequently compared to less active users. The unsolicited users enter to this chain of active users to execute their malicious activities [2]. Hackers possess as original users can have easy access to the important personal information such as bank account or passwords, available in social media account or those available in computing device (computer or mobile phone) [3]. Recent studies reveal that Twitter has become most preferred destination for cyber criminals to perform multiple malicious activities including spam, phishing and malware [4]. One of the instances of such activity was reported in March 2010 when using festive-themed messages, dangerous malware was spread

in Twitter. Later in September 2010, the malware has affected the millions of Twitter users including British Prime Minister [5-6]. Fig. 1 depicts the social- criminal ecosystem of social network especially Twitter site. As revealed in the figure, a separate community is formed by criminal users using a unique user ID along with a supporter community encircled by green dotted contour, which supports those users outside the community of criminal accounts [7]. It reveals two types of relationship among the networked community *viz.* inner and outer relationships. Inner relationship reveals the interrelation among the criminal accounts connected through social means while the outer relationship represents the interaction between the criminal accounts and his supporters, who maintain a close friendship with the criminal accounts. To propagate the threats identified in this research *viz.* malware and spam; the hackers post malicious links to unsolicited users for attracting user traffics.

The subsequent sub-sections will elaborate these both types of malicious activities for formulating further research work.

A. Spam

There is a general perception that spam mostly found in e-mails but there are instances where social networking sites frequently suffers from malicious software. Spam harms the users through various modes such as by sending undesired information in the form of advertisement or by sending messages continuously to the same e-mail or social media ID. The existing research detects the by analysing the features of the data. Beutel et al. (2013) has detected the spam by analysing the relationship between the users, social media pages and the time of instant at which the edge has been created in the social graph [8]. Another research performed by Ahmed et al. (2012) has used graph-based technique to show the relationship between the social nodes and their communication by edge of the graph [9]. The weight of edge represents the real and fake users' interactions in the form of shared URLs, pages, active friends. Here, spam detection has been performed using optimization-based machine learning approach. Sharma et al. (2014) have used Machine Learning to classify text containing spam as enunciated in the workflow [10].

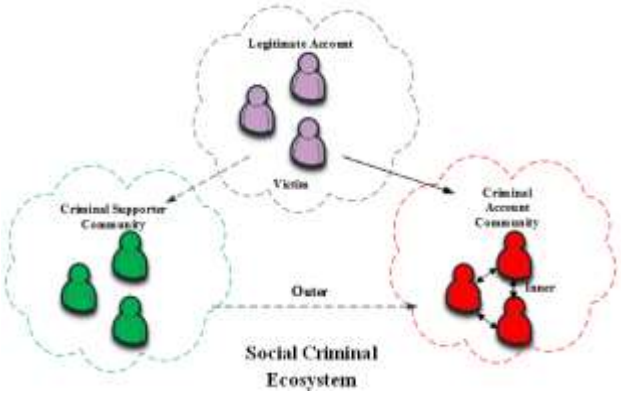

Fig. 1. Structure of the Social Criminal Ecosystem [7].

B. Malware

Malware is a type of code implanted into the network with the intention to affect the information of the legitimate users. It is injurious issue to a computer user and need to be resolved to safeguard his private and business rights. Recently there are examples where spammers have relied on outmoded social networks impersonated as e-mail to steal the information of normal users by inserting harmful worms. The Pay-Per Install (PPI) is the most amazing institute that spread malicious activities targeting the financial institutions and other websites like Facebook and Twitter. The modus operandi of Malware in affecting the information was characterized by Sanzgiri et al. (2013) [11].

This research intended to design a model to detect spam as well as malwares propagating in twitter contents through an approach based on finding similarity among the extracted features such as crime related keywords and normal keywords and the URL features. Based on the extracted features, the features are optimized using a novel fitness function of Genetic Algorithm (GA). Based on these optimized features the machine learning classifier such as ANN is trained based on the optimized features which enhance the accuracy of detection method [12]. The subsequent sections of this paper are arranged as follow: Section 2, describes the work accomplished by researcher community engaged in the regime of social network security and malicious attacks. The step by step description of the work proposed in this piece of research is presented in Section 3. The results obtained over simulation experiments and the examined parameters are discussed in Section 4. A conclusive discussion is carried out in Section 5, followed by the references consulted there in the paper.

II. Related Work

Social networking sites are naive form of socialization and hence facing out of security issues. A number of researchers have studied detection and protection of social networks against spam. Blanzieri et al. (2008) have surveyed the commonly known features which were further used to enlist the unsolicited methods of spam detection [13]. Further Sahamiet al. (1998) has deployed the unsolicited techniques such as content filtering for the same purpose [14]. In social media applications such as Twitter and Facebook the content-based methods are hardly effective because the spam contains only a few words along with the URLs. Therefore, some of the researchers have used URL blacklisting approach in order to filter the spam but this technique is not performing as per the user requirement as well as take large processing time as summarized by Grier et al. (2010) [15]. Song et al. (2011) have used relation features (distance and interconnection) among the transmitter and receiver social user for the detection of the spam in data. A list of spam and non-spam data has been created and then trained the classifier based on the extracted features. The results indicate that most of the spam has been generated by the account rather than receiver [16]. Lin et al. (2017) have presented a machine learning based approach to detect the spam based on ground truth value and provided

satisfactory performance. Also, the designed spam detection twitter model has been analysed for scalability and the performance has been measured in terms of true positive, False positive, F-score and accuracy of the system using different data size with small processing time [17].Gupta et al. (2018) have presented a spam detection framework through which the spam is identified based on user based and tweets' text-based features collectively. The use of text-based tweet features allows users to detect the spam tweet if the unsolicited user has created new account. The work has been verified based on four different classifiers such as Support Vector Machine (SVM), Neural Network (NN), Random Forest (RF) and Gradient Boosting and Neural Network based approach has achieved highest accuracy of 91.65 among all the methods [18]. Hanif et al. (2018) have introduced additional features to measure the countermeasure in the presence of spam in Twitter site. Hanif et al. (2018) have detected the spam and malware using four machine learning techniques such as RF, SVM, *K*-Nearest Neighbour (KNN) and Multi-Layer Perceptron (MLP). They have performed a series of experiments using two simulation tools such as WEKA and RapidMiner and better results in terms of detection accuracy of 95.44% has been obtained using RF as classifier on RapidMiner tool [19]. Hai and Hwang (2018) have used deep learning as a classification approach for the detection of malware based on their malicious activities. The detection accuracy of 98.75 % has been obtained, which is quite higher as compared to the other existing techniques [20].Kaur and Sabharwal (2018) have used feed forward neural network as a classifier, which was trained based on the extracted features (+ve and -ve) in social networks. To resolve the complexity of extracted features genetic optimization has been used as an optimization approach [21].

However, a lot of researches are available in literature which studied the issue in fragmented way but the technique offering a single framework to detect spam and malware affected tweets by utilizing a minimal number of feature set is still undiscovered. To address this issue this research intends to develop a novel model to filter out the spam and malware in the Twitter using machine learning approach.

III. Proposed Work

In this research, we have applied a hybridized approach that includes GA with ANN technique to detect spammed malware. The feature of tweets has been refined and optimized based on the fitness function of GA and used dynamically to trained ANN structure. In traditional methods, the features are refined using pre-processing technique and then applied to the whole dataset. ANN is a better approach for training the data in the sense that it dynamically selects the features for the individual user data instead of applying the same features to all users. It is effective strategy for the reason that each user possesses its own characteristic features and hence need to be segregated from each other. A study on Twitter reveals that fresh accounts and Spam accounts have higher link share than that of average link shares in normal accounts. Apart from its various studies yield that spam users share more images from news web sites and roll out lucrative advertisements to lure the innocent users. Therefore, mere filtering the users sharing more images will not be sufficient to detect the spam rather filtering the users sharing more images from new websites and issuing lucrative advertisements will serve the purpose better. This study more relied on the feature of individual user group identity and classifies each user based on the URL sharing. The identified grouped in each URL's are trained using ANN approach in addition to GA scheme. Fig. 2, presents a secure framework for detecting Spam and Malware in the Twitter network.

A. Upload Data

The data used for this research were initially obtained Kaggle database and only the data related to spam, malware and normal tweets have been processed for further use in study [22]. The dataset contains total of 200K tweets along with their URL. The study was initiated on the hypothesis that all tweets contain URLs with the aim to attract social users towards malicious sites such as spam and malware downloading. The hypothesis was contradicted to obtain the URLs with spam and malware and rest were discarded.

B. Stop Word Removal

The stop word is removed by comparing each row contains in the dataset with the stored stop word list available on GitHub Gist [23]. A few stop words used in the proposed work are listed in Table I. Initially, collected data is uploaded and compared with the list of stop words in the database.

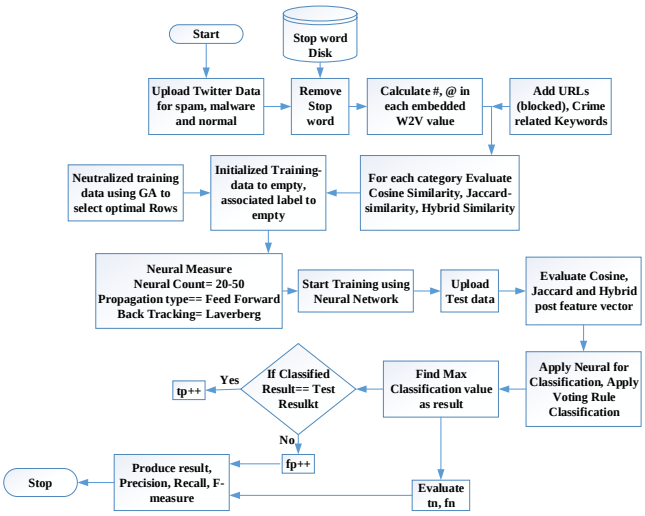

Fig. 2. Workflow Diagram of Proposed Model.

TABLE I. Stop Word List

"An"	"If"	"During"	"Before"	"After"	"Above"
"And"	"Or"	"Below"	"To"	"From"	"Up"
"But"	"Because"	"Down"	"In"	"Is"	"It"
"While"	"Until"	"Else"	"Than"	"Too"	"Very"
"Off"	"Of"	"Own"	"Can"	"Off"	"Will"
"The"	"At"	"Just"	"Don"	"Should"	"Now"

If the words in the uploaded data are matched with the database words, then these words are removed to obtain the data containing only the meaningful informative words. The mentioned algorithm is used to remove the stop word from database.

Algorithm 1: SWFD = Stop Word Removal (TUD)

Where, TUD → Twitter User Data (Individual user-wise)
SWFD → Stop Words Free Data sorted from main database

Start
Load Stop Word Dataset
Set, Count = 1
For I = 1 → All TUD
  For J = 1 → All SW
  If TUD (I, J) = SW (I, J)
  SWFD (Count) = TUD (I, J)
  Count = Count+1
  Else
  SWFD = ' '
  End – If
  End – For
End – For
Return: SWFD
End

C. Mention Ratio / URL as Content-Based Features

In this research mention ratio such as @ and # have been used as content features along with the URL. As these are the essential features used by the twitter and also used by the malicious users to misguide the normal tweet users. Therefore, it is necessary to remove these symbols from the tweet.

1) Mention Ratio: Generally, Twitter users are tagged using the '@' special character. Spammers and malware activists can also use the same special characters to trap the legitimate users. The malicious account holders entice normal users to attach with them. Equation (1) below, is used to calculate the mention ratio for each special characters.

$$Mention\ Ratio = \frac{Number\ of\ @\ present\ in\ the\ tweet}{Total\ tweets\ posted\ by\ the\ user} \quad (1)$$

2) URL Ration: Social media users generally share their thoughts and also give suggestion through tweets. The tweets posted by the sender may include URLs having a link to source pages encompassing complete information. The clever user intentionally enters a large number of URLs in their tweets to trap the legitimate users as their soft target. The URL ratio can be calculated using equation (2).

$$URL\ Ratio = \frac{Number\ of\ URL\ present\ in\ the\ tweets}{Total\ tweets\ posted\ by\ the\ Users} \quad (2)$$

3) Word to vector: The removal of stop words is followed by calculation of special characters (# and @) in the uploaded tweets, which can be applied on word to vector method. This scheme converts the text into its corresponding weighted value like as:

{-0.09450 0.16788 -0.14402 -0.0251 0.11355 -0.11794 -0.13871 -0.01607 0.1555 0.11695 0.05452 0.0936 0.08511 0.00671 -0.11653 -0.13014 0.12626 0.10248 -0.035507 -0.1523 -0.08457 0.089321 -0.01771 -0.07837 0.16123 -0.10844 -0.10118 0.03016 0.05699 0.03763 0.63156 0.06131 0.19388 -0.05652 0.1217 0.15755 0.01353 0.33352 -0.0223 -0.10877 0.11583 -0.07015 0.03653 0.05292 -0.0074 0.0242 0.08846 0.14987 0.12804 0.18679}.

The main purpose of word embedding is to study the vector representation obtained after word to vector method. One of the most commonly used word embedding method is word to vector, which maximize the probability of word condition, which is fitted in the window 'W'. After this, the crime related words appear in the URL are blocked.

On the basis of calculated value of 'W', the relationship between any two words such as W_i, W_j can be measured using hybrid similarity measures, which is a combination of Cosine Similarity and Jaccard Similarity index [24]. The similarity between tweets, which is being calculated using Cosine Similarity, is calculated using equation (3).

$$Sim_{i,j} = \frac{W_i \times W_j}{\|W_i\| \|W_j\|} \quad (3)$$

D. Cosine Similarity

Cosine similarity is a similarity analysis approach used to measure similarity score between two non-zero vectors. It is measured by the cosine of the angle between the two vectors and determines whether the two vectors point in approximately the same direction. It is often used in text analysis to measure similarity among documents. This method is used to determine the similarity and is a traditional approach, which is used in integration with the Term Frequency (TF). The text obtained after the filtering of crime related words appear in the tweet, cosine similarity is applied between the two vectors and then the multiplication of these two vectors value is being compared. Fig. 3 show the Cosine Similarity used in comparing tweets throughout this study.

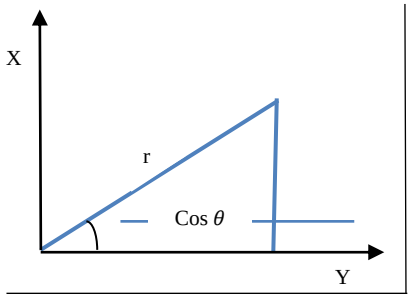

Fig. 3. Cosine Similarity.

Two tweets are declared similar if Cosine Similarity value is approaches to unity. The value of this factor approaches unity for 0^o and for other angles it is less than it [25]. The designed algorithm for Cosine Similarity is presented below:

Algorithm 2: Cosine Similarity

Required Input:	Data ← Raw data in which similarity needed
Obtained Output:	Sim_{Cos} ← Cosine similarity between data

Start
To store similarity create an empty array, Sim_{Cos}= []
Sim-count = 0
For m = 1 → Length (Data)
Current_Data = Data (m)
For n = m+1 → Length (Data)
Calculate the Cosine Similarity using given equation
L = |Cos (Current_Data) - Cos (Data (n))|
Sim_{Cos} [sim_count, 1] = Current Data
Sim_{Cos} [sim_count, 2]= Data(n)
Sim_{Cos} [sim_count, 3]=L
Incremental array →Sim-count = Sim-count + 1
End – For
End – For
Return: Sim_{Cos} as final output of cosine similarity between data
End – Function

E. Jaccard Similarity

Jaccard similarity is used to determine the similarity as well as the distinction among the documents based upon the attributes. Its value lies between 0 to 100 percentages. Higher percentage value represent more similar is the data while lower value infers least similarity. An effort has also made to determine the similarity using Jaccard Similarity with relationship between two tweets by calculating Jaccard Coefficient, basically utilized to compare data based on similarity, dissimilarity and distance bases [26]. The output obtained using Jaccard similarity is the rate of number of tweet features that are most common to the entire text with respect to the number of features present in the entire tweet. The measured similarity calculated using Jaccard similarity is given by equation (4).

$$J\left(W_i, W_j\right) = \frac{|W_i \cap W_j|}{|W_i \cup W_j|} \quad (4)$$

Following algorithm is implemented for Jaccard Similarity:

Algorithm 3: Jaccard Similarity

Required Input:	Data ← Raw data in which similarity needed
Obtained Output:	Sim_{Jac} ← Cosine similarity between data

Start
Create an empty array to store similarity, Sim_{Jac} = []
Sim-count = 0
For m = 1 → Length (Data)
Current_Data = Data (m)
For n = m+1 → Length (Data)
Union = (Cos (Current_Data) U Cos (Data (n)))
Intersection = (Cos (Current_Data) ∩ Cos (Data (n)))
Sim_{Jac} (sim_count) = $\frac{Count\ (Union)}{Count\ (Intersection)}$
Incremental array →Sim-count = Sim-count + 1
End – For
End – For
Return:Sim_{Jac} as Jaccard Similarity between data
End – Function

F. Genetic Algorithm (GA)

GA has been used as a feature selection algorithm in order to select the row features of the tweets obtained after hybridizing Cosine and Jaccard Similarity Index. Feature selection is one of the essential tasks, that helps to enhance the training accuracy of the classification algorithm such as Neural network is used to train the system based upon the optimized features obtained as per the designed fitness function as denoted by equation (5).

$$F(f) = \begin{cases} 1\ (Selected)\ if\ (1 - f_s);\ Max\ among\ all\ vectors < f_t \\ 0\ (Not\ Selected)\ Otherwise \end{cases} \quad (5)$$

Where,

F_t Generated mutation error

f_s : Current feature in FD

f_t: Threshold feature and it is the average of all FD

The implementation in GA can be accomplished in three steps depicted in the Fig. 4:

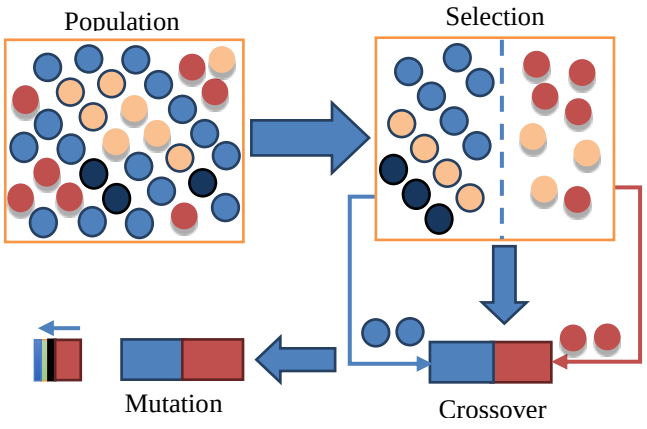

Fig. 4. Steps in GA Implementation Process.

GA is a basic heuristic algorithm that works on Darwin's theory of evolution and is also named as Evolutionary Algorithm that finds the best solution based on the natural selection and crossover as shown in Fig. 4. Genetic algorithms randomly generate a set of populations. A distinct gene is comprised by each individual and hence responsible for different solution to a particular problem, which is again encoded by the chromosomes. To solve the problem, a problem specific objective function is designed. GA mainly included three operators *viz.* (i) Selection (ii) Crossover and (iii) Mutation. Selection is used to choose individuals from the present generation, which is later used for next generation. At this stage the best one with high fitness values are selected. In chromosomes, it is responsible to suggest parents (two best chromosomes that are responsible for best generation). This process is repeated until the desired solution is obtained. The workflow of GA is written in algorithmic form as below:

Algorithm 4: Features Selection using GA

Required Input:	Feature Data ← Extracted feature from used Dataset
	Fitness Function ←Designed fitness function for feature selection
Obtained Output:	OFD←Optimized Feature Data

Start Feature Selection
Load Dataset, Feature Data (FD) = Load feature sets
To optimized the FD, Genetic Algorithm (GA) is used
Set up basic operators and parameters of GA:
Population Size (P) – Based on the number of properties
CO – Crossover Operators
MO – Mutation Operators
OFD – Optimized Feature Data
Calculate fitness function [F(f)]with usual terms

$$\mathbf{F(f)} = \begin{cases} \mathbf{1}\ (Selected) if (\mathbf{1} - \mathbf{f_s});\ Maxamongallvectors < f_t \\ \mathbf{0}\ (NotSelected) Otherwise \end{cases}$$

Set,Optimized Feature Data, OFD = []
For i in rang of R
F_s = FD (i) = $Selected_{Feature}$
$F_t = Threshold_{Feature} = \sum_{i=1}^{R} FD(i)$
$F(f) = Fit\ Fun\ (F_s, F_t)$
Nvar = Number of variables
$Best_{Prop}$ = OFD = GA (F(f), T, Nvar, Set up of GA)
End - For
Return: OFD as an Optimized Feature Data
End – Function

G. Artificial Neural Network (ANN)

After optimizing the features based on the fitness function of GA as according to equation (5), these features are used to train Neural Network as a classification algorithm. ANN is designed to work in the same way as that of human brain. Its working is inspired by the biological nature of cell known as Neurons or sometimes knows as nodes. The structure of ANN with 'N' number of data input and single output is shown in Fig. 5 while Fig. 6 shows the examined Mean Square Error (MSE) value during the training process of a spam and malware detection based social media system.

The figure shows that the desired value has been obtained after passing the 20 number of neurons to the hidden layer of ANN structure.

Fig. 5. Trained ANN Structure with MSE Value.

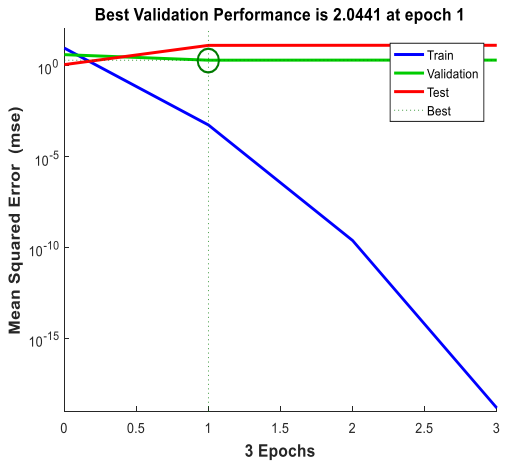

Fig. 6. Mean Square Error for Epoch.

Also, The MSE value examined during the training process ID indicated by the blue line, which is approaches to zero at 3rd iteration. The designed algorithm for ANN is represented as follows:

Algorithm 5: Training using ANN

Required Input:	OFD←Training Data as an optimized feature data
	C←Target/Category in terms of spam, malware and normal data
	N←Number of Neurons
Obtained Output:	Net ← Trained structure

Start Detection
Load Training Data, T-Data **= OFD**
Declare the initial parameters of ANN
- Epochs Counts: E
- Neurons Counts : N
- Performance Parameters: MSE, Gradient, Mutation and Validation
- Techniques Applied: Levenberg Marquardt Algorithm
- Data Division Strategy: Random

For i = 1 → T-Data
If T belongs to spam
Group (1) = Features (OFD)
Else if T belongs to malware
Group (2) = Features (OFD)
Else // Normal Case
Group (3) = Features (OFD)
End – If
End – For
Implement the ANN through Training data and Group
Net = Newff $(T - Data, Group, N)$
Setting training parameters as per the requirements and accomplish the train task
Net = Train (Net, T-Data, Group)
Return: Net value according to trained structure
End – Function

The testing of spam and malware detection social system has been performed by uploading the tweets as test data and then measure the similarity among the uploaded documents using Cosine with Jaccard as similarity measure. The data obtained are compared with the data stored into the ANN database by applying the voting rule as a cross-validation scheme. Here the voting classifier is used in addition to ANN classifier. If maximum value has been obtained, then calculate True Negative (T_n) and False Negative (F_n) values for the uploaded data. In case, if classified results are equal to test results then True Positive (T_p) and False Positive (F_p) has been calculated. Subsequent section 4 of this paper presents and discuss the results obtained for the parameters (T_n), (F_n), (T_p) and (F_p) in term of precision, recall, F-measure.

IV. Results and Discussion

The performance analysis of proposed model was carried out through simulation experiments conducted using standard settings considering optimization, classification with similarity measurement tools. A total of N-700 tweeter data were analysed over Simulink and Natural Language toolkit for parametric analysis and stop word removal respectively. The performance has been measured for three parameters precision, recall, F-measure using standard equations (6), (7) and (8) respectively. Where Precession signify the instances of correctness in the experiment, Recall signify the measure of correct hit and F-measure score is related to accuracy or correct prediction per unit of input.

$$Precision = \frac{T_p}{T_p + F_p} \quad (6)$$

$$Recall = \frac{T_p}{T_p + F_n} \quad (7)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

Where

T_p = Number of tweets that are actually spam/ malwares and also predicted as malicious.

F_n = Number of tweets that are being predicted as real but are spam and contains malwares.

F_p = Number of tweets that is actually real but predicted as affected one (Spam/ malwares).

T_n = Number of appropriately predicted real tweets.

The variation of precision values with number of tweets uploaded for various techniques *viz.* Cosine, Jaccard, Hybrid and GA with ANN approach are presented in Fig. 7. Figure illustrates that proposed work implementing GA with ANN in combination with hybrid similarity measure have highest Precession values for any number of tweets. The average precision computed for cosine, Jaccard, hybrid and GA with ANN approach are 0.746, 0.805, 0.885 and 0.963 respectively which reveals that the tweet that are filtered as a sub part of spam or malware for the tested dataset is maximum for GA with ANN approach (proposed in this research).

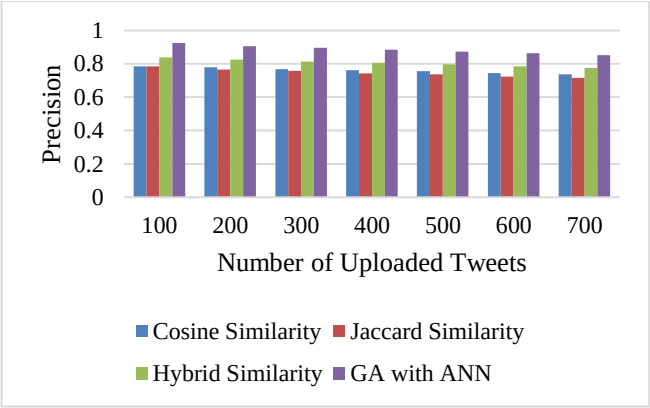

Fig. 7. Precision versus Number of uploaded Tweets (N=700).

The recall parameter represents the rate of tweets that are being posted by genuine user and have been predicted as spam or malware by the user accurately. The examined value of Recall for the uploaded tweet in the range from 100 to 700 is shown in Fig. 8. The average recall rate examined for the Cosine similarity, Jaccard Similarity, hybrid similarity and GA with ANN are 0.694, 0.785, 0.864, and 0.894 respectively which reveals that the tweet that are filtered owing to genuineness against spam or malware for the tested dataset is maximum for GA with ANN approach (proposed in this research).

To represent the arithmetic means of precision and recall of the examined values F-measure is illustrated in Fig. 9. F-measure basically envisages the accuracy of a model. The examined average values of F-measure for four different schemes *viz.* Cosine similarity, Jaccard Similarity, hybrid similarity and GA with ANN are 0.719, 0.822, 0.874, and 0.927 respectively which again reveals that the tweet that are filtered per unit of input due to spam or malware for the tested dataset is maximum for GA with ANN approach (proposed in this research).

A comparison of average values of the parameters under study for proposed model with the existing state-of arts in the area of research viz. K. Subba and E. Srinivasa (2019) [27] and Murugan and Devi (2018) [28] is tabulated in Table II and represented graphically in Fig. 10.

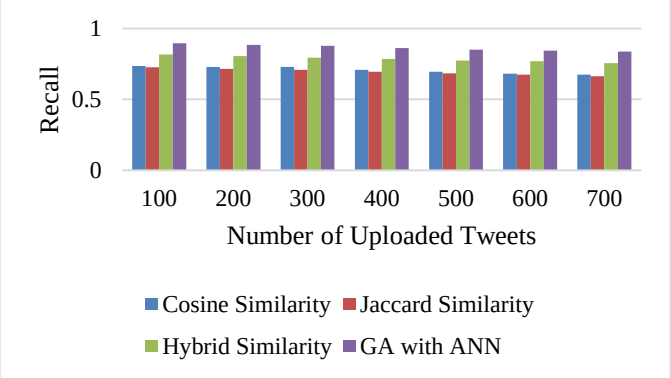

Fig. 8. Recall versus Number of uploaded Tweets (N=700).

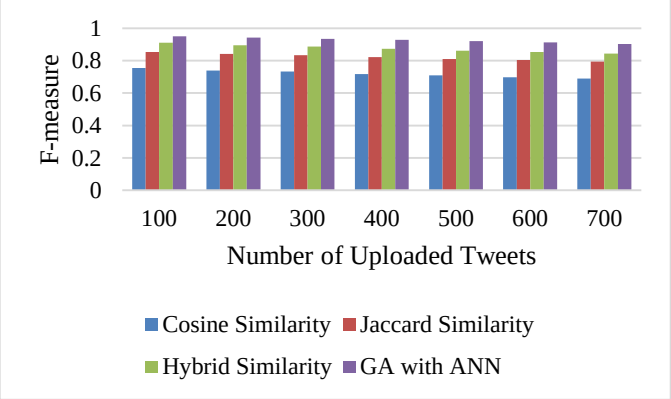

Fig. 9. F-Measure versus Number of uploaded Tweets (N=700).

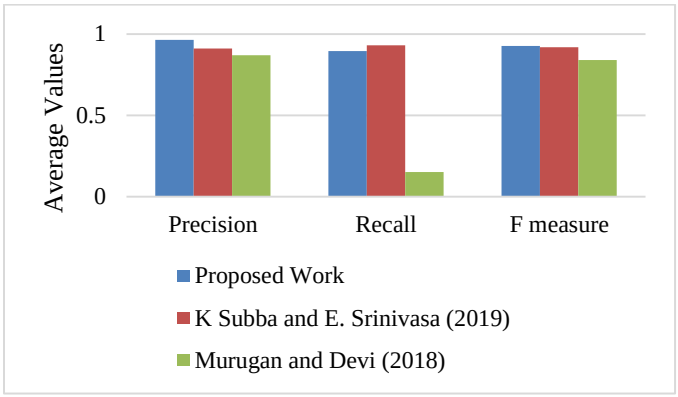

Fig. 10. Comparison of uploaded Tweets (N=700).

TABLE II. COMPARISON OF PARAMETERS

Parameters	Proposed Model	K. Subba and E. Srinivasa (2019) [27]	Murugan and Devi (2018) [28]
Precision	0.963	0.91	0.87
Recall	0.898	0.93	0.15
F-measure	0.927	0.919	084

A detailed look at the available literature reveals that the models established by K. Subba and E. Srinivasa (2019) [25] and Murugan and Devi (2018) [28] are state-of-arts exiting models having best performance so far. The comparison of performance of the existing state-of-arts with the model proposed in this research is shown in Fig. 10. Above results reveals that model proposed by us using the hybrid GA with ANN approach outperform the Murugan and Devi (2018) on all three examined parameter while it outperform the K. Subba and E. Srinivasa (2019) Model on the two parameters *viz.* Precision and F-measure and almost lessen Recall value. For quantitative purpose the precession in filtering the spam and malware for the proposed model is improved by 5.82 % and 10.69 % respectively from K. Subba and E. Srinivasa (2019) [25] and Murugan and Devi (2018) [28] models. Therefore, overall performance of the model proposed in this research is better than the existing models.

V. CONCLUSION

Presently Social networking sites are the most popular mode of network formation for the purpose of exchanging the information, advertise and the business purpose. Owing to their global popularity the Social Networking sites are at a great risk of having been used to misguide the genuine users from malicious activities of spammers and malwares. Therefore, to ensuring the data safety and privacy of the social media user is a need hour. In literature the measurement of Cosine Similarity, Jaccard Similarly and Hybrid Similarity has been carried out to evaluate the Precession, Recall and F-measure values for decide the effectiveness of a model in preventing the spam and malware but improving the performance is always remained an open challenge before research community working in this regime In this paper, we have designed a secure threat prevention (spam and malware) system for Twitter site.

We have used Machine Learning approaches such as GA and ANN in hybridization of existing models involving measurement of Cosine and Jaccard similarity. In our model novel GA approach and has been used for classification and ANN with voting algorithm is used for cross validation purpose. The simulation study carries out on N=700 tweets, reveals that average precision, recall and f-measure of 0.963, 0.894 and 0.927 has been achieved which is 5.82 % and 10.69 % higher than the other two models *viz.* K. Subba and E. Srinivasa (2019) and Murugan and Devi (2018); used as standard reference in research. This study reveals that the Machine Learning is an effective tool for prevention of legitimate users against attack of spam and malwares. Here in this research we have applied GA and MLL for filtering some stop words from Twitter and observed a promising result. In future we are planning to further investigate the similar issues using deep learning approach with complete text analysis with NLP in Twitter and social media sites. Such study may yield more effective results for preventing malicious attack of legitimate social media users.

REFERENCES

[1] A. Sanzgiri, A. Hughes, and S. Upadhyaya, "Analysis of malware propagation in Twitter," IEEE 32nd International Symposium on Reliable Distributed Systems. pp. 195-204, 2013.

[2] G. Fei, H. Li, and B. Liu., "Opinion Spam Detection in Social Networks," In Sentiment Analysis in Social Network, pp. 141-156, 2017.

[3] D. Niranjan Koggalahewa, Y. Xu, and E. Foo, "Spam Detection in Social Networks based on Peer Acceptance," In Proceedings of the Australasian Computer Science Week Multiconference, pp. 1-7, February 2020.

[4] B. A. Kamoru, A. Jaafar, M. B. Jabar, and M. A. Murad, "A mapping study to investigate spam detection on social networks," Int J Appl Inform Syst, vol. 11, no. 11, pp. 16-31, 2017.

[5] M. R. Faghani and H. Saidi, "Malware propagation in Online Social Networks," IEEE 4th International Conference on Malicious and Unwanted Software (MALWARE,) pp. 8-14, 2009.

[6] K. Nagaramani, K. Vandanarao, and B. Mamatha, "Machine Learning Algorithms for Spam Detection in Social Networks," Asian Journal of Computer Science and Technology, 8(S3), pp. 41-44, 2019.

[7] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, "Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter," In Proceedings of the 21st international conference on World Wide Web, pp. 71-8, April 2012.

[8] A. Beutel, W. Xu, V, Guruswami, C. Palow and C. Faloutsos, "Copycatch: stopping group attacks by spotting lockstep behavior in social networks," In Proceedings of the 22nd international conference on World Wide Web, pp. 119-130, May 2013.

[9] F. Ahmed and M. Abulaish, "An MCL-based approach for spam profile detection in online social networks," In 2012 IEEE 11th international conference on trust, security and privacy in computing and communications, pp. 602-608, June 2012.

[10] N. Sharma,and A.Verma, "Survey on Text Classification (Spam) Using Machine Learning," (IJCSIT) International Journal of Computer Science and Information Technologies, 5(4), pp. 5098-5102, 2014.

[11] A.Sanzgiri, A. Hughes, and S. Upadhyaya, "Analysis of malware propagation in Twitter," IEEE 32nd International Symposium on Reliable Distributed Systems, pp. 195-204, September 2013.

[12] F.J. Alqatawna, A. Madain, Z.A Ala'M and R. Al-Sayyed, " Online social networks security: Threats, attacks, and future directions," In Social Media Shaping e-Publishing and Academia, pp. 121-132, Springer, Cham, 2017.

[13] E. Blanzieri, and A. Bryl, "A survey of learning-based techniques of email spam filtering,"Artificial Intelligence Review, 29(1), pp. 63-92, 2008.

[14] M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz,"A Bayesian approach to filtering junk e-mail," In Learning for Text Categorization Papers from the 1998 workshop,Vol. 62, pp. 98-105, July 1998.

[15] C. Grier, K. Thomas, V. Paxson, and M. Zhang, "@ spam: the underground on 140 characters or less," In Proceedings of the 17th ACM conference on Computer and communications security, pp. 27-37, October 2010.

[16] J. Song, S. Lee, and J. Kim, "Spam filtering in twitter using sender-receiver relationship," In International workshop on recent advances in intrusion detection, pp. 301-317,Springer, Berlin, Heidelberg, September 2011.

[17] G. Lin, N. Sun, S. Nepal, J. Zhang, Y. Xiang, and H. Hassan, 'Statistical twitter spam detection demystified: performance, stability and scalability," IEEE access, 5, pp. 11142-11154, 2017.

[18] H.Gupta, S.M. Jamal, S. Madisetty, and S.M. Desarkar, "A framework for real-time spam detection in Twitter," 10th International Conference on Communication Systems & Networks (COMSNETS), pp. 380-383, IEEE, January 2018.

[19] M. H. M. Hanif, S. K. Adewole, B. N. Anuar, and A. Kamsin, "Performance Evaluation of Machine Learning Algorithms for Spam Profile Detection on Twitter Using WEKA and RapidMiner," Advanced Science Letters, 24(2), pp. 1043-1046, 2018.

[20] T.Q. Hai, and O.S. Hwang, "An efficient classification of malware behavior using deep neural network," Journal of Intelligent & Fuzzy Systems, 35(6), pp. 5801-5814, 2018.

[21] J. Kaur, and M. Sabharwal, "Spam detection in online social networks using feed forward neural network," In RSRI conference on recent trends in science and engineering 2, pp. 69-78, 2018.

[22] https://www.kaggle.com/uciml/sms-spam-collection-dataset accessed on January 01, 2020.

[23] https://gist.github.com/sebleier/554280 accessed on January 02, 2020.

[24] X. Yang, C. Macdonald, and I.. Ounis, "Using word embeddings in twitter election classification," Information Retrieval Journal, 21(2-3), pp. 183-207, 2018.

[25] R.A. Lahitani, E.A. Permanasari, and A.N. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 4th International Conference on Cyber and IT Service Managemen, pp. 1-6. IEEE, April 2016.

[26] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of Jaccard coefficient for keywords similarity," In Proceedings of the international multiconference of engineers and computer scientists 1 (6), pp. 380-384, March 2013.

[27] K Subba Reddy, E. Srinivasa 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) ISSN: 2278-3075, 8 (9), pp 6-12, 2019.

[28] S.N. Murugan, and U.G Devi, "Detecting streaming of Twitter spam using hybrid method," Wireless Personal Communications, 103(2), pp. 1353-1374, 2018.