International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

An Efficient Classifier for Spam Detection in Social Network

Amit Pratap Singh, Maitreyee Dutta

Abstract *The way to stay connects users with their friends provided by the Social networks. The popularity of social networks has been increasing day-by-day and enables users to collect huge volume of information about their friends. Various social networking sites are available among them, Twitter is fastest growing website. Due to its popularity various spammers are attracted towards it to utilized large amounts of spam to modify authorised users accounts. In this paper, the "Spam detection system in social sites" is developed for detection of spammer through the technique of machine learning. The work is initiated towards collection of data from H-Spam 14 site and then applied pre-processing mechanism such as conversion of data into lowercase and, removal of stop words etc. After this phase, the pre-processed data comes into the phase of feature extraction, which involves process of tokenization that used to split the entire sentences into a word-group and so the best features has been extracted from the raw data. To select the optimized value from extracted set of features, the optimization algorithm, Artificial Bee Colony (ABC) has utilized here to obtain the optimal sets of feature from spam along with non-spam data. The next process has to be done through Artificial Neural Network (ANN) to differentiate the spam and non-spam data. In the final process, the parameters for performance measure and compare the proposed and existing work to check the improvement in proposed work. In this proposed work the spam detection system gained higher accuracy, precision, recall and F-measure as compare to the previously utilized classifiers named as naïve Bayes and Support vector machine (SVM).*

Keywords: Spam detection, Twitter, ABC , AN, tokenization.

I. INTRODUCTION

Various online social networking sites such as Facebook, LinkedIn, and Twitter using these sites individuals are allows to connect new persons, stay in-touch with friends, make professional group, and many more[9]. As per the online report of social networking sites, among all Twitter is the fastest growing. Twitter provides most important service named as micro-blogging by which users can transfer messages among other users termed as tweets. The limitation of single tweets is 140 characters and tweets included only text and HTTP connections.

Revised Manuscript Received on November 30, 2019.
***** Correspondence Author
Amit Pratap Singh*, Asst. Prof. Dept. of Computer Science and Engineering, KEC 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

This exchange nature of tweets makes persons able to communicate and keep in touch with friends and colleagues [1]. The MICRO-BLOGGING is not limited to attract only legal users but also spammers. Now a day's spam becomes a growing issue of online social networks such as Twitter. Grier et al. have reported that 0.13% of the spam texts are posted on Twitter and two times as compare to the spam of email. According to the click rate on Twitter has increased rapidly, it becomes the interactive platform for the spammers. This increasing spam measures has adversary affected the experience and behaviour of users included as reviewing and recommending [2]. To take the actions against increasing threads of spammers, to outline the spam Twitter utilized various ways. The spam on Twitter can be reported easily through clicking on report as a spam link of website. The produces report is evaluated by the team of Twitter and if spam account is found then it will be blocked. There is also a different way is available to report about spam on Twitter is to send a tweet to the "@spam @username" format in which spam account are mentioned as @username. Nevertheless, this service is also exploited by a spammer. To write spam is also permitted by some Twitter applications [3].

The management team of the twitter site has authority to lock or close the doubtful accounts along with separate malicious tweets from the legal or, genuine tweets. It also possible to sometimes the legal user sends complaints that their twitter account has been stopped by Twitter's by mistake, recently team of Twitter started to close the spam accounts[4]. These discussed methods are particularly relying on the users experience to find the spam manually. But in today's era we need some tools or, techniques that automatically detect a close the spammers account. Additionally, we need some methods for spam detection by which dissatisfaction with authorized users have been avoided [5].

The main purpose f this paper is to produce an automatic detection system for identification and removal of spam from social networking sites. Mainly, the contribution of this paper is as discussed below:

- To design a novel pre-processing mechanism on the basis of the corpus method.
- To develop a new fitness function for ABC optimization technique.
- To perform the classification for spam and non-spam data, training based on ANN is designed.
- The efficiency of the developed model and the existing classifiers such as Naïve Bayes and SVM has compared.

II. BACKGROUND AND RELATED WORK

A. The Twitter

Twitter is very popular and fashionable online as well as micro-blogging service of social networking websites. Like the other online social media websites, Twitter included lower published barriers and allows users to post content in the structure of tweets. Twitter permits individuals to make a post text and attached with multimedia files such as images, URLs and videos as outside entities [6]. In the Fig. 1 the snapshot of the Twitter website included with different activities that could be done on the website is depicted.

The snapshot of a twitter program included with multiple activities that are performed on the website is shown in this Fig. given below. The text on the Twitter post termed as tweets having 140 characters and is also known as micro-posts [7].

Fig. 1: Snapshot of Twitter [16]

B. Related Work

Cao et al. (8, 2015) discussed about a technique utilized to identify the spam URLs included in social sites that are useful in to protect individuals from the links that contains malware or, any other low-quality of suspicious messages. The observation of this can be performed by utilizing two distinguished methods; firstly, go through the link posted by publication of Twitter. Secondly, how these links are accessible through users. Jain et al. (9, 2018) presented a CNN (Convolutional Neural Network) to perform classification for the detection of spam in social network. Proposed model is also called as SCNN (Semantic convolution neural network), since a semantic layer has been added in it. Word2Vec has been used to train the SCNN and gained semantic enrich words. The obtained accuracy is 94.40% as compare to the Twitter datasets. Ezpeleta, et al (10, 2018) developed a "Bayesian spam filtering" classification mechanism to detect the spam. To experimentally observe the result of the proposed algorithm utilized two different datasets such as "YouTube Comments dataset" and "YouTube spam collection dataset". Dwyer et al. (11, 2007) presented a classification scheme i.e. Bayesian to distinguished among legal and illegal behaviour of the users. The measurement of performance such as precision, F-measure has been computed and it has been concluded that the Bayesian classification scheme performs better as compare to other previous algorithm. Dutta et al. (12, 2018) produced an attribute selection scheme through the principle of rough set theory. This work has been done on five distinguished spam classification datasets and contrast the outcome with the proposed work. The selection of attribute is the main process in machine learning with data mining. Ala'M et al. (13, 2018) proposed a hybrid machine algorithm consisting of SVM as a classification mechanism and whale as an optimization algorithm. These algorithms have been used to recognize spammers in social networking sites. Aslan et al. (14, 2018) proposed an automatic detection scheme to provide security for Twitter users from spammers. In this paper, three classification approaches such as SVM, random forests and decision tree has been utilized, it has been also seen that when random forest utilized along with decision tree algorithm the complete accuracy obtained 95%. If random forest is utilized with behavioral feature, the accuracy of the detection system has been enhanced up to 97.877%.

III. PROPOSED SPAM DETECTION MODEL

This proposed spam detection algorithm consists of four modules such as input, processing, classification and evaluation as shown in Fig. 2. In this work it has to be assume that the ANN is trained according to the features of text data.

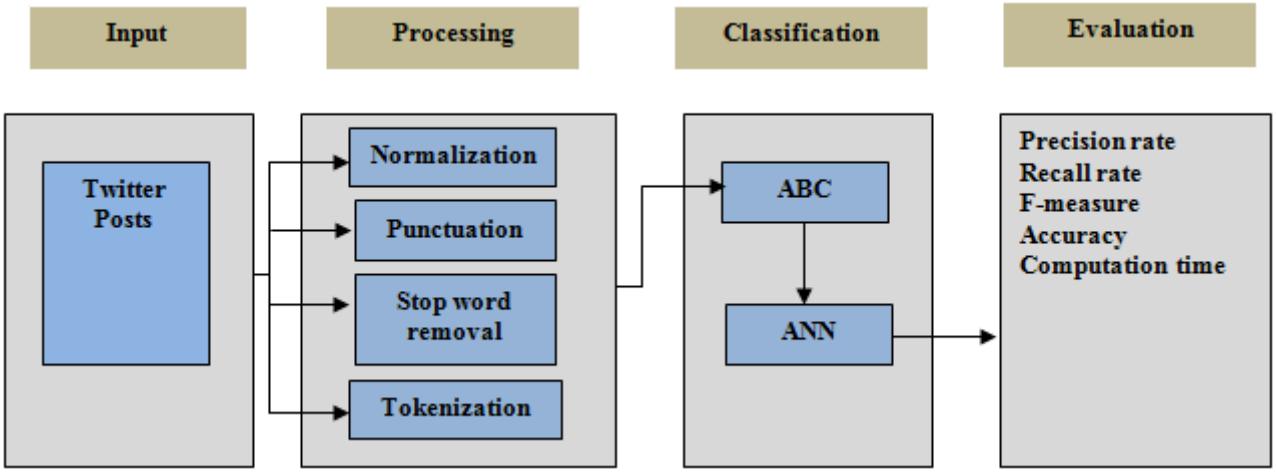

Fig. 2: Designed model

In the given model, the twitter posts are utilized as input data. The raw data can be involved in the post of twitter, so it becomes necessary to remove the irrelevant text and store the words that are compulsory for classification. In pre-processing mechanism various steps are included in it; (i) Normalization is utilized to convert the text into lower case (ii) Punctuation process removes the marks such as comma, full stop,

brackets etc that are basically used to separate the sentences have removed, (iii) In the process of stop word removal, the words like is, am, are, there, here, that that, those, which etc that are used commonly in the sentences are filtered out. (iv) at the end the pre-processing step, tokenization is utilized to produce features. In the process of Tokenization splits the sentences into single words, it is also helpful to compute the weight of the string as per the alphabet. The fitness function is defined as per the optimization rate of alphabets. If the feature value is higher than the fitness function is applicable as an input in the classification process; otherwise, it has to be avoided. The training of ANN has been done on the basis of optimized features of the spam-data and while the testing process has been performing. The text is contrast and observed based on the performance parameters as discussed in the evaluation block.

A. ABC

In this proposed work, ABC algorithm is used to discover the optimal features sets from spam and non-spam files. The features of the spam and non-spam message through employing a suitable fitness function. A threshold value will be selected by utilizing the fitness function based on the concept, by which the text has been classified as spam and non-spam. This algorithm has been utilized to optimize the extracted sets of features and the features that are not important are removed. The algorithm is written below in given steps:

ABC Algorithm:

Input: Feature of Spam & Non-spam data as per the token value and Fitness Function of ABC
Output: Token value that are optimized
Step 1: Initialization of ABC according to their operating functions – Bee Size
– Employee Bee
– Onlooker Eggs
– Scout Eggs
Step 2: Fitness function of ABC has defined for optimal feature selection

$$Fitness\_function = True; \quad if\ Bee_{Employee} > E$$
$$False; \quad otherwise$$

Step 3: Compute 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
Step 14: End

B. ANN

ANN (Artificial Neural Network) is utilized to distinguish between the spammer and the real user. This mechanism is a computer programs having biological inspirations are deployed to mimic the way of human information processes the information through human brain.

Input: Optimized token data as a Training Data (T), Target (G) and Neurons (N)
Output: Text Polarity as Spam and Non-spam

Step 1: Initialize ANN with parameters – Epochs (E)
– Neurons (N)
– Performance parameters: MSE Gradient, Mutation and Validation
– Training Techniques: Levenberg Marquardt (Trainlm)
– Data Division: Random
Step 2: For 1→ T
Step 3: If Training Data ε Spam
Step 4: $Group\,(1) = Categories\ of\ Trainingdata$
Step 5: Else if Training Data ε Non-spam
Step 6: $Group\,(2) = Categories\ of\ Trainingdata$
Step 7: Else
Step 8: $Group\,(3) = Categories\ of\ Trainingdata$
Step 9: End
Step 10: Initialized the ANN using Training data and Group
Step 11: Net = newff (Training Data, Group, *N*)
Step 12: Set the training parameters according to the requirements and train the system
Step 13: Net = Train (*Net, Training data , Group*)
Step 14: Classification Results = simulate (Net, Optimized Current Text Token Value)
Step 15: If Classification Results = True
Step 16: Show classified results in terms of their polarity
Step 17: Calculate the performance parameters
Step 18: End
Step 19: Return: Classified Results
Step 20: End

ANN gathered the knowledge by observing patterns and relationship of data and training or, learning at the time of experience, not through the programming. The developing process of ANN is varies from units, components of processing i.e. artificial neurons related to weights, that used to form the neural formation and arrangement of this in layers. The algorithm of ANN to detect the spam is explained below in steps:

C. Naïve Bayes

This approach is included in the machine learning, working of this algorithm based on Baye's theorem. Simply, Naives Bayes believes on the presences of a specific attribute in one class that depends on the presence of another attribute [15]. This model is very easy to install and preferable helpful in the large and complex data sets. Additionally, performance of this algorithm is better as compare to the high class comprehensive classification methods. For few of the probability models, naive Bayes classification has ability to receive higher producing training inside the controlled learning environment. In number of experimental usage, the parameter used for estimation in naïve Bayes model utilizes the maximum probability scheme. We can say that in other terms the Bayesian has ability to perform with the naïve Bayes model without accepting the possibility or,

utilizing Bayesian methods. Apart from simplicity of their loyal design and its appearance, this model performs well in various complicated real world scenarios. In this proposed work the optimized data obtained through ABC algorithm are utilized in the training of Naivy Bayes and this scheme is mainly used to classify spam from the text/

D. SVM

This approach is a supervised learning method, utilized to resolve the regression along with classification task. In this method, the data is plotted in the n-dimensional space and each data consisting of featured value as well as their co-ordinates. The hyper plane has utilized to distinguished non-spam and spam text. In this work, SVM is utilized to make a separation of normal text from spam text.

IV. RESULTS AND SIMULATION

The experimental result has been performed on Twitter dataset taken from HSpam 14. HSpam consists of more than 14 million tweets. Data was collected using the trending theme on Hashtags.org, 2019. In this research work, we use last couple of day tweets. Almost all tweets in HSpam14 are labeled as spam and ham, and the remaining small parts are classified as unknown since 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

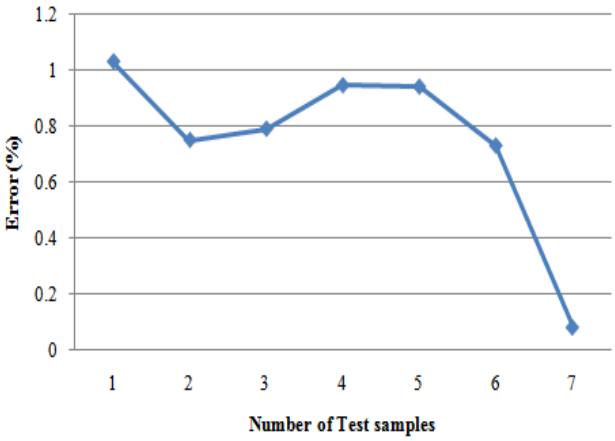

Fig. 3: Error (%) vs Number of test samples

Fig. 3 shows that error occurred during the classification process of spam. The average error observed for the seven different test data samples is approximately 0.752 (%), which is very small, which means that the spam model developed classifies spam and non-spam text with high accuracy.

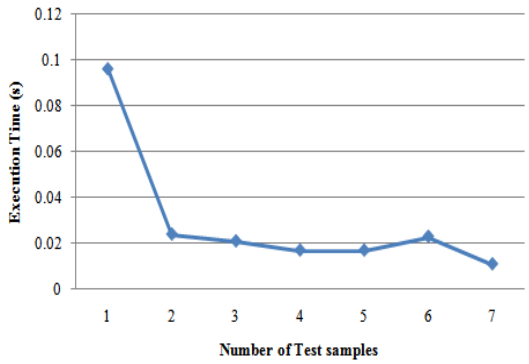

Fig. 4: Execution Time (s) vs Number of test samples

Execution time is defined as the total time provided by the model from the test process to the classification and evaluation process. The average execution time of seven test samples is approximately 0.029 s, which means that the device is quick enough to detect spam immediately.

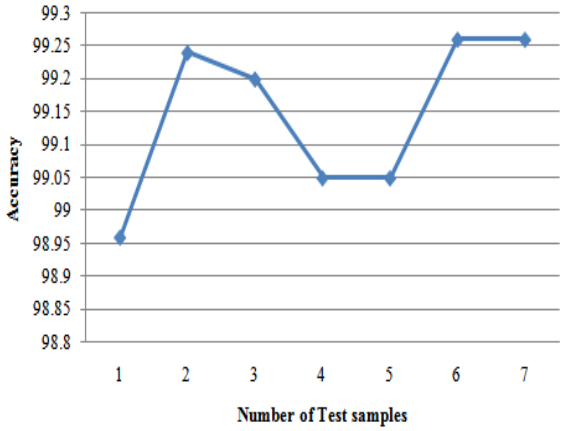

Fig. 5: Accuracy v/s Number of test samples

Fig: 5 shows the accuracy of designed system; the average accuracy is approximately 99.14% has been observed. After observed the accuracy result, the model adaptively learns new spam activities and maintains high accuracy for spam detection in a tweet post.

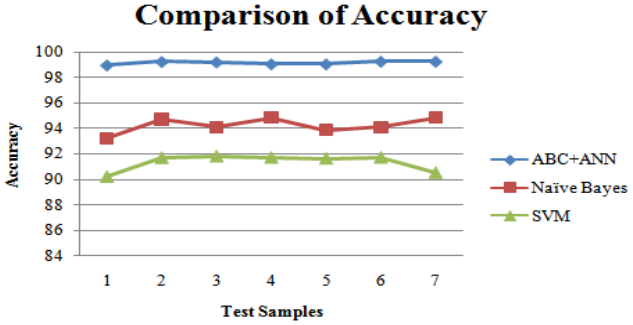

Fig. 6: Comparison of Accuracy

Fig. 6 shows the accuracy graph designed for proposed algorithm (ABC with ANN) along with other two existing algorithms named as Naïve Bayes and SVM (Support Vector System). The graph clearly described that when machine learning scheme is trained with optimized features; the accuracy of the system is higher as compared to the individual classifiers. The result shows that the average accuracy of proposed work, Naïve Bayes and SVM are 99.14, 94.26 and 91.36 respectively. After comparison of existing algorithm, thus there is an increase of 5.18 % from Naïve Bayes and 8.52 % from SVM.

Table II: 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

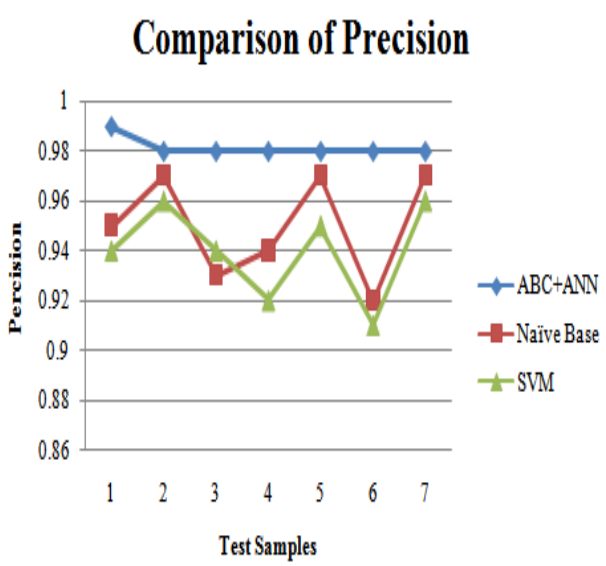

Fig. 7: Comparison of Precision

Fig. 7 shows the comparison of precision designed for the proposed model has been compared with the existing classifiers such as Naïve Bayes and SVM. The above graph, clearly describes that the detection of identifying spam using the ABC algorithm in hybridization with ANN which perform better as compared to Naïve Bayes and SVM. The average value of precision for proposed classifiers such as Naïve Bayes and SVm are 0.98, 0.95 and 0.94 respectively. Thus, there is an improvement of 3.16 % in the precision rate while ANN with ABC algorithm compared to Naïve Bayes and improvement of 4.26 % compared to SVM during the classification of spam in the designed model.

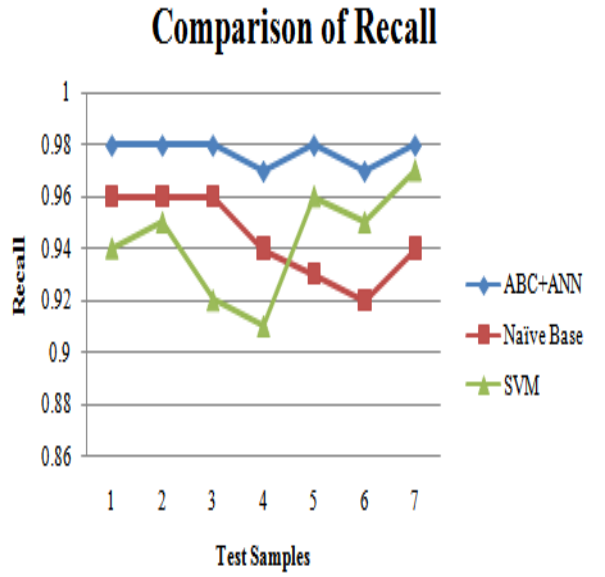

Fig. 8: Comparison of Recall

Fig. 8 represents the comparison of Recall, in this recall rate is computed by Naïve Bayes, SVM and proposed algorithm. The red line shows the Naïve Bayes, Blue line shows the proposed Algorithm (ABC+ ANN) and Green Line shows the SVM classifier. The blue, green and red line represents the recall rate observed for seven number of test samples determined for Naïve Bayes, SVM and proposed algorithm. The average value of recall for (ABC+ANN, Naïve Bayes and SVM) is 0.977, 0.944 and 0.942 respectively. The recall rate of proposed algorithm has been increased by 3.5 % from Naïve Bayes and 3.72 % from SVM approach.

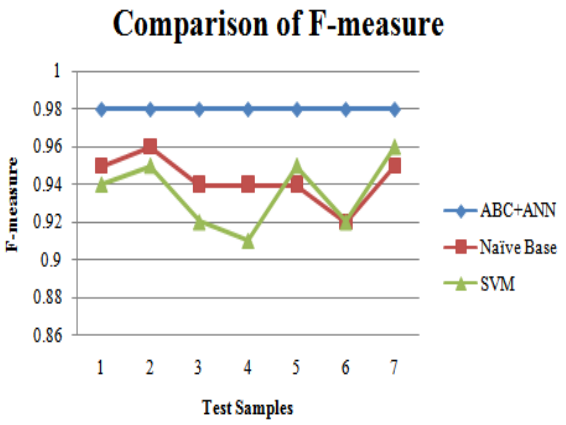

Fig. 9: Comparison of F-measure

Fig. 9 shows the comparison of F-measure. The F-measure is defined as harmonic means of precision and Recall. In above Fig., the values measured for the F-measure parameters for the proposed along with Naïve Bayes and SVM. The average value of F-measure determines the proposed algorithm is 0.98, Naïve Bayes is 0.94 and SVM is 0.93 respectively. It is observed form above Fig. that the proposed work has been increased by 4.26 % from Naïve Bayes and 5.38 % from SVM respectively.

V. CONCLUSION

In this study, we focused on the identification of social spam, particularly on the Twitter micro-blogging site. ANN approach with ABC has been used to distinguish between spam and non-spam tweets. According to the twitter spam policy, text data features are extracted using a tokenization process. The experimental results demonstrated the efficacy of the proposed work in terms of measured parameters such as error, execution time, accuracy, precision, recall and F-measure. The experiment result showed that pre-processing, optimization with the classification technique improved the accuracy of the spam detection system. The accuracy of the proposed spam detection program on the Twitter site was approximately 99.14 (%) achieved. Finally, a comparison was made between the proposed technique and the existing classification algorithms (Naïve Bayes and SVM). From the experiment result , it has been clearly observed that the proposed scheme (ABC with ANN) has perform well as compared to (Naïve Bayes and SVM) classifiers.

In the future, research can be enhanced through the use of other social media databases and the identification of 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

AUTHORS PROFILE

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 includes 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.