ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Spam Detection in Online Social Networks using Integrated Approach

Bhagyashri Toke[1], Dinesh Puri[2]

Research Scholar, Department of Computer Engineering, SSBT's COET, Maharashtra, India[1]

Assistant Professor, Department of Computer Engineering, SSBT's COET, Maharashtra, India[2]

ABSTRACT: Online Social Networks (OSNs) are considered to be the much in demand societal tool used by the masses world over to communicate and transmit information. On these platforms for seeking opinions, news, updates, etc. is increasing day by day. While it is true that OSNs have become a new medium for dissemination of information, at the same time, they are also fast becoming a playground for the spread of misinformation, fake news, unsolicited messages, etc. Consequently, OSN platform comprises of two kinds of users namely, spammers and non-spammers. Spammers, out of harmful intent, post either unwanted (or irrelevant) information or spread misinformation on OSN platforms. The proposed mechanisms to detect such users (Spammers) in Facebook social network (a popular OSN). The work is based on a number of features at post-level and user-level like Spam Words, Replies and Post. Two learning algorithms namely Naive Bayes and Rule Based are used. Furthermore, to improve detection of spammers, integrated approach is proposed which "combines" the advantages of the two learning algorithms mentioned above. On the basis of total accuracy, spammers detection accuracy and non-spammers detection accuracy the improvement of spam detection is measured. Results, thus obtained, show that integrated approach that combines all algorithms performs better than other classical approaches in terms of overall accuracy and detect non-spammers with 99% accuracy with an overall accuracy of 92.3%.

KEYWORDS: Online Social Networks, Spammers, Non-Spammers, Filtered Wall, Naïve Bayes, Rule based.

I. INTRODUCTION

Online Social Networks (OSNs) are a platform where people with common interests and interacts and connect. People visit OSN platforms to collect information relevant to them and also to build social and professional networks. OSNs like Facebook, twitter and LinkedIn are used by millions of users worldwide for making stronger interpersonal relationships and the number of users using these OSNs is increasing rapidly every day. These OSNs are becoming a new platform for dissemination of information, opinions and news. However, at the same time, some of the users, called spammers, are misusing these OSN platforms, thereby spreading misinformation, fake news, unsolicited messages, etc. Sometimes, the spamming is done with the intent of advertising and other commercial purposes, where spammers subscribe to various mailing lists and then send spam messages unsystematically to promulgate their interests. Such activities disturb the genuine users, called non-spammers and also decrease the reputation of OSN platforms.Therefore, there is a need to form mechanisms to detect spammers so that corrective actions can be taken thereafter.

The work focuses on detection of spammers over one of the most popular OSN platformslike, Facebook [1]. Facebook is viewed as one of the most common and much in demand online website utilize. Facebook is chosen as an OSN platform for the work because it offers a large number of users bases and also because information on Facebook is publicly available by default which can be accessed through APIs provided by Facebook. Being one of the most prominent OSNs, Facebook is always under attack by spammers. The spammers, thereafter, successfully proliferates spam messages among their highly connected communities. Another way in which spammers work is by sending the victim large number of direct messages called Direct Messaging (DM) spamming. This not only gives the identity of a spammer a legitimate appearance but also enables DM spamming, where the spammer can send direct messages that contain malicious content. There are many other techniques through which spammers can possibly gain popularity and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

spread malicious posts. In the work, the Facebook dataset is obtained from the authors of [2] and thereafter have performed preprocessing over it to obtain normalized set of features depend on which the activities of spammers were studied. The key features extracted were spam words, replies and post. After obtaining these features, two classical learning algorithms namely Naive Bayes, Rule Based were applied on these features to detect spammers in the dataset. Moving on, in order to improve detection of spammers, a novel proposed approach was devised which "combines" the advantages of the two classical algorithms (Naive Bayes, Rule Based).Finally, the improvement in spam detection is measured on the basis of accuracy parameters and the results, thus obtained show that novel integrated approach that combines all algorithms outperforms other classical approaches in terms of overall accuracy and non-spammer detection accuracy. The key contribution of work is the proposed integrated approach which combines the three learning algorithms namely Naive Bayes and Rule Based with an aim of improving spam detection accuracy.

II. RELATED WORK

The issue of spamming over emails and in many other forms is a well studied problem. Spam detection has been the area of interest of many researchers. Many solutions have been propounded in regard to spam detection. However, spam detection in the social networks, which is a recent phenomenon, has not been studied so widely. Also, the fact that post messages are small in size, restricted to 140 characters only (as opposed to email or web content), the problem of spam detection becomes more difficult. This section summarizes the main contributions of other researchers on spam detection in social networks. FabricioBenevenuto et al. [2] detected spammers by identifying various user social behaviors and the characteristics of post content. These characteristics were used in a machine learning approach to classify the users as spammers and non-spammers. De Wang et al. [3] in their study proposed a general framework to detect spam account across all the OSNs.The main contribution of their work was a new spam detected in any one social networking could be quickly identified across all other OSNs. Alex Hai Wang et al. [5] proposed a model which uses a directed graph that depicts the relationship between "friends" and "follower" relationship in facebook. Bayesian classifier was also used in his work, to detect spam accounts. Xin Jin et al. [6] propounded a method for detecting spam accounts in social media network. They employed a GAD Rule Based algorithm integrated with designed active learning algorithm to deal with spam accounts. M. McCord and M. Chuah [7] discussed various features related to user and post content which can be utilized in the detection of accounts intended for spamming.In their work, he evaluated four classifiers and compared there accuracies. Carlos Castillo et al. [9] constructed a spam detection system that exploits the linked dependencies of web pages. The algorithm assumed that the linked web pages belonged to same class, i.e. if one web page is spam than its linked web pages must also be spam. HongyuGao et al. [11] studied spam accounts in one of the popular OSN, Facebook. A dataset of "wall" messages between various Facebook users was used to identify spam accounts. In their study, they found out that spamming was most common during early hours, when regular users were asleep. Kenichi Yoshida et al. [15] worked on email spam.They used an unsupervised approach to detect spam accounts. Benjamin Markines et al. [16] devised a mechanism to detect social spam. In their work, six features were recognized which were used to distinguish between spammers and non-spammer users.The features were passed to various machine learning algorithms which further classified the spam and non-spam accounts [20].

To the best of knowledge, some of these works applied each of the machine learning algorithms separately but not in a combined manner which has resulted in improving spam detection, one of the key contributions of the work.

III. PROPOSED APPROACH

An overview of the complete process of spam detection is shown in the diagram in Figure 1, each of whose steps are explained in this section. The preliminary step for the detection of spammers in any OSN is data collection and necessary preprocessing to convert it into a form, which can be used by the learning algorithms.

A. *DATA SET DESCRIPTION:*

In the work, Facebook dataset obtained from FabricioBenevenuto et al. [2] which consists of labeled record of 5000 Facebook users. Dataset comprises of 62 features containing user specific and post specific information. The spammer accounts comprised of around 50% of the dataset. Also, as per [2], the users were chosen randomly and not based on

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

any of their characteristics. They [2] have used SVM based machine learning approach as opposed to the work in which other learning approaches namely Naive Bayes, Rule Based are used and finally combined all ofthem together to achieve a higher spam detection accuracy.

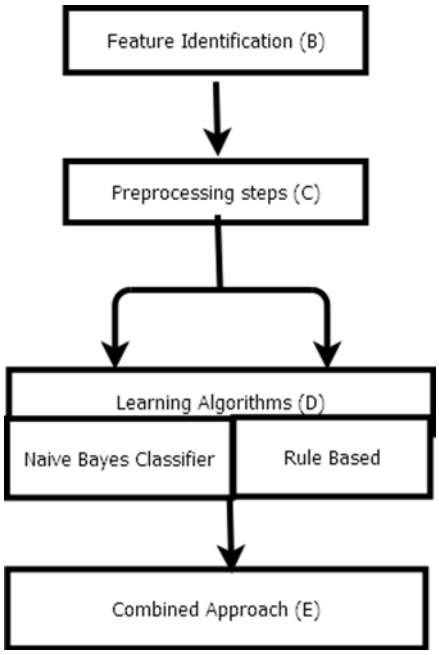

Figure 1: Proposed Spam Detection Approach

B. ***FEATURE IDENTIFICATION:***

Since, spammers behave differently from non-spammers; therefore it can identify some features or characteristics in which both these categories differ. Various features are used to detect spam accounts includes:- Spam Words: An account with spam words in almost every post can be considered to be a spam account. Therefore, "Fraction of post with spam words" can be considered as an important factor for detecting spammers. Replies: Since, information or message sent by a spammer is useless, therefore people rarely replies to its post. On the other hand, a spammer replies to a large number of posts in order to get noticed by many people. This pattern can be used in the detection of spammers.

C. ***PREPROCESSOR:***

Facebook user accounts in the dataset [2], labeled as spammer and non-spammers, and were used for training the learning algorithms and also in accuracy calculations. In preprocessing step, all the continuous features were converted into discrete. The procedure adopted to select the intervals for a particular feature was obtained from [4] according to which all user accounts are arranged in increasing order of their feature values. Processing begins from the first account, if it encounters an account whose category is different from the category of the next account, and then an interval is created as a mean of both the feature values.

D. ***LEARNING ALGORITHMS:***

There are various different classification algorithms, which can be used to classify an account as "spammer" or "non-spammer". In the work, learning algorithms such as Naive Bayes, Rule based are used. Although, each of these approaches can be solely used to classify user accounts, but in order to increase the accuracy, the integrated algorithm is used by combining these two approaches. In proposed approach as outlined in Figure 1, the preprocessed data is first

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

classified using different learning algorithms to predict the class {"Spammers" or "Non- Spammers"} of all Facebook user accounts. In Naive Bayes approach, accounts were classified by calculating the probability of the given account to be P (C/F)=P(F/C)*P(C)/P(F), Here,F is a vector of the features of a user C is the class (Spammer/Non-Spammer) of the user.If the calculated probability of an account to be spam is more than 0.5, than that account is classified as spammer, otherwise it is considered as a non-spammer (genuine account).

Rule based is basically an unsupervised learning technique. Unlike Naive Bayes, a separate training data need not be prepared for this algorithm. On the basis of similar feature values (like similar kind of reply trend, similar usage of spam words in post), this algorithm could classify the entire set of accounts unto two classes. One of this class was labeled as spammer and another as non- spammer.

E. *COMBINED APPROACH:*

As part of combined approach, it compares the classification results of any two learning algorithms, if both the learning algorithms predict the same result, then it finalizes the class of the Facebook account under investigation. Otherwise, if the predicted class of both the classification techniques differs, then it uses the prediction of combination algorithm as the final class. Results obtained following this approach show an improvement in spam detection.

IV. EXPERIMENTS AND EVALUATION

For the corroboration of the accuracy of the algorithm the results obtained from the algorithm were compared with the labels (Spammers/Non-Spammers) in the dataset [2]. Output of various algorithms implemented is depicted in tables below.

It is evident that the proposed algorithm was able to successfully identify an account as spammer or non-spammer with 88% accuracy. The algorithm's accuracy of detection of non-spammers was higher (99.1%) as compared to the accuracy of detection of spammers (68.4%). This integrated algorithm was then compared with each of the learning algorithm, Naive Bayes, Rule Based. The results showed that Rule Based algorithm performs better in detection of non-spam accounts but was very poor in detecting spam accounts. The algorithm was able to maintain the high accuracy of rule based algorithm in detecting non-spam and at the same time, retain the accuracy of Naive Bayes in detecting spammers accounts thereby, increasing the overall accuracy. The graphical representation of the above data is shown in Figure 2.

	Predicted values	
Actual Values	Non-Spammers	Spammers
Non-Spammers	2015/2500	485/2500
Spammers	1865/2500	635/2500

Table 1: Naïve Bayes Results

	Predicted values	
Actual Values	Non-Spammers	Spammers
Non-Spammers	1365/2500	1135/2500
Spammers	1250/2500	1250/2500

Table 2: Rule Based Results

	Predicted values	
Actual Values	Non-Spammers	Spammers
Non-Spammers	2475/2500	25/2500
Spammers	2308/2500	192/2500

Table 3:Integrated Approach Results

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

	Precision	Recall	F-Measure
Naïve Bayes	80.6	74.6	77.46
Rule Based	54.6	50	52.03
Integrated	85.6	77.2	81.17

Table 4: Overall Result

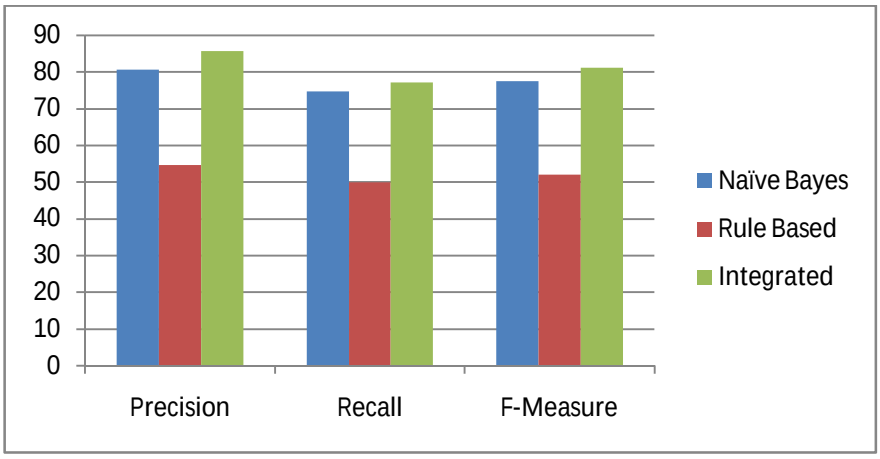

Figure 2: Result graph

V. CONCLUSION AND FUTURE WORK

An algorithm, combining two different learning algorithms (namely Naive Bayes, Rule Based) was implemented. The integrated algorithm categorizes an account as spammer/non-spammer with an overall accuracy of 92.3%. Finally, this algorithm was compared with the two learning algorithms taken alone. It was observed that the combined approach could give best results in terms of overall accuracy and in detection of non-spammers.

In future, the new kind of rules for the new kind of spams and different integrated approach with it will be the possibility for the improvement of the spam detection.

REFERENCES

1. "Facebook" https://facebook.com
2. FabricioBenevenuto, Gabriel Magno, Tiago Rodrigues, and Virgilio Almeida, "Detecting Spammers on Facebook", Proceedings of Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), 2010.
3. De Wang, DaneshIrani, and CaltonPu, "A Social-Spam Detection Framework", Proceedings of Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), 2011.
4. Dewan Md. Farid, NouriaHarbi, and Mohammad ZahidurRahman, "Combining Naive Bayes And Decision Tree For Adaptive Intrusion Detection", International Journal of Network Security & Its Applications (IJNSA), Volume 2, Number 2, April 2010.
5. Alex Hai Wang, "Don't Follow Me: Spam Detection In Facebook", Proceedings of Security and Cryptography International Conference (SECRYPT), 2010.
6. Xin Jin, Cindy Xide Lin, JieboLuo and Jiawei Han, "A Data Mining-based Spam Detection System for Social Media Networks", Proceedings of the VLDB Endowment, Volume 4, Number 12, August 2011.
7. M. McCord and M. Chuah, "Spam Detection on Facebook Using Traditional Classifiers", Proceedings of Autonomic and Trusted Computing International Conference (ATC), 2011.
8. Kurt Thomas, Chris Grier, Vern Paxson and Dawn Song, "Suspended Accounts in Retrospect: An Analysis of Facebook Spam", Internet measurement conference (IMC), 2011.
9. C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri, "Know your neighbors: Web spam detection using the web topology", in Int'l ACM SIGIR, 2007.
10. G. Stringhini, C. Kruegel and G. Vigna, "Detecting Spammers on Social Networks", Proceedings of ACM ACSAS, 2010.
11. H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Zhao, "Detecting and characterizing social spam campaigns", Proceedings of the Internet Measurement Conference (IMC), 2010.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

12. F. Benevenuto, T. Rodrigues, V. Almeida, J. M. Almeida, C. Zhang, and K. W. Ross, “Identifying video Spammers in online social networks”, in AIRWeb, pages 45–52, 2008.
13. C. Pu and S. Webb, “Observed trends in spam construction techniques: a case study of spam evolution”, Proceedings of Conference on Email and Anti-Spam (CEAS), 2006.
14. Leyla Bilge, Thorsten Strufe, DavideBalzarotti and EnginKirda, “All your contacts are belong to us: automated identity theft attacks on social networks”, Proceedings of ACM World Wide Web Conference, 2009.
15. K. Yoshida, F. Adachi, T. Washio, H. Motoda, T. Homma, A. Nakashima, H. Fujikawa, and K. Yamazaki, “Density-based spam detector”, Proceedings of the Tenth ACM SIGKDD International Conference, 2004.
16. B. Markines, C. Cattuto, and F. Menczer, “Social spam detection”, in AIRWeb, pages 41–48, 2009.
17. H. Drucker, D. Wu, and V. Vapnik, “Support vector machines for spam categorization”, IEEE Transactions on Neural Networks, 10(5): pp. 1048–1054, 1999.
18. S. Webb, J. Caverlee, and C. Pu, “Introducing the webb spam corpus: Using email spam to identify web spam automatically”, Proceedings of the Conference on Email and Anti-Spam (CEAS), 2006.
19. I. Drost and T. Scheffer, “Thwarting the negritude ultramarine: Learning to identify link spam”, Proceedings of the European Conference on Machine Learning (ECML), 2005.
20. A. Gupta, and R. Kaushal, "Improving Spam Detection in Online Social Networks," IEEE Cognitive computing and information processing, March 2015 , pp. 1-6.
21. B. Toke, and D. Puri, “Review on Spam Detection in OSN using Integrated Approach”, International Research Journal of Engineering and Technology, May 2016