Feature Subset Selection for Twitter Spam Detection

Harshita Tiwary
Research Scholar,CSE,FET MRIU
Kolkata

Indu Kashyap
Associate Professor, Dept of CSE,FET MRIU
Faridabad

ABSTRACT
Rapid growth of social networking have had an immense effect on today's general public and Web stage.Social networking sites are developing in both size and prevalence with a high rate in recent years. Twitter is one of the quickest developing Social Networking Sites.With the measure of information developing in Twitter lately, detection of spam in real time has become a challenging task for researchers as well asfor Twitter itself. Enormous work is being done towards spam detection.

The work done previously was not giving the appropriate results in the context of content based spam discovery on Twitter.In this paper accuracy is analyzed by using Classical approaches like Naïve Bayes and Random forest algorithm.It is observed that these algorithms are not giving accurate results.With a specific end goal to increase the accuracy of spam detection Random forest with Feature Subset Selection have been used.

Here the aim is to propose a Feature Subset Based Classification Approach where a set of features will be tested using Random Forest Classifier for twitter spam detection. In this paper the capabilities of Random Forest Classifier has been extended for detecting spam by including Feature Subset with it.

General Terms
Intelligent Decision Support System,Security,Algorithms et. al.

Keywords
Labeled Dataset, Feature subset selection,Random Forest

1. INTRODUCTION
Online social networking websites, such as Twitter,Facebook and LinkedIn, are now part of manyindividuals' day by day schedule: from posting their recentencounters,discovering what companions are up to and monitoring the most sultry patterns, to review fascinating photographs. Twitter is additionally a microblogging service, established in 2006, where clients can post 140 character messages called tweets.The objective of Twitter is to permit friends to communicate and remain associated through the exchange of short messages.

Dissimilar to Facebook and MySpace, Twitter is coordinated, implying that a user can follow another user, but the second user is not required to follow back. Most accounts are public and can be followed without requiring theowner's approval. Twitter has followed some safety efforts to anticipate spam yet spammers are discovering an ever increasing number of procedures to trap genuine clients.Along these lines, fundamentally this paper rotates around the strategy of identifying spam in tweets.

Twitter has its own categorization. This segment characterizes Twitter Taxonomy [7]:-

a)Tweets[7]:Twitterallows the communication of short messages.The maximum character length allowed is 140 which the user can use to post messages.

b)Followers[7]: The follower of a user is the group of users whoget a tweet when posted. If a tweet has been posted by a user onhis home page, all of his followers get the same tweet on their home pages as well.

c)Friends[7]:Friends are the set of users an accountsubscribes to in order to obtain access to status updates.

d)Hashtags[7]: Hashtags is being demonstrated by a # symbol and these arecombined with keywords that indicate a theme of interest.

e)Trending Topics[7]:These are the prevalent hashtags thatshow up on the main page of Twitter and can expand the quantity of tweets containing that topic.

Spam Tweets are increasing at a very high pace.It isunable to differentiate between the spam tweets and ham tweets.This growing misconception is very dangerous for us.It need to be tracked seriously.Even the various machine learning algorithms are not up to the mark.These algorithms are too lagging behind in terms of differentiating between the two.In order to increase the accuracy of spam classification a few subset of features havebeen selected that would be tested using Random Forest Classifier.Each feature selected will be evaluated and based on the evaluation the subset of features will be formed.These feature subset would help in increasing the accuracy of spam detection.

2. LITERATURE SURVEY
Shukla Twinkle and Shirsagar D.B.K[6] have used Naive Bayes classifier to detect the spam tweets.They had created two modules in their work.One is the user module and the other is admin module.Each user logged in is allowed to post comments.The framework will create a record containing legitimate and spam post from the current client.The admin can view list of valid and invalid spam post.Some of the parameters analyzed by the authors were number of followers,number of followings,number of records,number of tweets,number of digits,number of characters etc.However the parameters they had used were user specific and thus could not give much accurate results in differentiating between spam tweets and ham tweets.There was a need to go beyond the user specific method to increase the accuracy rate.

Khurana Girish and Kumar Marish[5] have investigated the existing techniques for identifying spam users in twitter social network.They had used both the substance based feature and user based componentsin order to detect spam. Classicevaluation metrics was used by them to compare the execution of different traditional classification methods like Decision Tree, Support Vector Machine (SVM), Naive Bayesian, and Neural Networks and among all Bayesian classifiers have been judged the best as far as performance is concerned.Some of the parameters that had been considered

by them for spam detection were the following- number of hashtags per number of words in each tweet, number of URLs per word, number of words of each tweet, number of characters of each tweet, number of URLs in each tweet, number of hashtags in each tweet, number of numeric characters that show up in the content, number of clients mentioned in each tweet, number of times the tweet has been retweeted.Even here the same problem was faced.The rate of accuracy was too low.There had to been some alternative solution.

Chen Chao,Wang Yu,Zhang Jun,Xiang Yang and Zhou wanlei and Min Geyong[1],have used random forest algorithm to detect spam tweets.They had collected and labeleda genuine dataset ,which contains 10 continuous days tweets with 100k spam tweets and 100k non spam tweets in each day.They have worked on user specific features and not the content specific features.They have then compared the performance of the algorithm by Naive Baiyes and Decision Tree based algorithm.They also have used 12 different features like the days of an account record since its creationuntil the time of sending the latest tweet,the number of follower of this twitter user,number of followings/companions of this twitter user,the number of retweets,the number of hashtags and so on.

Twitter has turned into an objective stage for both promoters and spammers to spread their messages, which are more unsafe than conventional spamming methods. Recently, large amounts of campaigns that contain lots of spam or promotion accounts have emerged in Twitter. The campaigns cooperatively postunwanted information, and thus they can infect more normal users than individual spam or promotion accounts. Organizing the campaign and participating in it has become one of the main technique to spread spam or promotion information in Twitter.X.Zhang,S.Zhu and w.Liang[2] had worked towards recognizing spam and promotion campaigns in Twitter. Theirsystem comprises of three stages: in the first step it helped in determining the accounts that post URL for the purpose of promoting spam; in the second step the data had been extracted that may had been for spam or promotion purposes; and in the third step it classified the data into spam or normal. The main aim was to gauge the comparability between the various accounts for posting URLs for the similar purpose.They had considered two important parameters for measuring the similarity the first one uses the URLs posted by the users, and the second one considers both URLs and timestamps.However there is no consideration of the content in the work they had proposed.It had only talked about URL posting and the time of posting the URL.This itself does not give much of information.

k.Thomas,C.Grier,J.Ma,V.Paxson and D.song [3] have worked on URL filtering.They had tried to determine whether a URL submitted for web services is a spam or not.A real time system called Monarch was created by the authors that used to crawl URL as they were submitted to web services and decides if the URL lead to spam.If the URL submitted lead to a spam then that request used to get filtered.No response was to be given for such a request.This method also had many drawbacks.We cannot determine if the text or web page of that URL also contains spam just by determining that particular URL can lead to a spam.It was not preferred as one of the really good method or approach in differentiating whether a particular data contained spam text or not.

H.Gao,J.Hu,C.Wilson,Z.Li,Y.Chen and B.Y.Zhao[4] had worked specifically in detecting spam campaigns on facebook.They had gathered an extensive dataset of "wall" messages between facebook clients. All wall messages gotten by around 3.5 million Facebook users which is more than 187 million messages on the whole had been analyzed.A set of automated techniques had been used by the authors to characterize coordinated spam campaigns.It was detected that over 70% of the malicious wall posts advertised phishing websites.It was a good approach made by the authors in detecting spam onfacebook.This work is similar to the approach where spam detection is based on Feature selection in Twitter .

3. TWITTER SPAM DETECTION

Table.1 Original Dataset

Even my brother is not like to speak with me. They treat me like aids patent.	Ham
URGENT! You have won a 1 week FREE membership in our ??100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18	Spam
Is that seriously how you spell his name?	Ham
As a valued customerI am pleased to advise you that following recent review of your Mob No. you are awarded with a ??1500 Bonus Prizecall 09066364589	Spam
Lol your always so convincing.	Ham
Thanks for your subscription to Ringtone UK your mobile will be charged ??5/month Please confirm by replying YES or NO. If you reply NO you will not be charged	Spam
I see the letter B on my car	Ham
Congrats! 1 year special cinema pass for 2 is yourscall 09061209465 now! C Suprman VMatrix3StarWars3,etc all 4 FREE! bx420-ip4-5we. 150pm. Dont miss out!	Spam
Sorry my roommates took foreverit ok if I come by now?	Ham
Urgent UR awarded a complimentary trip to EuroDisinc TravAco&Entry41 Or ??1000. To claim txt DIS to 8712118+6*??1.50(moreFrmMob. ShrAcomOrSglSuplt)IOLSI 3AJ	Spam

A dataset of 10 tweets has been taken whose class has been given.classification of these 10 tweets are done separately by using Naïve Bayes,then by using Random Forest and then later on by our proposed technique.It is analysed how the results vary for all the three different algorithms and how Random Forest with FSS is better than Naïve Bayes andRandom Forest approach.

3.1 Classification of tweets by Naïve Bayes Classifier

Naive Bayes is a basic procedure for building classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class names are drawn from some limited set. It is not a single algorithm for preparing such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers accept that the value of a specific feature is autonomous of the value of any other feature, given the class variable.

Table.2 Classification of tweets by Naïve Bayes

Tweet	Class	naïve bayes	Remarks
brother speak me. treat aids patent.	ham	ham	Correctly classified
URGENT! won 1 week FREE membership ??100,000 Prize Jackpot!	spam	spam	Correctly classified
seriously spell name?	ham	ham	Correctly Classified
valued customerI am pleased advise following recent review Mob No.	spam	spam	Correctly classified
Lol convincing.	ham	ham	Correctly classified
Thanks subscription Ringtone UK mobile charged ??5/month	spam	ham	Incorrect classification
letter car	ham	spam	Incorrect classification
Congrats! 1 special cinema pass 2 yourscall 09061209465 now!	spam	spam	Correctly classified
Sorry roommates foreverit ok now?	ham	ham	Correctly classified
Urgent UR awarded complimentary trip EuroDisinc TravAco&Entry41	spam	spam	Correctly classified

It is seen that the Naïve Bayes has not classified all the tweets correctly.It givestwo incorrect classification.

This means that it is not the efficient algorithm for detecting the spam in tweets. Thus Random Forest has been used to check if it is better than Naïve Bayes.

3.2 Classification of tweets by Random Forest Classifier

Random forests are a combination of tree predictors such that each tree relies upon the values of a random vector inspected independently and with the same distribution for all trees in the forest. Random forests or random decision forests are a group learning strategy for classification, regression and other tasks, that work by building a multitude of decision trees at preparing time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Table.3 Classification of tweet by Random Forest

Tweets	Class	Random Forest	Remarks
brother speak me. treat aids patent.	ham	ham	Correctly classified
URGENT! won 1 week FREE membership ??100,000 Prize Jackpot!	spam	ham	Incorrect classification
seriously spell name?	ham	ham	Correctly classified
valued customerI am pleased advise following recent review Mob No.	spam	spam	Correctly classified
Lol convincing.	ham	ham	Correctly classified
Thanks subscription Ringtone UK mobile charged ??5/month	spam	spam	Correctly classified
letter car	ham	ham	Correctly classified
Congrats! 1 special cinema pass 2 yourscall 09061209465 now!	spam	spam	Correctly classified
Sorry roommates foreverit ok now?	ham	ham	Correctly classified
Urgent UR awarded complimentary trip EuroDisinc TravAco&Entry41	spam	spam	Correctly classified

It has been analysed that the Random Forest classifier is better than the Naïve Bayes but it is not the best.There is an incorrect classification of tweet by Random forest approach. Thus a new technique has been proposed that is Random Forest with Feature Subset Selection.

4. PROPOSED TECHNIQUE

Feature subset selection (FSS) assumes an imperative part in the fields of data mining and machine learning. A good FSS algorithm can effectively remove irrelevant and redundant features and take into account feature interaction. Random Forest approach is better than the Naïve Bayes approach but it is also not very efficient in detecting spam on twitter.Thus there was a need to use different technique for spam detection by increasing its accuracy.

Table.4 Classification of tweets by Random Forest with FSS

Tweets	Class	RF+FSS	Remarks
brother speak me. treat aids patent.	ham	ham	Correctly classified
URGENT! won 1 week FREE membership ??100,000 Prize Jackpot!	spam	spam	Correctly classified
seriously spell name?	ham	ham	Correctly classified
valued customerI am pleased advise following recent review Mob No.	spam	spam	Correctly classified
Lol convincing.	ham	ham	Correctly classified
Thanks subscription Ringtone UK mobile charged ??5/month	spam	spam	Correctly classified
letter car	ham	ham	Correctly classified
Congrats! 1 special cinema pass 2 yourscall 09061209465 now!	spam	spam	Correctly classified
Sorry roommates foreverit ok now?	ham	ham	Correctly classified
Urgent UR awarded complimentary trip EuroDisinc TravAco&Entry41	spam	spam	Correctly classified

It can be seen that the accuracy rate of spam detection is higher by using this approach as compared to the Naïve Bayes approach and the Random Forest approach.

5. EXPERIMENTAL ANALYSIS

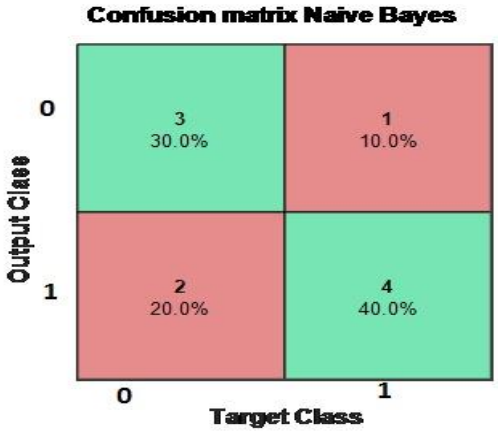

Fig.1 Confusion Matrix for Naïve Bayes

The confusion matrix for Naive Bayes is producing less accurate result.It is predicting 30% of the spam tweets as spam tweets and 10% of the spam tweets as ham tweets.It is misclassifying 20% of the ham tweets as spam.Hence the accuracy rate is getting reduced to a large extent. Another classifier is needed to improve its accuracy.

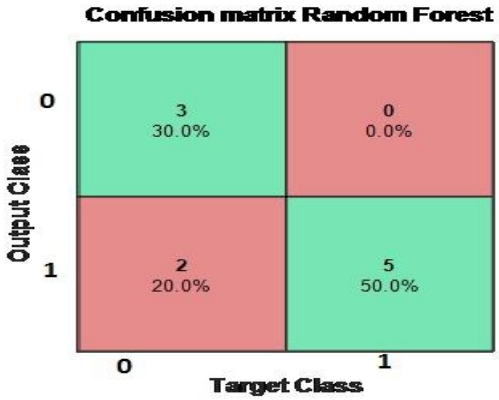

Fig.2 Confusion Matrix for Random Forest

The accuracy of Random Forest is better than the Naïve Bayes.It is predicting 30% of the spam tweets as spam tweets which is a larger percentage and 50% of the ham tweets as ham which is a good percentage.Thus the accuracy of the Random Forest classifier is increased and improved as compared to the Naïve Bayes approach.

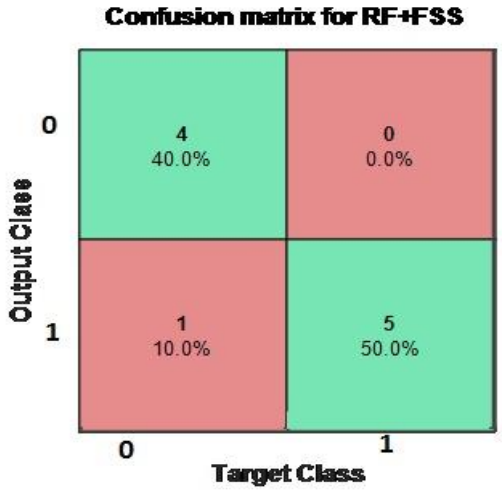

Fig.3 Confusion Matrix for Random Forest+FSS

The accuracy rate is further increased when Random Forest with Feature Subset Selection is used.It is predicting 40% of the spam tweets as spam and 50% of the ham tweets as ham which is a larger number as compared to Naïve Bayes and Random Forest.

After the creation of confusion matrix various measures for Naïve Bayes,Random Forest and Random Forest with FSS are calculated like accuracy,specificity,precision,recall and F-measure.

Table.5 Various measures for Naïve bayes,Random Forest and RF+FSS

Measure	naïve bayes	random forest	FS+RF
Accuracy	0.7	0.8	0.9
Sensitivity	0.8	1	1
Specificity	0.6	0.6	0.8
Precision	0.6667	0.7143	0.8333
Recall	0.8	1	1
F_Measure	0.7273	0.8333	0.9091
Gmean	0.6928	0.7746	0.8944

Based on these above calculated measure a chart is drawn representing different measures for Naïve Bayes,Random Forest and Random Forest with Feature Subset Selection.

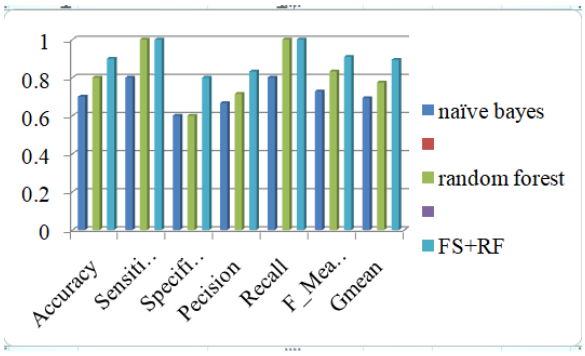

Fig.4 Chart representing different measures for Naïve Bayes,Random Forest and Random Forest with FSS

When the various measures for Naïve Bayes,Random Forest and Random Forest with FSS are calculated separately it isconcluded that the Random Forest +FSS is performing better in all terms.

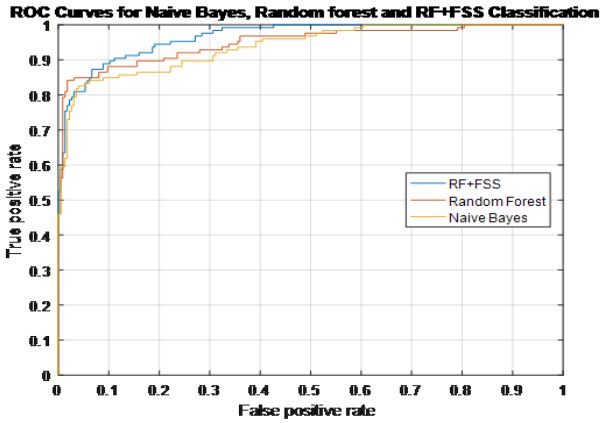

Fig.5ROCCurve

Receiver Operating Characteristic(ROC) curve is a plot of thegenuine positive rate against the false positive rate for the distinctive possible cutpoints of aanalytic test.

An ROC curve exhibits a few things:

(a)The nearer the curve takes after the left-hand border and then the top border of the ROC space, the more precise the test.

(b) The closer the curve goes to the 45-degree inclining of the ROC space, the less precise the test.

(c) The area under the curve is a measure of content precision.

In the ROC curve it is seen that the RF+FSS is covering more True positive rate area than the Naïve Bayes and Random Forest.Hence it is giving more accurate result in detecting spam in twitter.

6. CONCLUSION AND FUTURE SCOPE

Spam is actually unsolicited, unwanted messages typically delivered with commercial intent. The exercise of delivering spam messages is known as spamming.The sheer number of mass mails received is on the rise very year because spamming is financially viable as marketers do not have running costs beyond the handling of their mailing lists plus it is difficult to keep senders accountable due to their mass mailings. Spamming has already been the topic of legislation in lot of jurisdictions.

Additionally, spamming has actually become increasingly difficult to identify as the hackers become smarter.Feature subset selection (FSS) plays an vital act in the fields of data excavating and contraption learning. A good FSS algorithm can efficiently remove irrelevant and redundant features and seize into report feature interaction. This also clears the understanding of the data and additionally enhances the presentation of a learner by enhancing the generalization capacity and the interpretability of the discovering mode.

In this work,tweets has been characterized into various classes of spam and ham.Initially the labelled tweets are extracted and pre-processed to refine the tweets and later on these tweets have been used to train the classifier.Then, subset of features are created which helps in modifying the classifier further.After performing the classification the accuracy, precision and recall values are calculated whichclarifies that

how precisely spamhas been classified.It is concluded that the proposed technique gives more accurate results than the Naïve Bayes and Random Forest classifier. This research work can be additionally reached out by expanding the quantity of content features used.Also,this technique can be used to detect spam in any other social networking sites.

7. REFERENCES

[1] Chen Chao,Wang Yu,Zhang Jun,Xiang Yang,Zhou Wanlei,Min Geyong,"Statistical Feature Based Real Time Detection of Drifted Twitter Spam",IEEE Transactions on Information Forensics and security,2015

[2] X. Zhang, S. Zhu, and W. Liang. Detecting spam and promoting campaigns in the twitter social network. In Data Mining (ICDM), 2012 IEEE 12th International Conference on, pages 1194–1199, 2012.

[3] K. Thomas ,C. Grier, J. Ma, V. Paxsonand D. Song. Design and evaluation of a real-time url spam filtering service. In Proceedings of the 2011 IEEE Symposium on Security and Privacy, SP '11, pages 447–462, Washington, DC, USA, 2011. IEEE Computer Society.

[4] H. Gao, J. Hu ,C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao. Detectingand characterizing social spam campaigns. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, IMC '10, pages35–47, New York, NY, USA, 2010. ACM.

[5]KhuranaGirisha,Mr. KumarMarish,"Review:Efficient Spam Detection on social Network",International Journal for Research in Applied Science and Engineering Technology(IJRASET),vol 3,Issue VI,June 2015

[6]Miss.Shukla Twinkle Kailas,Prof.Shirsagar D.B.K,"Design of Machine Learning Approach for Spam Tweet Detection",vol-2,Issue 5,2016

[7]R .Kumar Arun , Mittal Shruti ," Twitter Spamming:Techniques and Defence Approaches",International Journal of Applied Engineering Research , vol 7,No.11,2012

[8] E. M. Clark, J. R. Williams, C. A. Jones, R. A. Galbraith, C. M.Danforth, and P. S. Dodds. Sifting robotic from organic text: A natural language approach for detecting automation on twitter. Journal of Computational Science, 16:1 – 7, 2016.

[9]S. Yardi , D. Romero, G. Schoenebeck, and D. Boyd. Detecting spam in a twitter network. First Monday, 15(1-4), January 2010.

[10]A. H. Wang. Don't follow me: Spam detection in twitter. In Security and Cryptography (SECRYPT), Proceedings of the 2010 International Conference on, pages 1–10, 2010.

[11] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou. 6 million spam tweets: A large ground truth for timely twitter spam detection. In IEEE ICC 2015 - Communication and Information Systems SecuritySymposium (ICC'15 (11) CISS), pages 8689–8694, London, United Kingdom, June 2015.

[12]C. Chen, J. Zhang, Y. Xiang, W. Zhou, and J. Oliver. Spammers are becoming smarter on twitter. IT Professional, 18(2):14–18, Mar.-April. 2016.

[13]J. a. Gama, I. Zliobait˙e, A. Bifet, M. Pechenizkiy, and A. Bouchachia.A survey on concept drift adaptation. ACM Comput. Surv., 46(4):44:1 44:37, Mar. 2014.

[14] K. Huang, Z. Xu, I. King, M. Lyu, and C. Campbell. Supervised selftaught learning: Actively transferring knowledge from unlabelled data.In Neural Networks, 2009. IJCNN 2009.International Joint Conference on, pages 1272–1277, June 2009.

[15] R. Jeyaraman. Fighting spam with botmaker. Twitter Engineering Blog,August 2014.

[16] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: socialhoneypots + machine learning. In Proceedings of the 33rd internationalACM SIGIR conference on Research and development in information retrieval, SIGIR '10, pages 435–442, New York, NY, USA, 2010. ACM.

[17] S. Lee and J. Kim. Warningbird: A near real-time detection system for suspicious urls in twitter stream. IEEE Transactions on Dependable andSecure Computing, 10(3):183–195, 2013.

[18] J. Oliver, P. Pajares, C. Ke, C. Chen, and Y. Xiang. An in-depthanalysis of abuse on twitter. Technical report, Trend Micro, 225 E. JohnCarpenter Freeway, Suite 1500 Irving, Texas 75062 U.S.A., September2014.

[19] A. Comparatives.Whole product dynamic real-world protection test Technical report,AVComparatives,http://www.avcomparatives.org/wpcontent/uploads/2016/07/avc prot 2016a en.pdf, July 2016.

[20] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In Proceedings of the 26th Annual Computer SecurityApplications Conference, ACSAC '10, pages 1–9, New York, NY, USA, 2010. ACM.

[21]Dheeraj Pal,Alok Jain,Aradhana Saxena and Vaibhav Agarwal,"Comparing Various Classifier Techniques for Efficient Mining of Data",Proceedings of the International Congress on Information and Communication technology,pp.191-202,2016

[22]S.Dinh,T.Azeb,F.Fortin,D.Mouheb and M.Debbabi,"Spamcampaig detection,analysis,investigation",vol.12,pp.S12-S21,2015.