IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 4, August 2017
ISSN (Online): 2277-5420
www.IJCSN.org
Impact Factor: 1.5

A Survey on Spam Detection Methodologies in Social Networking Sites

[1] K Subba Reddy; [2] Dr E Srinivasa Reddy

[1] Research Scholar, Computer Science and Engineering
Acharya Nagarjuna University, Guntur , AP, India

[2] Principal, Anucet, Computer Science and Engineering
Acharya Nagarjuna University, Guntur, Ap, India

Abstract - Conventional media, such as television or newspapers, essentially transmits information in one direction. Social media is a two-way form of communication that allows users to interact with the information being transmitted. Social media encompasses a wide variety of online content, from social networking sites like Facebook . online social networks are becoming popular among internet users. The internet users spend more amount of time on popular networking sites like Facebook, Twitter, google+ etc. Huge information available on these sites attracts the spammers who misuse the valuable information on these sites. Spammers send unwanted messages, share malicious links, develop malicious apps and sometimes create fake accounts. A lot of research has been done to detect spam on social networking sites. In this paper we have reviewed different research papers on spam detection. Our study provides techniques used, dataset and accuracy of various spam detection methodologies.

Keywords - *Social Media, Facebook, Twitter, Spammer, Internet*

1. Introduction

Social media are computer mediated technologies that facilitate the creation and sharing of information, ideas, career interests and other forms of expression via virtual communities and networks. Social media use web based technologies, desktop computers and mobile technologies to create highly interactive platforms through which individuals, communities and organizations can share, co-create, discuss and modify user generated content posted online. Social media differ from paper based media or traditional electronic media in many ways including quality, reach, frequency, usability, immediacies and permanence. Various familiar social networking sites are Facebook, Twitter, LinkedIn, sinaweibo, YouTube, WhatsApp, instagram, Skype, pinterest, snapchat etc.

Facebook is an online social media and online social networking site. Facebook can be accessed by desktops, laptops and smart phones over the internet and mobile networks. The users registered with the social site and then create user profiles. Users can add other users as friends, exchange messages, post status updates, digital photos, share digital videos, links and also receive notifications when others update their profiles or make posts. Users may join common-interest use groups and users can complain about or unpleasant people. Facebook has more than 1.86 billion monthly active users as of December-31,2016. Facebook was the most popular social networking site based on number of active user accounts. Twitter is an online news and social networking service. In this media users interact with tweets. The tweets are restricted to 140 characters. Registered users can post tweets but unregistered users can only read the tweets. The users can access Twitter through website interface, SMS or mobile App. Twitter has more than 319 million monthly active users. Tweets are publicly visible by default, but sender can restrict message delivery to just their followers. Users may subscribe to other users tweets. Individual twe

ets can be forwarded by other users known as retweet. Users can like individual tweets, can update their profiles via smart phones.

Social spam is unwanted content appearing on social networking services. The social spam can be in many ways including bulk messages, profanity, insults, hate speech, malicious links, fraudulent reviews, fake friend's etc.

Social networking experts estimate that 40% of social network accounts are used for spam. The spammers can utilize the social network tools to target certain segments, fan pages or groups to send embedded links to pornographic or other product sites designed to sell something from fraudulent accounts. So, spam detection is the very critical step in social networks. Spam can be detected by user based, content based or relation based techniques. Many research papers have been published to detect spam on social network sites. In this paper, we have done a survey of research papers. Our paper aims to define the various types of techniques used to detect spam in various social networks. This paper also aims to give a review of how these techniques have been

implemented by various researchers.

2. Literature Survey

Doaa Hassan studied a methodology to detect spam Emails. Nowadays Emails have been an easy and fast tool of communication among people. Spam Emails have been a huge problem that spreads widely on the Internet. They are represented as unsolicited (junk) messages that are sent to a large number of users. In this methodology text clustering and classification have been used to identify spam Emails. The basic idea of text mining is to break the documents into words and then treat each word as an attribute in the feature vector of machine learning model. The term weight W, refers the number of occurrences of a specific word in the document.

In this methodology all Email documents in the dataset are parsed to calculate the frequency words and spam or non spam mails. With using these words frequency table is constructed. To measure word frequency, used TF*IDF weighting scheme. This methodology also used stop list removal mechanism on the word frequency table to remove generic words like I,am,for,is etc, and also applied stemming removal mechanism to reduce words into their root forms by removing prefix, suffix and infixes. The word frequency table is represented as Email matrix that contains class label of each training sample. This Email matrix is an input to the clustering or classification algorithm.

In this methodology used K-means clustering algorithm on Email matrix to divide email dataset into two clusters ie spam or non spam clusters. The original email matrix is augmented with an extra feature called cluster. This feature has two values called cluster1 to indicate non spam emails and cluster2 indicate spam emails.

The classifier is trained and tested with expanded email matrix. On this model 10 fold cross validation is applied for training and testing. For each validation 90% of random dataset as selected for training and remaining 10% dataset as selected for testing. In this approach they had used WEKA free data mining software for Email text pre-processing and performing email text mining with clustering and classification. For each trail they have been build a learning model that combines k means clustering algorithm and various classifier algorithms including Naive Bayes, support vector machines, logistic regression and decision tree algorithms. The conjunction of logistic regression(LR) with clustering(C-LR) outperforms using LR alone with an average overall accuracy is 93.99% for C-LR and 93.79% for LR.

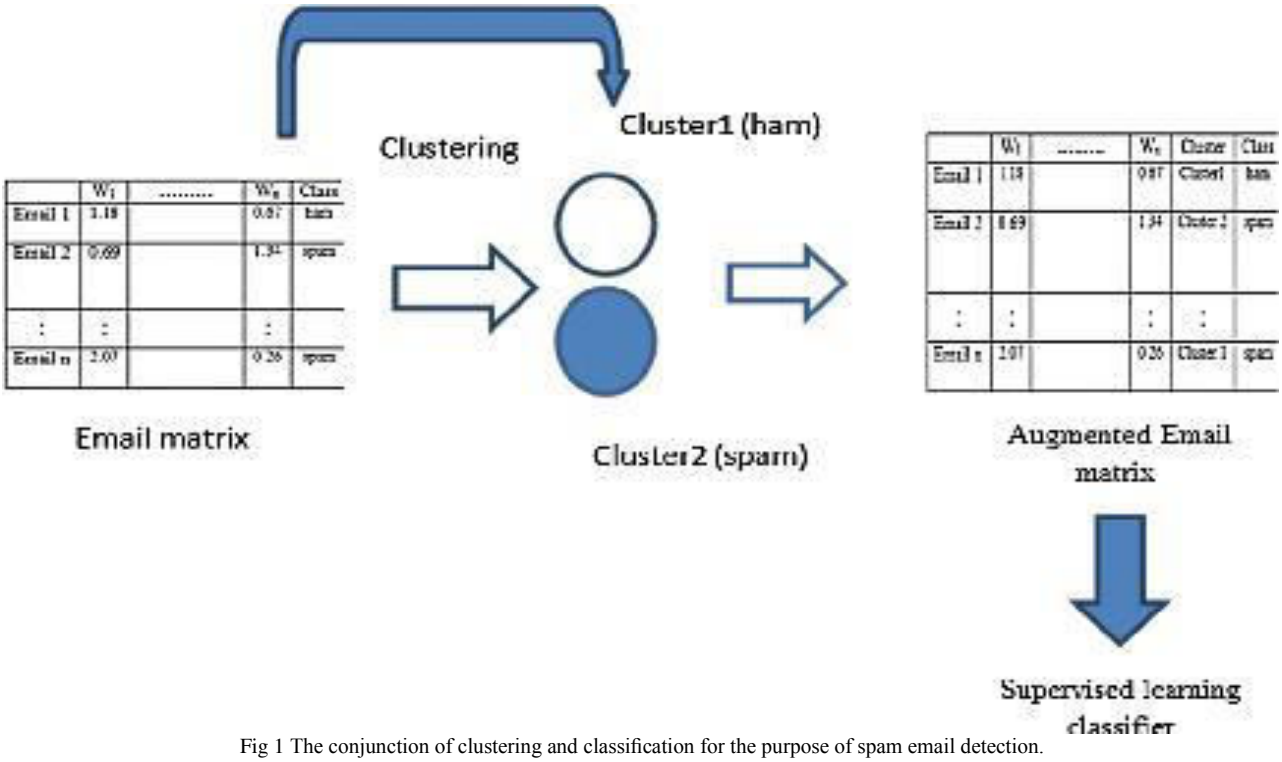

Fig 1 The conjunction of clustering and classification for the purpose of spam email detection.

The average time required to construct a classification model with clustering is 243.39 and without clustering is 25.51 for linear regression model (LR). If the original email dataset size is very large then the model construction time is also high. In this methodology used only k means clustering algorithm for conjunction..

Arushi Gupta et al proposed a methodology to detect spam in Twitter social network. Online social networks are most interactive platforms used by the users to communicate and share information. Twitter is one of

the most familiar online website used for information sharing. Twitter information is publically available, which can be accessed through API's provided by Twitter. Twitter messages are also called tweets. Tweet messages are small in size and restricted to 140 characters. Twitter is continuously under attack by spammers. These spammers spread malicious messages, malicious links and advertisements through tweets to the normal users. The authors have been proposed a methodology to detect spammers in Twitter social network. In this methodology they have been used 1064 users data. The user's dataset comprises of 62 features and these features containing user specific information and tweet specific information. To construct spammer detection model in twitter network used Naive bayes, clustering and decision tree learning algorithms. To get highest spam detection accuracy combined these three learning algorithms in the model.

In their proposed approach first identify the various features like followers, followees, URL's, spam words, Replies and hash tags. These features are used to detect spam accounts in twitter dataset. Manually all user accounts are represented as spammers and non spammers. In pre-processing step all continuous features are converted into discrete features. In Naive bayes approach user accounts are classified as spammers and non spammers by calculating the probability of user account. Clustering is an unsupervised learning approach, based on similar feature values the entire dataset is classified as spammer or non spammer classes. In decision tree learning approach, a decision tree structure was prepared and the decision was made at every level of tree to classify data set as spam or non spam dataset. To improve the accuracy of spam detecton these three approaches are integrated. The integrated approach was identifying the given dataset as spammer or non spammer with 87.9% accuracy.

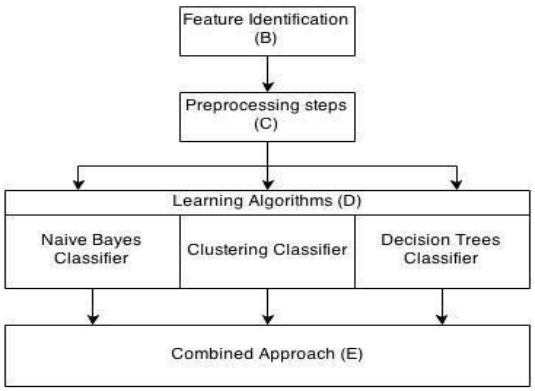

Fig 2.Proposed Spam Detection Approach

Himanshi Agrawal et al. proposed a methodology to detect relatedness between Facebook public page posts and comments. On online social networks spammers share malicious link looking like genuine one, place discount messages on their wall. Intruders use the portal for spreading rumours and overload the forums with off-topic comments. When readers are only interested in reading only on-topic information and unrelated comments create confusion. So it is very important to analyze the unrelated content on online social media. They had analyzed two public facebook pages categorized as entertainment website ie india-forum.com and non profit organisation ie Wikipedia. They had collected 225 posts and 607 comments from india-forum.com public pages, 783 posts and 6313 comments from Wikipedia public pages. Initially these posts and comments are simplified. For analysing these posts and comments used string similarity index formulas and corpus based similarity measures. String similarity index measures give index value between 0 and 1 and corpus based similarity measures give similarity index between -1 and 1. For analysing both measures similarities used precision formula. Precision value of result is number of non zero value index out of total values.

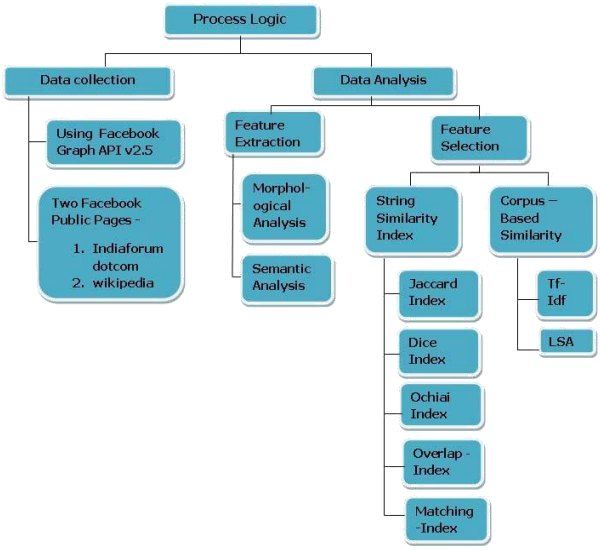

Fig 3. Process logic

Malik Mateen et al.studied an approach for spam detection in Twitter network. To detect spam in Twitter dataset used different kind of features like user based features, content based features and graph based features.user based features are based on users relationships and properties of user accounts. The spammers has to reach large number of profiles to spread misinformation. Different user account related features are Number of followers, Number of following, age of account, FF ratio and reputation. Content based features are related to tweets posted by user. Different features are total number of tweets, hash tag ratio, URL's ratio, mentions ratio, tweet frequency and spam words. Graph based features are used to identify spammer behaviour. Different features are in/out degree

and betweenness. In the proposed methodology used Twitter dataset consist of 10,256 users and 467480 tweets. To develop a spam detection model used J48, decorate and NaiveBayes classifiers. These three classifiers are individually trained on various dataset features and classify the dataset as spam or ham dataset. Out of these three classifiers J48 classifier highest accuracy to classify the data as spam or non spam. Content based features are best suitable for classifying the dataset. To classify the dataset with highest accuracy combine the content, user based and graph based features. The combined feature set is given as input to the three classifiers. But decorate and J48 classifiers have given highest accuracy upto 97.6%.

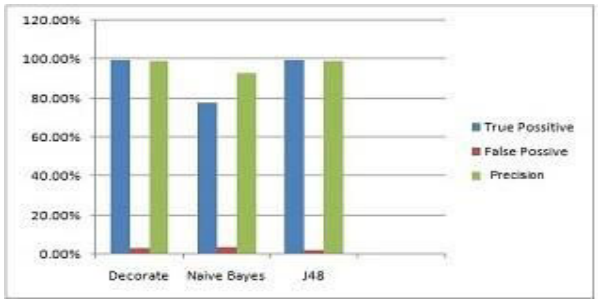

Fig 4. Classification results using user based features, content based features and graph based features

Hailu Xu et al. studied a methodology to detect spam across online social networks. This methodology focuses on combining spam in one soial network to the another social network. They had used 1937 spam tweets and 10942 ham tweets and 1338 spam posts and 9285 ham posts. In TSD, out of 1937 spam tweets, 75.6% spam tweets contained in URL links, 24.4% spam tweets contained in words. From 10942 ham tweets, 62.9% tweets are in URL links and words, remaining 37.1% consist of only words.

For the spam posts of FSD, 32.8% spam posts consists of URL links and words, 67.2% of spam posts consist of words. For ham posts 95.1% consist of URL links and 4.9% only consist of words. They had used top 20 word features from Twitter spam data and Facebook spam data. They had split the TSD and FSD into training and test data sets .The training and test data sets of TSD, FSD are used to train and test various classifiers like Random forest, logistic, random tree, BayesNet, Naive bayes.

After analyzing the classifiers accuracy, then combine the spam of facebook dataset into twitter training dataset and spam of Twitter dataset into Facebook training dataset. The combined dataset is used to train the classifiers and test the classifiers. Finally compare the results of classifiers on these two social networks. To measure the performance of classifiers used accuracy , precision ,recall and FM measures. The combined approach has given more accuracy in spam detection

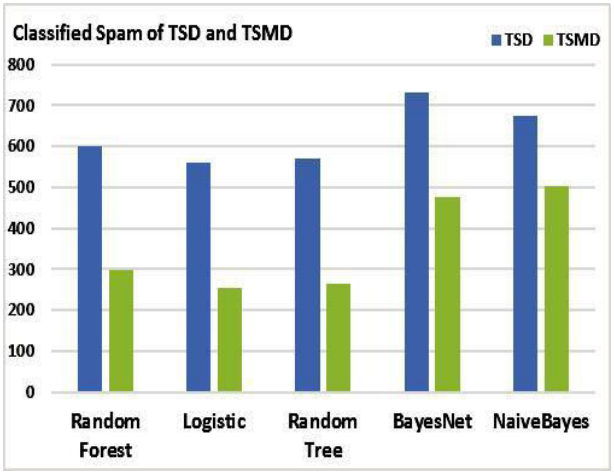

Fig 5 . Number of Classified Spam in TSD and TSMD

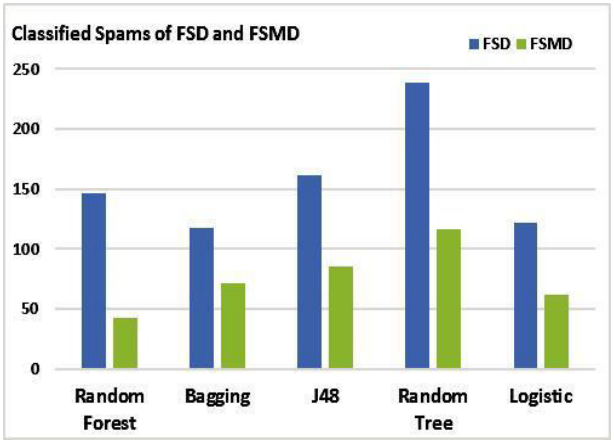

Fig 6 Number of Classified Spam in FSD and FSMD

Xianghan Zheng et al. described a procedure to detect spammers in social networks. Social network users spend plenty of time on social networks to interact with friends. These social networks also attracted by many of abnormal users called spammers. These spammers post the malicious information, advertisements etc in social networks. In this methodology used sina weibo social network and support vector machines (SVM) algorithm to detect spammers. To develop a model they had used 16 million messages from various users in weibo social network. In this model 18 features are used to construct a feature vector. The network users are classified as spammers and non spammers by manually. From the labelled dataset, 80% spammers and non spammers are selected randomly as training dataset and remaining dataset considered as test dataset. The network users behaviour is analysed with content based features and user based features. The feature vector dataset is given as input to the model for training. To gain highest spam

detection accuracy of this model used 1:2 ratio between spammers and non spammers of training dataset. With this 1:2 spammer to non spammers ratio the model classify the dataset with 99.5% spammers accurately and 99.9% non spammers accurately.

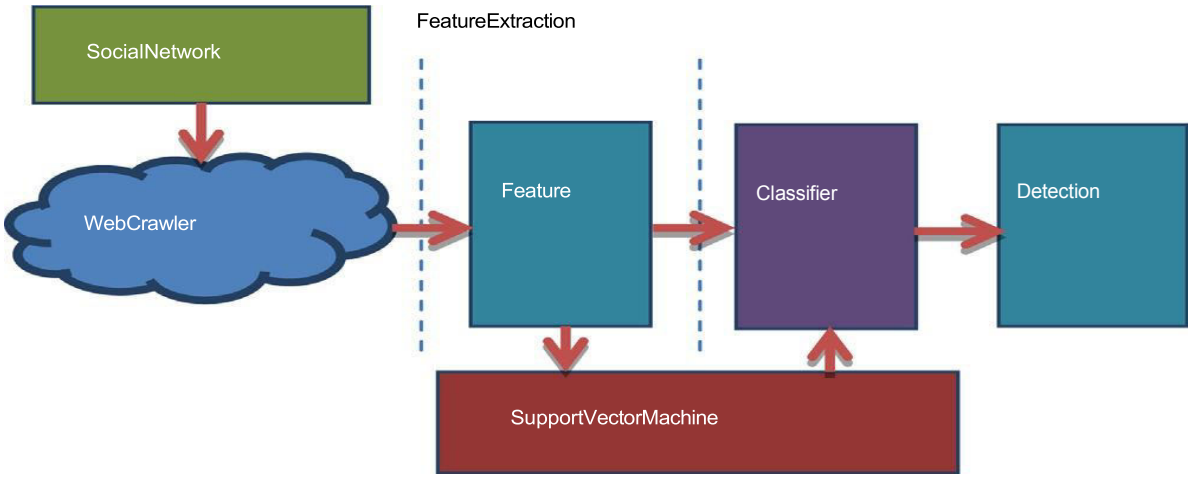

Fig 7.used methodology

3. Outline of Spam Detection

Sno	Title	Methodology	Type	Dataset	Result
1	Investigating the effect of combining text clustering with classification on improving spam email detection	K means clustering algorithm and classification algorithms ie NaiveBayes, support vector machines,logistic regression,decision tree	Content based	17171 spam mails,16545 non spam mails	K means clustering with logistic regression Has best model
2	Analysis of text mining techniques over public pages of Facebook	Compared string similarity indexs ie jacard index, dice index, ochiai index, overlap index, string matching index,LSA and TF-IDF	Content based	1008 posts and 6920 comments.	LSA has highest preision
3	A hybrid approach for spam detection for Twitter	Compared J48, Decorate and Naive Bayes classifiers	Content based,user based and graph based	10256 users and 467480 tweets	J48 and decorate has 97.6% precision
4	Efficient spam detection across online social networks	Compared random forest,logistic, random tree,bayesnet and naivebayes	Content based	1937 spam tweets,10942 ham tweets and 1338 spam posts,9285 ham posts	Random forest has the highest accuracy
5	Detecting spammers on social networks	Support vector machines(SVM)	Content based feature and user based features	30116 user accounts	99.9%
6	Improving spam detection in online social networks	NaiveBayes,clustering and decision trees algorithm	User based and content based	1064 twitter users data	Integrated approach has given 87.9% accuracy

4. Conclusion

From the research papers reviewed it can be concluded that spam detection techniques are categorized based on what attributes they used. These techniques can identify the spam users or spam messages. From these papers we have been observed that SVM model can identify both spam users and spam messages in better way. In future work we are going to proposed an efficient methodology to detect spam users and spam messages in social networking sites.

References

[1] De Wang, Danesh Irani, and Calton Pu. A Social-Spam Detection Framework. *CEAS 2011 - Eighth annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference* Sept. 1-2, 2011.

[2] Benjamin Markines,Ciro Cattuto,Filippo Menczer. Social Spam Detection**.** AIRWeb '09, April 21, 2009

[3] Enhua Tan1, Lei Guo2, Songqing Chen3, Xiaodong Zhang2, and Yihong (Eric) Zhao UNIK: Unsupervised Social Network Spam Detection

[4] Kyumin Lee, James Caverlee, Steve Webb:Uncovering Social Spammers: Social Honeypots + Machine Learning SIGIR'10, July 19–23, 2010, ACM

[5] 5.Xin Jin, Cindy Xide Lin, Jiebo Luo, Jiawei Han, SocialSpamGuard: A Data MiningBased Spam Detection System for Social Media NetworksThe 37th International Conference on Very Large Data Bases,August 29th September 3rd 2011,*Proceedings of the VLDB Endowment,* Vol. 4, No. 12

[6] Xueying Zhang, Xianghan Zheng, A Novel Method for Spammer Detection in Social Networks, IEEE-2015

[7] Yin Zhuy, Xiao Wang, Erheng Zhongy, Nanthan N. Liuy, He Li, Qiang Yang: Discovering Spammers in Social Networks, Association for the Advancement of Artificial Intelligence (www.aaai.org).,2012

[8] Jyotika varma, Dr sanjeev Dhawan: detection of spam in social networks using clustered k-nearest neighbour,international journal of advanced research in computer science and software engineering,ijarcsse,volume 5,issue 3,march 2015

[9] Gianluca stringhini,Christopher kruegel,Giovanni vigna,detecting spammeres on social networks:ACM-2010

[10] Faraz ahmed,Muhammad abulaish: An MCL Based Approach for spam profile detection in online social networks.IEEE 11th international conferene on trust,security and privacy in computing and communications-2012

[11] 11.Chengo cao,james caverlee: detecting spam URLs in social media via behavioural analysis, springer international publishing pp 703-714,2015

[12] 12 .Saini Jacob soman, Dr s murugappan; detecting malicious tweets in trending topics using clustering and classification, ieee international conference on recent trends in information technology -2014

[13] 13. Fabricio benevenuto, Gabriel magno, tiago rodrigues,and virgilio almedia: detecting spammers on twitter-ceas electronic messaging, antiabuse and spam conferene july 13-14,2010

[14] 14.Sajid Yousuf Bhat1, Muhammad Abulaish, Abdulrahman A. Mirza:Spammer Classification using Ensemble Methods over Structural SocialNetwork Features, 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)

[15] 15.Hailu Xu,Weiqing sun,Ahmad javaid: Efficient spam detection across online social networks,IEEE-2015

[16] 16.Arushi gupta,rishabh kaushal: Improving spam detection in online social networks-IEEE-2015

[17] 17.M soirayan, S Thanalerdmongkol,c chantrapornchai :using a data mining approach :spam detection on facebook, International Journal of Computer Applications (0975 – 8887) Volume 58–No.13, November 2012

[18] 18. Krishna Chaitanya T, HariGopal Ponnapalli, Dylan Herts and Juan Pablo:: Analysis and detection of modern spam techniques on social networking sites 2012 Third International Conference on Services in Emerging Markets IEEE

[19] 19. Xianghan Zheng a,b, ZhipengZeng a,b, ZheyiChen c, YuanlongYu a,b,n, ChunmingRong: Detecting spammers on social networks, Neurocomputing159(2015)27–34

[20] 20.De wang :Analysis and detection of low quality information in social networks, IEEE-2015

WEBSITE REFERENCES

1. https://en.wikipedia.org/wiki/Social_media
2. https://en.wikipedia.org/wiki/Facebook
3. https://en.wikipedia.org/wiki/Twitter
4. https://en.wikipedia.org/wiki/Spamming