Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 32 – 41

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Twitterati Identification System

Winnie Main[a]*, Dr. Narendra Shekokhar[b]

[a]*P.G Student, Dwarkadas J Sanghvi College of Engineering, Mumbai, India*
[b]*Head of Computer Engineering Dept. , Dwarkadas J Sanghvi College of Engineering, Mumbai, India*

Abstract

Twitter is an online service playing dual roles of social networking and micro blogging. Communication with other twitter users is carried out by publishing text and media based posts called tweets. Lately, Twitter has attracted a large number of automated programs, known as bots. Generally bots are used to generate a large amount of benign tweets delivering news and updating feeds, whereas some bots are being created to spread spam or malicious contents. To assist human users in identifying who they are communicating with, this project focuses on the classification of human and bot accounts on Twitter. We collected twitter statistics of a number of twitter users, their tweets, bot tweets, features, characteristics, etc. The data is then analyzed based on statistics to create a known training set of bots and humans. The proposed classification system uses a number of twitter attributes where every stage of the system makes a decision about the users of Twitter. Based on the statistical training data a decision tree is generated. Rules are formed using the decision tree to detect the user of twitter as a human or a bot. The various properties based on twitter features help distinguishing a human from a bot are discussed and implemented in this paper. Based on the results obtained it can be concluded that the more number of attributes, better is the detection mechanism. The statistical training data set is consistent for varying sizes of the test data.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords:- Bots; Phishing; Mining; Tweets; Twitter; Spam.

* Corresponding author. Tel.: 919987543430; fax: +0-000-000-0000 .
*E-mail address:*winniemain@gmail.com, nshekokhar@yahoo.co.in

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.076

1. Introduction

Twitter is a free micro blogging service founded in 2006 [1]. Remarkable simplicity is its distinctive feature. At its heart are 140-character bursts of information called tweets. Users can incorporate links to different content in their tweets, and tweet broadcasts can be public or private. Public figures like celebrities, statesmen, journalists, and others have established significant followings on Twitter. Media outlets mainly use Twitter as a way to broadcast breaking news. Twitter has been described as the "SMS of the Internet".

Twitter users have to adhere to certain formats to be able to publish tweets. Hashtags, namely words or phrases prefixed with a # symbol, helps group tweets by topic. For example, #ElectionsInIndia and #T20. The symbol @ followed by a username is used to mention or directly post replies to users. Twitter's user relationship is direct and consists of two ends, friend and follower [3]. From the standpoint of information flow, tweets flow from the tweeter to the follower. More specifically, when a twitter user posts tweets, these are displayed on the twitter user's timeline as well as those of his followers.

The growing popularity of twitter has made it vulnerable to exploitation by automated programs called bots. Bots are a common feature used in most web applications and are mainly used to assist humans to be able to do work more efficiently. The bots keep posting news feeds on the twitter page of the channel and it is programmed to do so at set time intervals. However bots may be exploited to carry out malicious attacks. The legitimate bots may be used to generate a large number of benign and junk tweets which may cause hindrance to twitter becoming a news publishing site. Also along with legitimate bots we may have malicious bots which may be used by spammers to spread spam, carry out phishing attacks etc. Therefore it becomes very important to address the problem of menacing bots so as to ensure that the users of twitter or twitterati enjoy a bot free tweeting experience.

The paper has been organized into the following sections. Section II covers related work that has been carried out on the topic. Section III proposes the technique for classification of twitterati. Section IV gives a detailed insight on the experimental results. Finally, Section V concludes the paper and its related future scope.

2. Related Work

The popularity of Twitter has naturally resulted in a number of studies. Studies have been carried out focusing on the social networking aspects as well as those evaluating twitter as a platform for dissemination of information.

Chao Yang et al [5], discusses how most analysts concentrate on creating Bot detection mechanisms. However, spammers these days are intelligent enough to bypass the mechanisms in place, Chao Yang et al makes a comprehensive and empirical analysis of the evasion tactics utilized by Twitter spammers. 24 new detection features have been proposed by the author. Through experiments carried out, these new features are found to be more effective on the spammer community. Zi Chu et al [4], proposes a classification system to identify Twitter accounts as either human or bots. A large scale user account set has been generated and various tweeting behaviours, content and properties have been analyzed. The combination of features is used to determine the account type. The tests carried out with the classification model have been found to be efficient in tackling the bot problem among Twitter users.

Steven Gianvecchio et al [6] first conducts a series of measurements on a commercial chat network. These measurements then capture and identify 14 different types of Bots. A classification system is proposed based on the measurements to separate out the humans from the bots. The entropy-based classifier is more accurate to detect unknown chat bots, whereas the machine-learning-based classifier is faster to detect known chat bots. The tests are found to be successful in differentiating humans from bots. Ke Tao et al [14] studies the unique problem of duplicate tweet detection. Most intelligent bots now modify tweets before reposting them. Their proposed framework compares the posted tweets.

3. Proposed Solution

In this section the technique for classification of twitterati has been proposed. The system classifies twitterati into humans and bots. The system uses different twitter attributes for effectively carrying out the classification. The system consists of the components Inter Tweet Delay (ITD), Spam Detection (SPD), Near Duplicate Tweet, Klout score, Tweeting Device and the final decision maker algorithm. The design of our proposed twitterati classifier is shown in Fig 1. We start by extracting tweets of users by creating a twitter API. The tweets are then filtered per user basis. The different features considered for classification of users are given below.

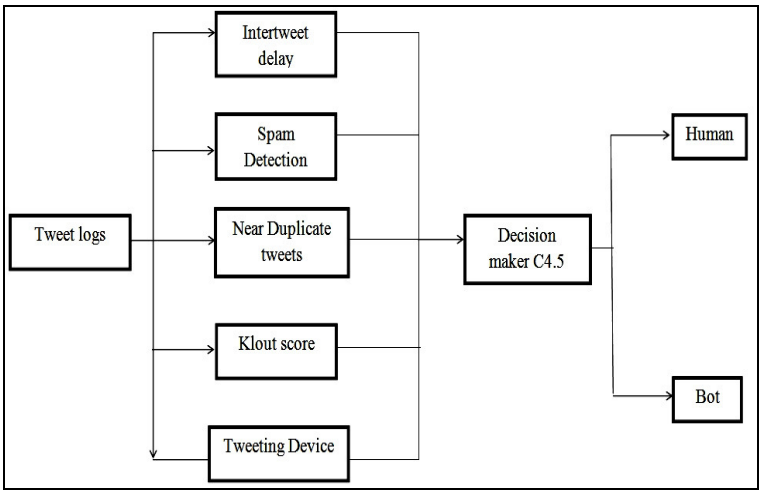

Fig.1. Twitterati Classification System

3.1. Inter Tweet Delay (ITD)

Inter tweet delay is one of the methods to identify bots. As the same suggests, it is the delay between consecutive tweets posted by a twitter user. Most bots use an application to post tweets. These applications generally allow the bot to select and schedule the day and time to post these tweets. These schedules are set to repeat at specific times each day. As a result, we can that users posting tweets with a fixed or fairly standard inter tweet delay are more likely to be bots.

To compute the inter tweet delay we compute the entropy for a given user's tweets. The entropy rate is a measure of the complexity of a process. A random process X= {Xi} is defined as a sequence of random variables. The entropy of such a sequence of random variables is defined as

Where, $$H(X1,....,Xm) = -\sum_{i=1}^{m} P(xi)\log P(xi) \quad (1)$$

$P(xi)$ is the probability $P(Xi = xi)$ (2)

A high entropy value indicates the user is a human whereas a low entropy value indicates the user is a bot.

3.2. Spam Detection (SPD)

Spam is sending unsolicited or irrelevant messages sent over the Internet. Typically, spam is sent by a large numbers of users, for the purposes of advertising, phishing, and spreading malware. One of the most common ways of disrupting a twitter user's experience is by posting spam. Twitter users indulging as spammers are considered to be bots with high spam content in their tweets. Google spam words are generally used references for classifying words as spam. There could be instances of a legitimate tweet containing a spam keyword. In this case, the application platform will need to be learned enough to make that judgment, while also considering other twitter user attributes.

To carry out spam detection we use the Bayesian Spam Classification Technique. We filter out particular users tweets. These tweets are then checked for their content by comparing the words in the tweets with known spam and ham data sets and computing the probability of occurrence of the word in a spam set. Based on the probability of occurrence of spam words in a tweets we can conclude whether the user is a spammer or not. Let's suppose the suspected tweet contains the word "lottery". Most people who are used to receiving tweets know that a particular tweet is likely to be spam, more precisely a proposal to lure people and trap them in to making money by winning lottery at the cost of divulging their personal details. The spam detection component, however, does not "know" such facts; all it can do is compute probabilities. The formula used by the component to determine that is derived from Bayes' theorem.

$$\Pr(S \mid W) = \frac{\Pr(W \mid S) \cdot \Pr(S)}{\Pr(W \mid S) \cdot \Pr(S) + \Pr(W \mid H) \cdot \Pr(H)} \tag{3}$$

where

$\Pr(S \mid W)$ is the probability that a tweet is a spam, knowing that the word "replica" is in it.

$\Pr(S)$ is the overall probability that any given tweet is spam.

Pr(W|S) is the probability that the word "replica" appears in spam tweets.

$\Pr(H)$ is the overall probability that any given tweet is not spam (is "ham").

$\Pr(W \mid H)$ is the probability that the word "replica" appears in ham tweets.

Recent statistics [15] show that the current probability of any message being spam is 80%, at the very least $\Pr(S) = 0.8$, $\Pr(H) = 0.2$

3.3. Near Duplicate Tweets(NDT)

While twitter introduced measures to filter out duplicate tweets posted by users [13], intelligent bots have found another way. Bots replace a small part of the earlier tweet to make it seem like a new tweet. Though well designed solutions can detect this small change, bots seems to come up with newer alternatives. The most common being that of replacing the mentioned user in a tweet with another mentioned user. The base content of the tweet remains the same. Other alternatives include [14] replacing same content with different hash tag or trending topics or mentioning random verified account with unrelated tweet contents.

3.4. Klout Score

Klout [7] is a popular website and mobile app that uses social media analytics to rank its users according to online social impact via the "Klout Score", which is a numerical value between 1 and 100. In determining the user score, Klout measures the size of a user's social media network and correlates the content created to measure how other users interact with that content. Klout measures influence by using data points from Twitter, such as following

count, follower count, retweets, list memberships, how influential the people who retweet you are and unique mentions. A user having higher Klout score e.g. 99 is supposed to be more influential than a user who has a Klout score of 5. Klout measures the user's influence online. This influence is measured primarily as the ability to drive others to action.

3.5. Tweeting Device

As per the twitter statistics [9], 77.3 % of tweets are sent from official Twitter Apps while the remaining 22.7% of tweets from third party apps. By allowing third-party developers partial access to its API, Twitter allows them to create programs that incorporate Twitter's services [10]. Once the manual authentication is allowed, automated programs are easily able to use these APIs to access twitter services. The easiest way for bots to use and misuse twitter is via the APIs. We consider this attribute while distinguishing between humans and bots.

3.6. Decision Maker

Table.1. Attribute Table

Attribute	Description	Data type	Possible values
Intertweet Delay	Delay between consecutive tweets of a single user	Numeric	0-1
Spam Detection	Contents of a tweet whether spam or not	Numeric	'SPAM', 'NOT SPAM'
Near Duplicate tweets	Similarity content between two tweets	Numeric	'YES','NO'
Klout score	Social impact or influence of the user	Numeric	1-100
Tweeting Device	Type of device used eg. API, phone, browser.	String	'API', ' NOT API'
Result	Final classification	String	'HUMAN','BOT'

The different attributes required for the system have been tabulated above. The list consists of the various attributes, their description, their data type and the possible values the attribute can contain. We evaluate user tweets and evaluate all attributes specified. A training set is provided to the weka decision maker taking into account different twitter statistics. To carry out the classification of human and bots we use the C4.5 classifier which comes from the decision tree family. The C4.5 technique is one of the decision tree families that can produce both decision tree and rule-sets; and construct a tree for the purpose of improving prediction accuracy [12]. This algorithm is an extension of the ID3 algorithm.

Algorithm C4.5 begins the process of learning through establishing a decision tree from above to below. The training data set T =t_1, t_2, t_3… consists of classified samples. Each sample t_i consists of a p-dimensional vector (x1, x2, x3…) in which x_j stands for attributes of the sample and the class in which t_i falls. In the decision tree, at each node, C4.5 chooses the best attribute that successfully divides its set of samples into subsets supplemented in one class or another. This dividing criterion is the normalized information gain, also known as difference in entropy. Hence the best selected attribute with the highest normalized information gain is then chosen to make the right decision. The C4.5 algorithm then follows this repeated procedure on the smaller lists.

The homogeneity of the data set across a certain attribute would first be tested by calculating its entropy, $E(A)$, given by the formula,

$$E(A) = \sum pi \log_2 pi \quad \ldots\left(i = 1,2\ldots.m \right) \tag{4}$$

Where pi is the list of probabilities of m values of the attribute A

The information gain is described as the function $Gain(A)$ which is showed as below

$$Gain(A,C) = E(A) - E(Aij) \quad \ldots (j = 1,2....v) \tag{5}$$

Where $E(Aij)$ is the list of entropies for j values of the child attribute.

The information gain is thus calculated for each attribute after which the attribute having the highest gain value is selected. This step is repeated continuously until a decision tree containing all the attributes is formed. Based on the decision tree constructed rules are formed which help take a decision to classify the user of twitter as human or bot.

4. Results

A training set is created taking into account different twitter statistics. Based on these statistics a data set of 8000 cases was generated. The classification system considers 8000 cases of training data to generate a decision tree and carry out the classification. This statistical training set considers all different components and takes into account statistics to carry out the final decision of the user as a human or a bot.

Five attributes in total have been considered to carry out the final decision. An experiment was carried out in which first a decision was taken considering the two major attributes that is the inter tweet delay and the spam detection component. And then all five attributes including the remaining three attributes i.e the duplicate tweets, klout score and tweet device.

The experiment was carried out using the weka tool. The total number of correctly classified instances to the total number of incorrectly classified instances is compared. The results were tabulated and it has been observed that though a decision can be carried out using the main attributes the decision making process improves its accuracy when supported with extra attributes. The results are as follows.

Fig.2 shows the performance comparison based on the number of attributes. The graph shows that there is a 30 % increase in the number of correctly classified instances when we use all 5 attributes whereas there is a 20% decrease in the number of incorrectly classified instances on increasing the number of attributes.

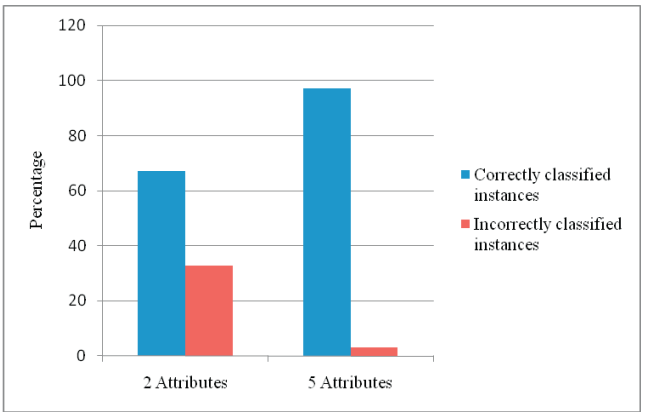

Fig.2. Performance comparison based on number of attributes

Experiments were carried out to check the efficiency of the system in detecting humans and bots by considering two v/s five attributes. The training data efficiency is measured based on three parameters namely

- False positives-The total number of cases that were false but incorrectly classified as true
- True positives-The total number of cases that were true and were correctly classified as true
- Precision-The appropriateness of the system to correctly classify the data.

The results obtained are described in Fig. 3, and Fig. 4.

4.1. For Human Detection

It is observed from Fig.3, given in that increasing the number of attributes improves the true positive rate and the precision and causes a drop in the false positive rate. This shows that there is an improved detection rate for a human and the possibility of him being falsely detected as bot is reduced.

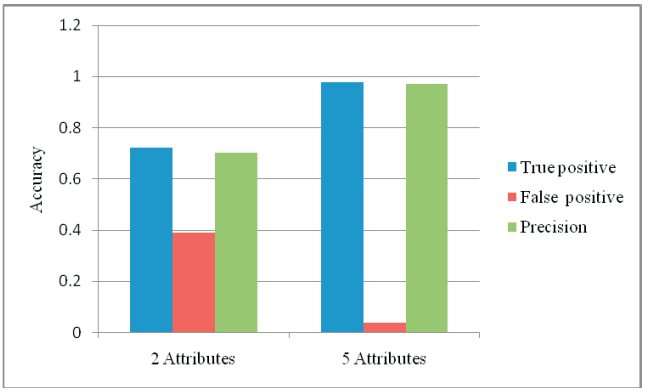

Fig.3. Accuracy comparison for human detection

4.2. For Bot detection

For detection of bots from the graph in Fig. 4, we see that the true positives obtained are close to one and false positives are close to zero when five attributes are used indicating better detection result for bots. Hence we can conclude from the graphs that more and better the number of attributes you use we obtain better detection and higher precision rates.

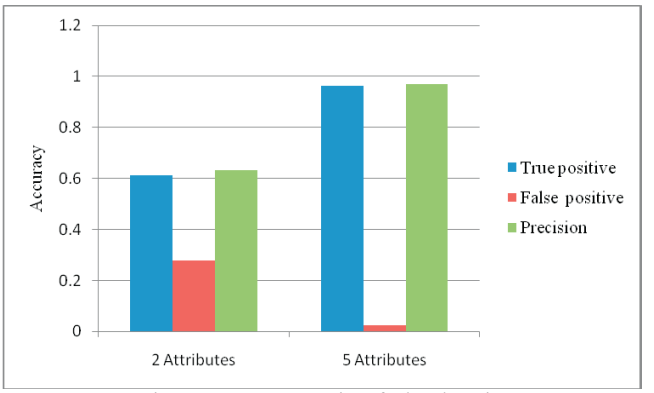

Fig.4. Accuracy comparison for bot detection

The second type of results we have obtained is by varying the size of the test data. We take variable sizes of the test data in proportion with the training data set. The different ratios of test to training data set are 1:8, 2:8 and 4:8. We check how well the test data can be classified using the generated training decision tree. This is done to check the accuracy, quality and efficiency of the training data set which aids in our final classification of human or bot.

Using the weka classifier we checked the accuracy of our training set by testing it with three different sizes of the test sets. Tests were carried out first for detection of bots and then for detection of humans.

For detection of Bots, it has been obtained that there is a very marginal increase in the True positives as seen in Fig. 5 and the precision by varying the test data size and a marginal decrease in the false positives has been observed in Fig. 6. So from the graphs we conclude that our training set works consistently for varying sizes of test data for detection of bots.

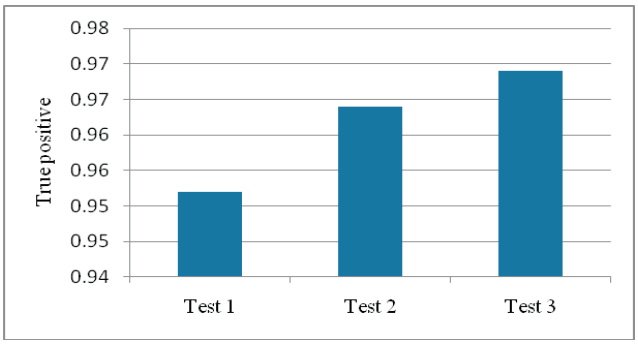

Fig.5. True positives for bot detection

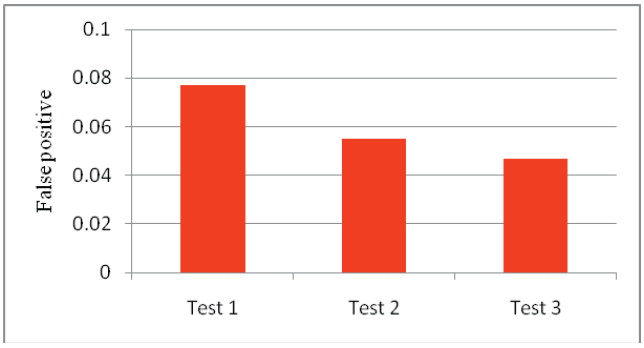

Fig.6 False positives for bot detection

For detection of Human as seen in Fig.7, it has been obtained that there is an increase in the True positives which is better than what has been obtained for bots as seen in Fig. 6, and the precision by varying the test data size and there is a decrease in false positives as seen in Fig. 8, which is better compared to Fig.6. So from the graphs we conclude that our training set works consistently for varying sizes of test data for detection of humans.

The findings of the results prove the efficiency of our statistical training data and the effectiveness of the Twitterati classification system.

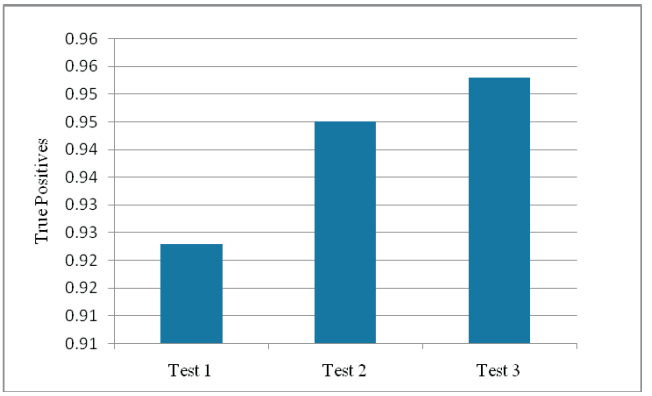

Fig.7. True positives for human detection

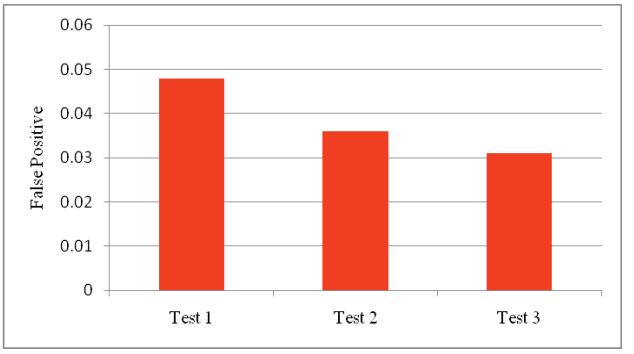

Fig.8. False positives for human detection

5. Conclusion & Future Scope

The popularity of Twitter has made it a target of malicious attacks. This combined with the fact that for most part it relies on open source authentication practices has caused automated programs to take advantage. These automated programs or bots cause massive amount of spamming and needs to be kept in check.

The paper classifies twitter users as either human or bots. A number of user behaviors are analyzed as attributes to make such a classification. A log analyzer stores the tweets to be analyzed. Bots which are automated programs repetitively keep posting malicious content. The Inter Tweet Delay (ITD) is a factor that the program utilizes. A number of traits of the user profile, the tweeting time intervals, account properties, etc are unique when identifying Humans from Bots. These traits are to be analyzed in a five component system. A decision maker finally judges a user profile to be either human or bot.

Bots are ever evolving in nature, learning new evasive techniques to counter detection algorithms. Previously designed algorithms need to be updated based on new and modified Twitter features and functionalities. A future scope of this paper would be to create a web crawler to gather twitter details and re-run our algorithm and re-check our observations for a larger sample size data.

References

1. Mashable.(Aug.2014).Twitter[Online].Available:http://mashable.com/category/twitter.
2. Statistic Brain. (Jun. 2014). Twitter Statistics[Online]. Available:http:// www.statisticbrain.com/twitter-statistics.
3. Alexa.(Jul. 2014). The top 500 sites on the web[Online]. Available http://www.alexa.com/topsites.

4. Zi Chu, Steven Gianvecchio, Haining Wang, Sushil Jajodia, "Detecting Automation of Twitter Accounts:Are You a Human, Bot, or Cyborg?," IEEE Transactions on Dependable and Secure Computing, Vol. 9, No. 6, November/December 2012.
5. Chao Yang, Robert Harkreader, and Guofei Gu,"Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers", IEEE Transactions On Information Forensics And Security, Vol. 8, No. 8, August 2013.
6. Steven Gianvecchio, Mengjun Xie, Zhenyu Wu, and Haining Wang, "Measurement and Classification of Humans and Bots in Internet Chat" Proceeding SS'08 Proceedings of the 17th conference on Security symposium.
7. Klout. (Aug. 2014). Be Known For What You Love[Online]. Available:https://www.klout.com.
8. Weka. (Apr. 2014). Data Mining with Open Source Machine Learnin[Online].Available:https://www.cs.waikato.ac.nz/ml/weka.
9. Twitter.(Jul. 2014). Twitter[Online].Available:https://www.twitter.com.
10. Twitter.(Jul. 2014).Twitter Fact Sheet [Online].Available: https://about.twitter.com/company.
11. Leonardo Fabrico.(Aug 2014). Data Mining Classification[Online]. Available:http://www.courses.cs.washington.edu/courses/csep521/
12. 07wi/prj/leonardo_fabricio.pdf.
13. C4.5 (Mar 2014). [Online].Available:https://ww.mgt.ncu.edu.tw/ ~wabble/School/C45.ppt.
14. Ke Tao, Fabian Abel, Claudia Hauff, Geert-Jan Houben, Ujwal Gadiraju "Groundhog Day: Near-Duplicate Detection on Twitter", World Wide Web Conference Committee (IW3C2). IW3C2 WWW 2013, May 2013, ACM 978-1-4503-2035-1/13/05.
15. Michal Prilepok, Jan Platos, Vaclav Snasel, Eyas El-Qawasmeh, "The Bayesian Spam Filter with NCD" CEUR Workshop Proceedings, Vol 837, Paper 18.
16. Twitter help centre.(Jun. 2014) Twitter Verified Accounts FAQs [Online].Available: https://support.twitter.com/articles/ 119135-faqs-about-verified-accounts.
17. Beevolve. (Oct. 2012).Twitter Statistics[Online].Available http://www. beevolve.com/twitter-statistics.
18. How stuff works.(Jul. 2014).Twitter programs for social networking [Online].Available http://computer.howstuff
19. works.com/ internet/ social-networking/networks/twitter2.htm.
20. Mashable.(Jun. 2009). 18 million twitter accounts in 2009 [Online].Available http://www.mashable.com/2009/09/14/twitter-2009-stats/.
21. Twopcharts(Jun. 2014).Number of registered Twitter accounts.[Online].
22. Available : http://www.twopcharts.com/twitteractivitymonitor.
23. Duplicate Tweet content (Dec.2010) [Online].Available https://www.tropo.com/2010/12/reminder-beware-of-duplicate-tweets-when-testing-twitter-apps-on-tropo/.