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. 5, Issue 7, July 2017

Machine Learning Approach For Spam Tweets Detection

Jagtap Kalyani Laxman[1], Prof.B.A.Khansole[2]

P.G Student, Department of Computer,Chh. Shahu College of Engineering, Kanchanwadi, Aurangabad, Dr. BAMU, Maharashtra, India. [1]

Professor, Department of Computer, Chh. Shahucollege of Engineering, Kanchanwadi, Aurangabad, Dr. BAMU, Maharashtra, India. [2]

ABSTRACT: Twitter has grown tremendously over the past few years. With sites such as Google, YouTube, Twitter and Facebook, amongst them twitter is ranked in the top 10 most visited sites [2]. In February 2009, twitter was the fastest-growing website with a growth rate of 1,382 per. [3]. In 2011, people sent about 140 million tweets per day and 460,000 new accounts were created per day [4]. The enormous growth of twitter allows many users to share their information and communicate with each other. Spammers have several goals, which are phishing, advertising, or malware distribution. These goals are similar to traditional spam in email or blogs, but twitter spam is different. Twitter limits the length of each message to less than 140 characters. Because of this limitation, spammers cannot put enough information into each message. To overcome this restriction, spammers usually send a spam containing URLs that are created by URL shortening services. When a user clicks the short URLs, he will be redirected to malicious pages. Since the messages are short and the actual spam content is located on external spam pages, it is difficult to apply traditional spam filtering methods based on text mining to twitter spam. In system there is an offline dataset of tweets which contain the 200 twitter user tweets. User dataset has some feature labeled attributes. Spam detection built the model which includes the binary classification and these issues is solving by machine learning approach. The machine learning algorithms such as Naïve Bayesian (NB) classifier or Support Vector Machine (SVM) classifier reported the behavior of models. System reported the impact of the data related factors, such as spam to non-spam ratio, timely tweets. The feature of implemented system is simple and time varying spam tweet detection. The system shows as spam detection is big challenge and it bridge the gap between the performance evaluations and mainly focus on the data, feature and model to identify the genuine user and report the spam user. The new contribution of this system is that real time tweets are captured and performance evaluation is carried out by NB and SVM classifier also by comparing their result it calculates theaccuracy.

KEYWORDS: Feature Extraction, Machine Learning, SPAM, Performance Evaluation.

I. INTRODUCTION

Twitter has grown tremendously over the past few years. With sites such as Google, YouTube, Twitter and Facebook, amongst them twitter is ranked in the top 10 most visited sites. In February 2009, twitter was the fastest-growing website with a growth rate of 1,382%. In 2011, people sent about 140 million tweets per day and 460,000 new accounts were created per day. The enormous growth of twitter allows many users to share their information and communicate with each other. Owing to the popularity of Twitter, malicious users often try to find a way to attack it. Most common forms of Web attacks, including spam, scam, phishing, and malware distribution attacks, have appeared on twitter. Because tweets are short in length, attackers use shortened malicious URLs that redirect twitter users to external attack servers. Spam is defined as the use of electronic messaging system to send unsolicited bulk messages. With the rise of OSNs, it has become a platform for spreading spam. Spammers intend to post advertisements of products to unrelated users. Some spammers post URLs as phishing websites which are used to steal user's sensitive data. Social networking sites such as Twitter, Facebook, Instagram and some enterprise of

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. 5, Issue 7, July 2017

online social network have become extremely popular in the last few years. Individuals spend vast amounts of time in OSNs making friends with people who they are familiar with or interested in. Twitter, which was founded in 2006, has become one of the most popular micro blogging service sites. Around 200 million users create around the 400 million new tweets per day the growth of spam. Twitter spam, which is referred as unsolicited tweets containing malicious links that directs victims to external sites containing malware spreading, malicious link spreading etc. has not only affected a number of legitimate users but also polluted the whole platform.

II. RELATEDWORK

A. Twitter Sentiment in Data Streams with Percepetron[6]

In 2014, Nathan Aston et al. analysed the huge increase in popularity of twitter in recent years, the ability to draw information regarding public sentiment from twitter data has become an area of immense interest. Numerous methods of determining the sentiment of tweets both in general and in regard to a specific topic have been developed.

B. Aiding the Detection of Fake Accounts in Large Scale Social Online Services[2]

OSNs suffer from abuse in the form of the creation of fake accounts, which do not correspond to real humans. Fakes can introduce spam, manipulate online rating, or exploit knowledge extracted from the network. OSN operators currently expend significant resources to detect, manually verify, and shut down fake accounts.

C. Spam Filtering in Twitter using Sender-Receiver Relationship[3]

Song et al. Extracted the distance and connectivity between a tweet sender and receiver to determine whether tweet is spam or not. System use distance and connectivity as the features which are hard to manipulate by spammers and effective to classify spammers.

D. Detecting Spammers on Social Networks[4]

In 2010, although there are few works such as, which uses content and account features such as account age, number of followers and followings, URL ratio and length of tweets to distinguish spammers and non-spammers, System then analysed the collected data and identified anomalous behaviour of users who contacted profiles. Based on the analysis of this behaviour, system developed techniques to detect spammers in social networks.

III. PROPOSEDALGORITHM

A. ***Naïve Bayes Algorithm:*** This is mainly used for filtering the spam tweets and also used in text classification. Naive Bays classifiers work by correlating the use of tokens (typically words, or sometimes other things), with spam and non-spamtweetsandthenusingBayes'theoremtocalculateaprobabilitythatatweetsisorisnotspam.

Step 1: Convert the data set into a frequency table.

Step 2: Create Likelihood table by finding the probabilities.

Step 3: Use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

$$P(A|B) = P(A|B) * P(A) / P(B) \quad \ldots\ldots\ldots (I)$$

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. 5, Issue 7, July 2017

B. ***Support Vector Machine Algorithm:*** This mainly helps in data classification. The classification step is built after the training process of tweets. Timely captured tweets also label in this classifier. This classifier gives the output in the form of 0, 1, and 2. The label 0 is mainly for true positive, 1 for true negative and 2 for neutrallabels.

Step 1: Maximize the margin of hyper-
plane Assume linear reparability
for now:
– In 2 dimensions, can separate by aline
– In higher dimensions, need hyperplanes
Can find separating hyper plane by linear programming (e.g. perceptron):
– Separator can be expressed as ax + by
=c Step 2: Function specify trainingsample.
Step 3: Quadratic programming
problem Step 4: Text classification
method

IV. RESULTS ANDANALYSIS

Experimental Setup:

A. **Software and Hardware:** The system has been implemented Java (JDK 1.7) with Eclipse JEE-Indigo-SR2-win64 and Apache-Tomcat-7.0.42. With JSP server serving java technologies. The system is tested on Intel(R) Core(TM) i3 2330M CPU @ 2.20 GHz and 4 GB RAM. The System uses My-SOL, an open source relational database management system that uses SQL for adding, managing and accessing content in adatabase.

B. **Dataset Used:** Standard dataset contain 1 million twitter user tweets from that we take 200 twitter user tweet data with feature attribute as tweet user name, nick name, tweet content, no of follower, no of following, country, place, tweet URL. For the NB and SVM classifier training dataset used which contain nearly 630 trained tweets with 336 spam and 295 non spam tweets. Also we used additional 200 tweets as a testing used for classify. For real time tweets are collected are stored in synthetic dataset and this data of user tweets used fortesting.

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. 5, Issue 7, July 2017

C. Results andAnalysis:

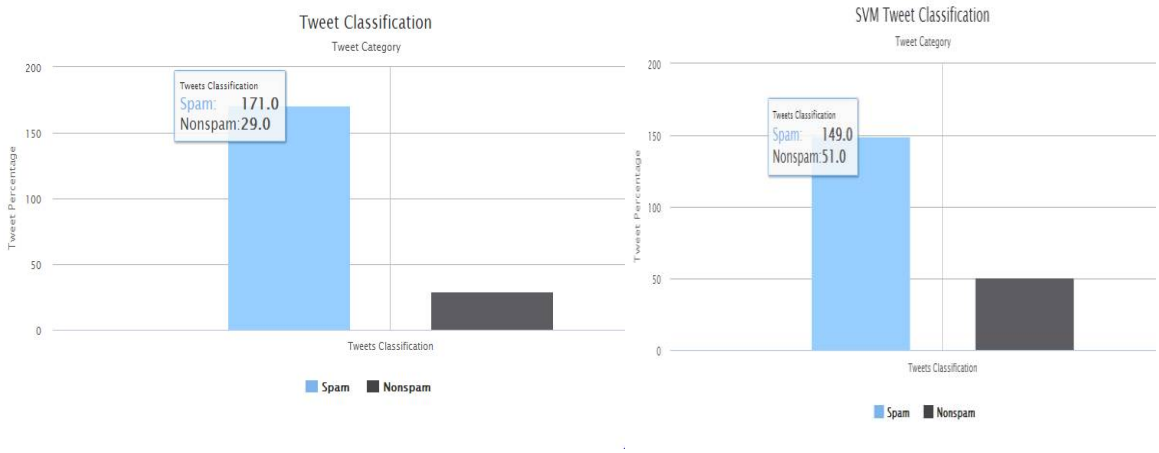

Fig: 1 Impact of Spam to Non spam ratio for NB and SVM

Above fig. shows impact of spam to non-spam ratio for NB and SVM classifier, For Naïve Bayes, it is 171:29 and for SVM, it is 149:51.

In following fig, NB classifier performance is shown by tabular value along with graph. The below table 1 contains details about all performance parameter of NB classifier like TPR, FPR, Precision, Recall, F-measure etc. on offline dataset of 200 user. Here, The NB classifier gives accuracy 90.7%.

Table:1 NB Classifier Performance

Performance Parameter	Parameter Value
TPR	85.5
FPR	79.6
Precision	60.2
Recall	85.5
F-measure	70.7
Accuracy	90.7

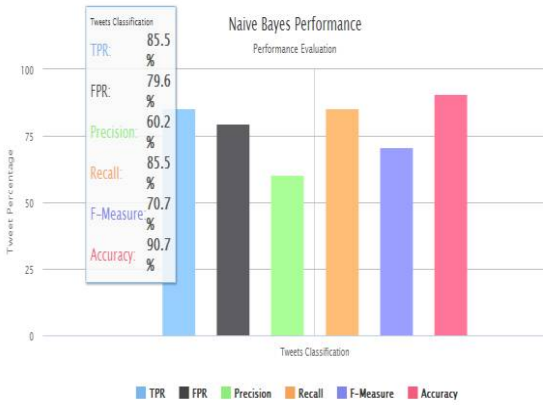

Fig: 2 Analysis Graph for Naïve Bayes

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. 5, Issue 7, July 2017

In following fig, SVM classifier performance is shown by tabular value along with graph. The below table 2 contains details about all performance parameter of SVM Classifier like TPR,FPR, Precision, Recall, F-measure etc. on offline dataset of 200 user. Here, The SVM classifier gives accuracy 79.4%.

Table:2 SVM Classifier Performance

Performance Parameter	Parameter Value
TPR	74.5
FPR	48.0
Precision	76.0
Recall	74.5
F-measure	75.3
Accuracy	79.4

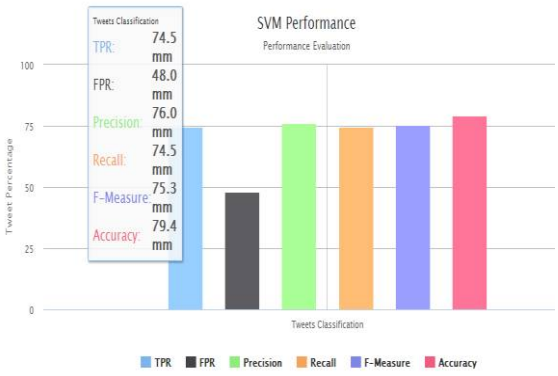

Fig: 3 Analysis Graph for SVM

The below table 3 contains details of comparison between NB and SVM with respect to performance parameter.

Table:3 Comparison table for both classifier

Performance Parameter	NB Classifier	SVM Classifier
TPR	85.5	74.5
FPR	79.6	48.0
Precision	60.2	76
Recall	85.5	74.5
F-measure	70.7	75.3
Accuracy	90.7	79.4

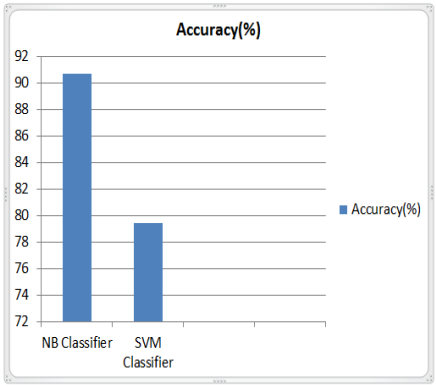

Fig: 4 Comparison graph

The below table contains details of NB parameter value and SVM parameter value in real stream tweet collections. Table 4 and 5 compare details of NB performance parameter values with SVM parameter value for 5 days.

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. 5, Issue 7, July 2017

Table 4: Day-wise estimation for NB Classifier

Days	D1	D2	D3	D4	D5
Spam	133	140	137	156	151
Non-spam	41	31	37	28	22
Precision	59.11	56.2	55.5	54.9	53.9
Recall	76.43	81.9	83.5	84.8	87.3
F-measure	66.7	66.7	66.7	66.7	66.7
Accuracy	84.6	88.9	90.1	91	92.7

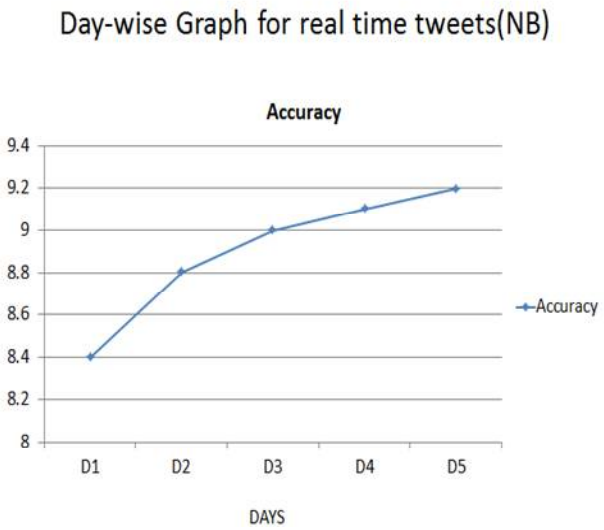

Fig: 5 Day-wise analysis graphs for NB classifier

Table 5: Day-wise estimation for SVM Classifier

Days	D1	D2	D3	D4	D5
Spam	135	128	117	104	133
Non-spam	50	46	40	33	41
Precision	79.41	78.0	76.0	73.23	61.86
Recall	72.97	73.6	74.5	75.91	76.43
F-measure	76.05	75.7	75.2	74.54	68.37
Accuracy	77.27	78.1	79.4	81.14	83.98

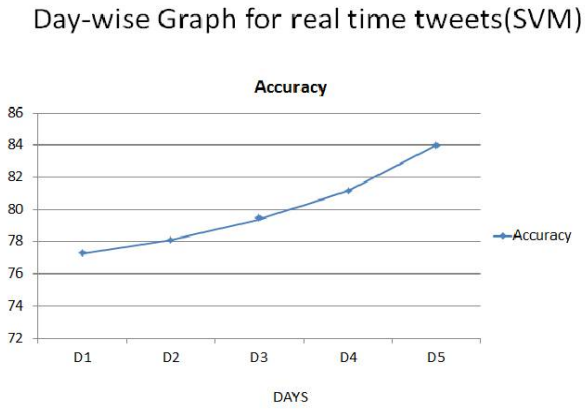

Fig: 6 Day-wise analysis graphs for SVM classifier

Fig. 7 shows the Accuracy comparison between NB and SVM classifier in both Real time and Offline dataset. As shown in graph, NB gives the higher accuracy in both offline dataset and real timedataset.

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. 5, Issue 7, July 2017

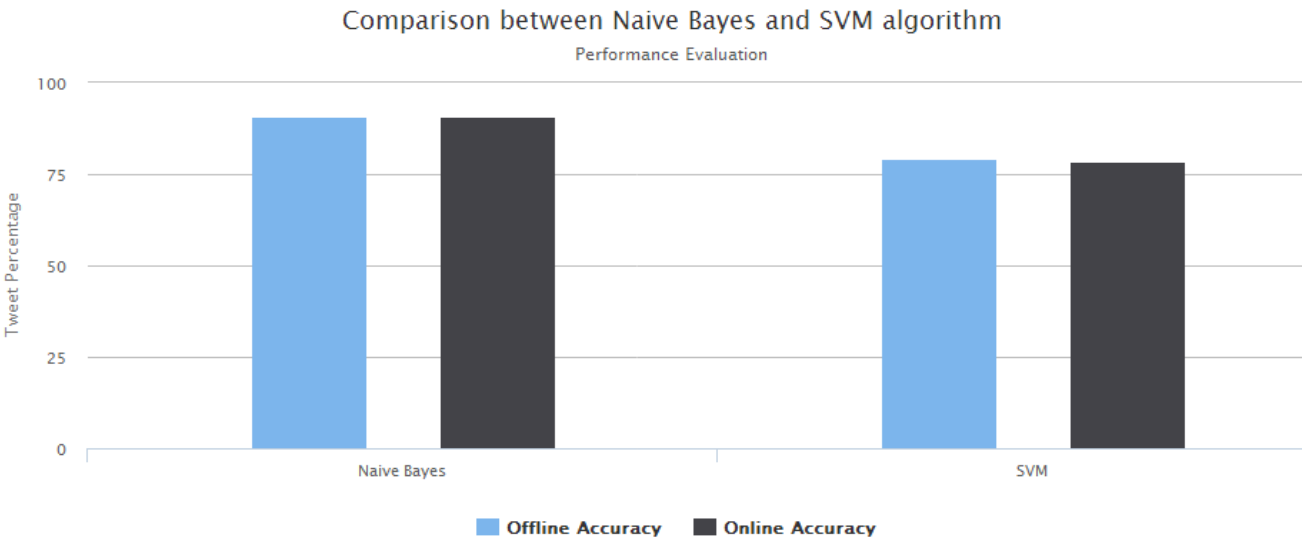

Fig: 7 Accuracy graph for offline and online accuracy

Fig. 8 shows the Day-wise accuracy comparison between NB and SVM classifier.

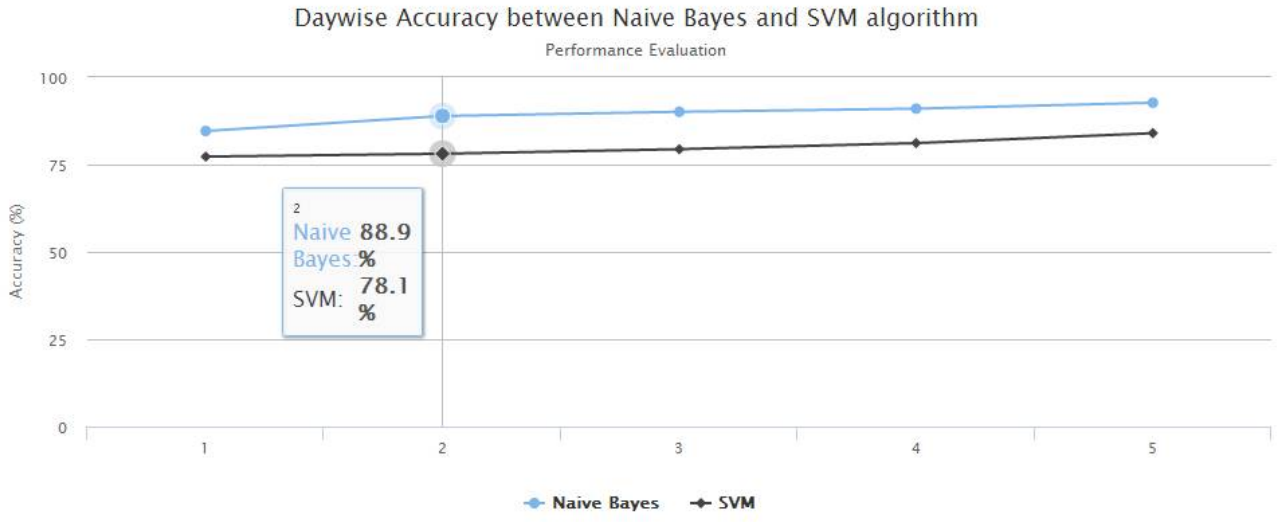

Fig: 8 Day-wise Accuracy graph for NB and SVM classifier

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. 5, Issue 7, July 2017

V. CONCLUSION ANDFUTURE SCOPE

In this dissertation, System provides a fundamental evaluation of ML algorithms on the detection of streaming spam tweets. In this evaluation, system works on offline tweets and real time tweets which are timely updated. System identified that Feature discretization was an important pre-process to ML-based spam detection. System should try to bring more discriminative features or better model to further improve spam detection rate. System applies two classifier and compares their result and measure the performance for them. As the result of this dissertation shows the work of NB classifier is better than SVM classifier for giventweets.

In future system will work on the categorization of tweets. Also system will extend more database value for better result. System may work on the Content and Relation Features of the tweets to achieve the accuracy.

REFERENCES

[1] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. Symp.Netw. Syst. Des. Implement. (NSDI)*, 2012, pp. 197–210.

[2] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proc. 26th Annu.Comput. Sec. Appl. Conf.*, 2010, pp.1–9.

[3] J. Song, S. Lee, and J. Kim, "Spam filtering in Twitter using sender receiver relationship," in *Proc. 14th Int. Conf. Recent Adv. Intrusion Detection*, 2011, pp. 301–317.

[4] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots + machine learning," in *Proc. 33rd Int. ACM SIGIR Conf.Res.Develop.Inf.Retrieval*,2010,pp.435–442.

[5] Nathan Aston, Jacob Liddle and Wei Hu*, "Twitter Sentiment in Data Streams with Perceptron," in *Journal of Computer and Communications*, 2014, Vol-2 No-11.

[6] K. Thomas, C. Grier, D. Song, and V. Paxson, "Suspended accounts in retrospect: An analysis of Twitter spam," in *Proc. ACM SIGCOMM Conf. Internet Meas.*, 2011, pp. 243–258.

[7] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time URL spam filtering service," in *Proc. IEEE Symp. Sec. Privacy*, 2011, pp. 447–462.

[8] X. Jin, C. X. Lin, J. Luo, and J. Han, "Socialspamguard: A data mining based spam detection system for social media networks," *PVLDB*, vol. 4, no. 12, pp. 1458–1461, 2011.

[9] S. Ghosh *et al.*, "Understanding and combating link farming in the Twitter social network," in *Proc. 21st Int. Conf. World Wide Web*, 2012, pp.61–70.

[10] H. Costa, F. Benevenuto, and L. H. C. Merschmann, "Detecting tip spam in location-based social networks," in *Proc. 28th Annu.ACM Symp. Appl. Comput.*, 2013, pp. 724–729.

BIOGRAPHY

Kalyani L. Jagtapis a PG Student,**B. A. Khansole**is Professor in the Computer Engineering Department, Chh. ShahuCollege of Engineering, Kanchanwadi, Aurangabad, BAMU. KalyaniL.Jagtap pursuing Master's Degree in Chh.Shahu college of Engineering,Kanchanwadi,Aurangabad,BAMU, India.