JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REVIEW PAPER ON REAL TIME SPAM DETECTION IN TWITTER

[1]Shruthishree S.H, [2]P Sravani, [3]Akashdeep, [4]Satyaphani pavan,[5]Yadhunanda kishore*[5]

*1Assistant professor, Department of Information and Technology ,FET, Jain University, Kanakpura, Ramanagara ,Bangalore, India

*2.*3,*4,*5: B-Tech, Dept. of ISE Jain university, Faculty of Engineering and Technology, Jakasandra post, Kanakpura Taluk ,Bangalore.

Abstract: Due to increased popularity in online social networks ,spammers used to find these platforms easily accessible to find users in malicious activities by posting spam messages. for this we have taken twitter platform and performed spam tweets detection.

To stop spam comments, twitter bot maker and google safe browsing will detect and will block spam tweets. these tools will block spam links as well ,still they cant protect the user in real time.so that researchers applied many approaches to give spam free social platform. In that some are based on user based features and others are based on tweet based features. we can solve this issue by proposing a framework which will take user and tweets features along with tweet text to make tweets classified. the advantage of using tweet text is we can easily find the spam tweets even if spammer can create new account which cannot be possible only with user and tweet based feature. we find the solution with four machine learning algorithms i.e,, Random forest ,gradient boosting and support vector machine (SVM) , & neural network. In the recent news Elon musk has virtually put on hold of $44 billion twitter deal as he seeks clarity on spam bots. Musk asked twitter INC to show proof that spam bots account for less than 5%of its total users. This deal cannot move forward until twitter CEO shows proof.

KEYWORDS: Random forest, gradient boosting, support vector machine, spam bots, hacking, daniel of services, stop-word, spam drift, real time spam detection, porter stemming.

I. INTRODUCTION

In last few years ,online social platforms like twitter , Facebook , Instagram becomes most prevailing social platforms which are crucial part of people daily life. Now-a-days people will spend more time in blogging website to post their comments/messages & to make friends worldwide and to share ideas .due to increasing in this trend ,these platforms attracts a large number of users and also fake accounts/spammers. In recent time twitter's 400 million users are generating 600 million new tweets per day .This much expansion of twitter platform influences more number of fake accoutres or spammers to generate more spam links which will contain malicious links that directs user to a sites which contains drug sales , scams and phishing. these may lead to temporary shutdown of internet services all over the world as it is not only damaging user experience but also the internet. As twitter and researchers came up with spam detection solutions to make spam free social networking platform. Twitter built twitter botmaker to fight spam in twitter.so that by launching bot maker they have seen 40 to 45% of reduction in spam metrics , but the main disadvantage of botmaker is it fails/it cannot protect a victim from new spam.it is not correct tool for real time spam tweets detection.

II. REVIEW OF LITERATURE

Ford and gordon proposed a typology which is consisted of two classifications: i.e,,

1) The discrete/singular events facilitated by the introduction of programs (malware) like virus ,key strokes loggers and the root kits.

2) These are facilitated by programs which are not classified as crime-ware and there are many repeated events/contacts from perception of the user .

There are many more broader classifications which is recommended by wall, proposed three categories: The very first one is computer integrity crimes which includes the illegal activities **hacking** and **denial of services**. The second category is computer assisted crimes which offences of scams ,virtual robberies are included in this. In third category the computer content crimes is included of offensive communications.

III. MATERIALS & METHODS

Algorithms/Technology used:

The algorithm used here is Support Vector Classification Algorithm:

An Support vector machine(SVM) classifier is a linear classifier in which it helps to separate the hyper plan so that the expected classification error for unseen test patterns is minimized. SVM is a strong classifier.It is also used to apply for gender classification problems by many researchers successfully.

Porter Stemming Algorithm:

Porter stemming is a method or process for to remove common morphological and inflexional end from the english words.

Natural Language Processing :

NLP is a high level language i.e English processing technique of communication with processes the data entered by user in high level language and give back the appropriate response.

IV. RESULT & DISCUSION

Pre-processing techniques are applied on dataset to get clean data.
First we will scan the whole document of input dataset and after comparing with some other documents , if we get same it will remove it and also it will check in twitter available comment with some other USER ID.
The next step , other preprocessing is remove the special characters and symbols where the whole input document gets scanned and compared with symbols are deleted.

In this for their move to the **stop-word** removal being the next step in data. Stop-word removal exactly means that from the whole statement after scanning it removes the words like and, is, the, etc and only keeps noun and adjective to from the statement.

RESULTS:

This is the web app which we created for our Twitter Spam detection.
Scenario for spam tweet:

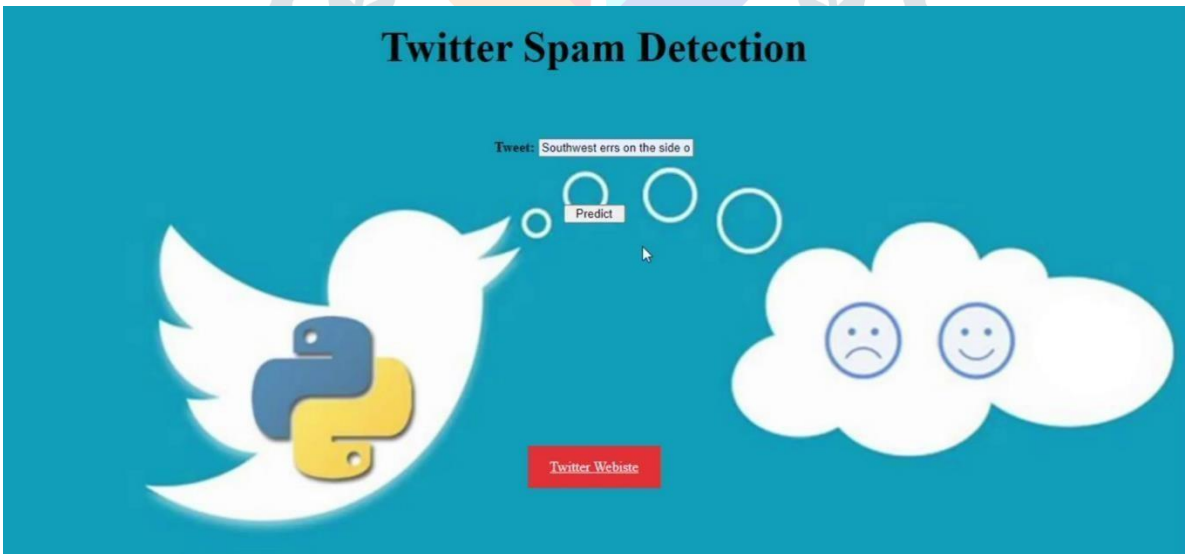

Figure : 1.1 web Search

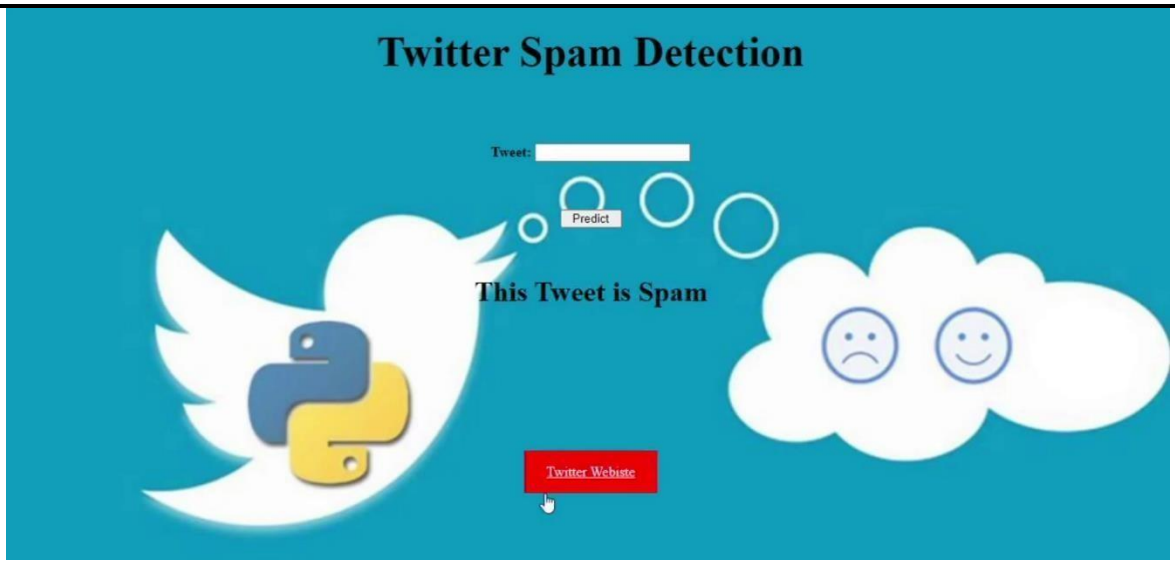

Figure : 1.2 Input of a Spam tweet

The output of the web application.

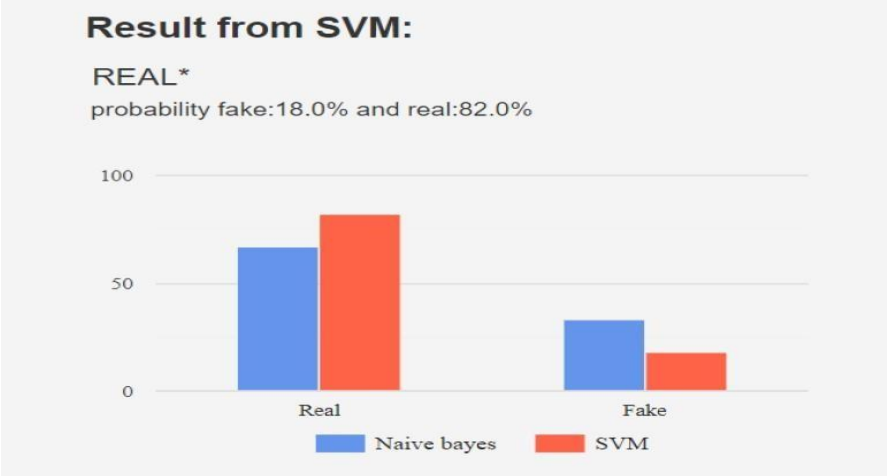

Figure: 1.3 Result of the input Spam (real time)

Here the model predicted the input article and returned the output from SVM.

The probability of each class is displayed in a graph separately for the comparison purpose.

Figure 1.4 Final output Spam tweet

Real Time Tweet

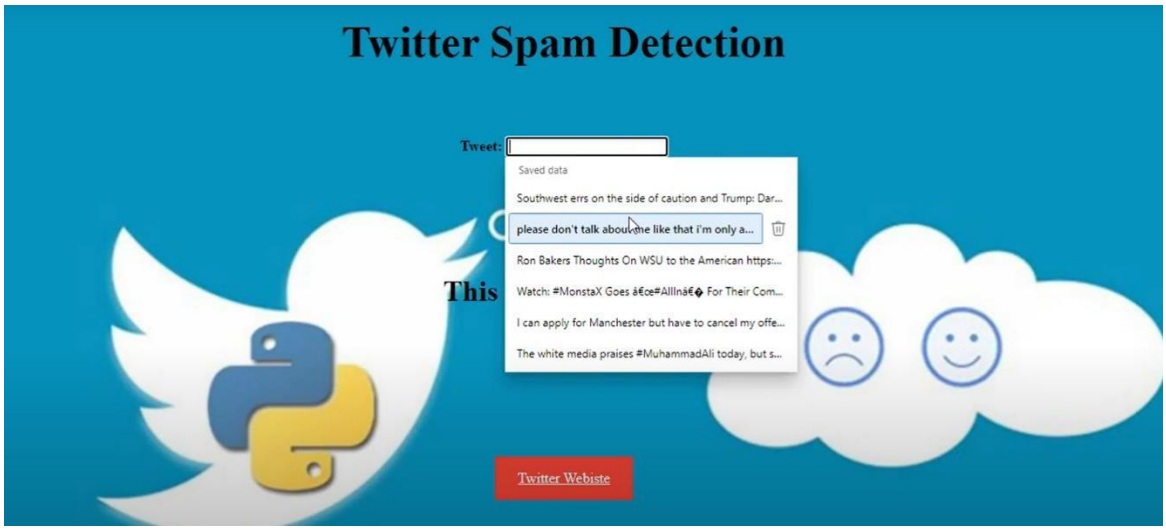

Figure: 1.5 Twitter Spam Detection Real Tweet

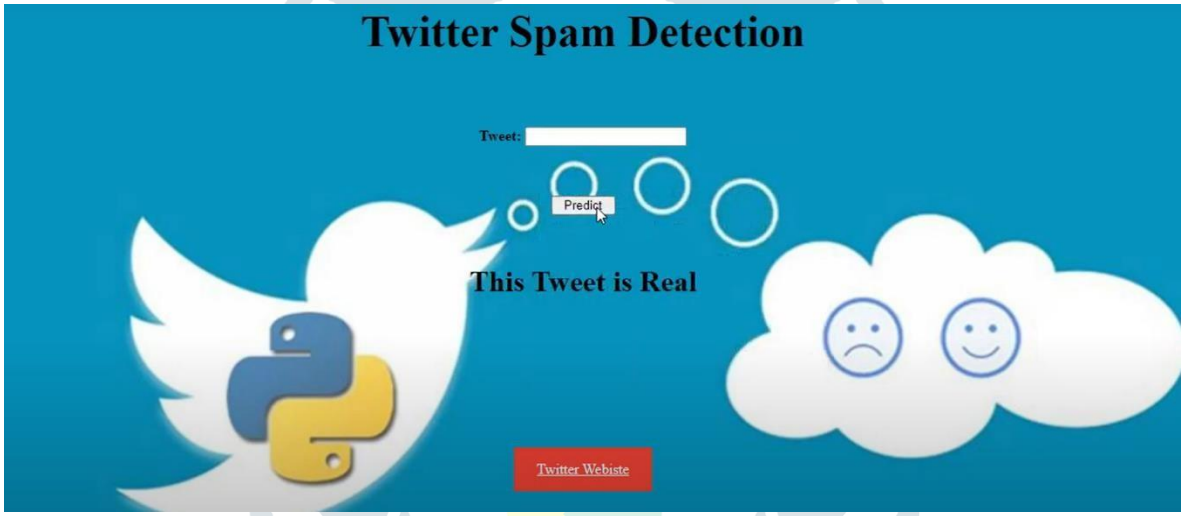

Figure: 1.6 Twitter Spam Detection Final Output Real tweet

V. CONCLUSION

In this paper, we have presented a real time spam detection in twitter. From this we have collected big number of 400,000 tweets from public. dependent on tweets text we have extracted top 30 words which can give the highest information in order to classify tweets. And also we have tested our approach in real time tweet detections. In the real world , spam features of tweets keep changing in an unanticipated way. this is referred as spam drift. By implementing self learning algorithm we will keep on updating our model based on new spam tweets. And also we have observed that in our dataset 80%of spam tweets contains a malicious links.so we will perform URL crawl mechanism to detect twitter spam. In real time we have frequent pattern mining of tweets text which is main aspect to distinguish twitter spam.

VI. REFERENCES

[1] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, “6 million spamtweets: A large ground truth for timely twitter spam detection,” in 2015.IEEE International Conference on Communications (ICC), June 2015,pp. 7065–7070.
[2] A. Greig, “Twitter Overtakes Facebook as the Most Popular Social Network for Teens,According to Study, DailyMail, accessed on Aug. 1, 2015 ,” http://www.dailymail.co.uk/news/article-2475591/Twitter-overtakes-Facebook-popular-socialnetwork-teens-according-study.html, 2015, [Online].
[3] H. Tsukayama, “Twitter turns 7: Users send over 400 million tweets perday,” https://tinyurl.com/ybsaq7e7, 2013, [Online].
[4] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, “Detectingspammers on twitter,” in In Collaboration, Electronic messaging, Anti_Abuse and Spam Conference (CEAS, 2010.
[5] C. Pash., “The lure of Naked Hollywood Star Photos Sent the Internet into Meltdown in New Zealand, Bus. Insider, accessed on Aug. 1, 2015,” https://tinyurl.com/yc93ssj4, 2014, [Online].
[6]“BotMaker,https://blog.twitter.com/engineering/enus/a/2014/fighting-spam-with-botmaker.html, [Online].
[7] K. Thomas, C. Grier, D. Song, and V. Paxson, “Suspended accounts inretrospect: An analysis of twitter spam,” in Proceedings of the 2011ACM SIGCOMM Conference on Internet Measurement Conference, ser.IMC ’11. New York, NY, USA: ACM, 2011, pp. 243–258. [Online].