

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

Spam Detection on Collection of Twitter Data Using Naive Bayes Algorithm

Ms. Ashwini Athawale¹, Mrs. Deepali M. Gohil²

P.G. Student, Department of Computer Engineering, DYPCOE, Akurdi, Maharashtra, India¹

Associate Professor, Department of Computer Engineering, DYPCOE, Akurdi, Maharashtra, India²

ABSTRACT: Considering today's state of people communicating with each other using online social network like twitter. We could conclude lot of spams present in twitter site. For reducing the unwanted threats we need to first identify spammers in twitter accounts. Our proposed approach to identify spam in twitter using content, user based feature to analyze spam behavior of user. In this project we use Twitter API (Application Protocol Interface) to get all details of twitter user and then generate analysis and match with predefined result. After collection of tweets stopword removal is being done, then the classification is done using bagging technique in the algorithm. We use Naive Bayes algorithm to match feature and identify spam. In result, the comparison of the algorithm with existing method is visualized.

KEYWORDS: Anomaly detection; Feature extraction; Naive Bayes Classification; Training; Testing; Twitter

I. INTRODUCTION

Identify Anomalous User Behavior on Twitter is web application used for detecting the anomalous data on the twitter. It will help you to find out the malicious behavior of the user account on the twitter, manage the visualization graph for detection of user suspicious.

Everyday, new and advanced form of technology broadening the mindsets to cope up with the increasing necessity of information and technology. This may include the daily usage of loads of information transferring through emails, messages on social media networks. Hence the risk involved of potential threats to the society by spammers and frauds could affect the actual functioning of the network. So it is needful to identify the threats for the betterment of the information safety over network. Some spammers also determined to deteriorate sending piles of junk files leading to the frustration of user. There are some studies which have shown the behavioural features of anomalies present over network. Eg., Some spammers work early in the morning while the common user is sleeping. In Machine learning we need train machine to predict appropriate result to show spammers. Machine learning divided into two parts supervised learning and non-supervised learning. In supervised learning we train classifier and unsupervised learning we didn't require to learn classifier. But supervised learning give better accuracy as compare to unsupervised learning.

II. REVIEW OF LITERATURE

Nan Cao et al. used unsupervised machine learning algorithm (TLOF- Time adaptive native Outlier Factor) anomaly detection supported behavioral options of user accounts and so analyzing whether or not it's spam or not. They created novel visual analysis system, Targetvue, to visually summarize a users behaviors that effectively gift the users communication activities, features, and social interactions[1].

H.Gao et al. introduced novel framework (Tangram) for spam filtering system that detects the spam in twitter and that they find out distance between sender and receiver message. during this paper, they need by trial and error analyze the matter pattern of an outsized assortment of on-line Social Network (OSN) spam. puzzle mechanically divides OSN spam into segments and uses the segments to construct templates to filter future spam. Experimental results show that puzzle is very correct and might apace generate templates.[2].

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

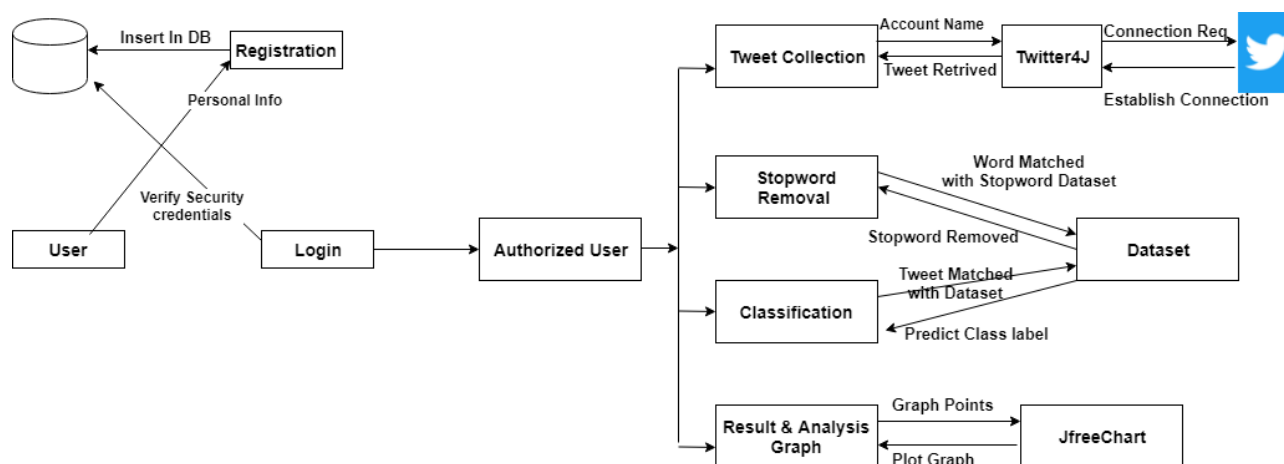
J. Hu, C. Wilson et al. they bestowed a study to live and analyze the tries to unfold malicious content on OSNs. They discovered in their study that, spamming dominates actual wall post activity within the early morning hours, once traditional users square measure asleep. They used user primarily based feature extraction and classify those feature with area mathematician algorithmic rule and so determine given content is spam or not spam[3].

D. Palsetia et al. analyze to acknowledge spam uniform resource locator they check spam uniform resource locator to spot similarity between uniform resource locator and content in uniform resource locator. They projected to reconstruct spam messages into campaigns for classification instead of examine them on an individual basis. thence endeavor the message labels 'spam' before delivering the meant user and thence protective from the frauds[4].

S. Lee et al. analyze the user account supported past messages and retweet and post done by user. They projected WARNINGBIRD, a suspicious uniform resource locator detection system for Twitter. they need collected an outsized variety of tweets from the Twitter public timeline and trained a applied mathematics classifier with options derived from related URLs and tweet context info. K. Thomas et al. given tweet text compare with predefined dataset of word. In dataset is assortment of word that is take into account as spam. If input tweet contain word gift then given tweet take into account as spam [5].

III.SYSTEM OVERVIEW

A. System Architecture



1. Data collection : To fetch a data from twitter we need access twitter, access obtained by creating a twitter application. Whenever we create application we get four access keys from twitter. They are four required for integrate twitter:

- Consumer Key
- Consumer Secret
- OAuth Access Token
- OAuth Access Token Secret

By using this key in Java program we are able to collect user data. System collect the input as twitter data and later use for classification using Nave bayes.

2. Preprocessing: The twitter tweets contain noise. That may decrease the accuracy of the system hence there is necessity to remove noise from the tweets. We will be collecting stopword dataset from internet and later tokenizing each tweets collected from twitter. Then this tweets words matched with dataset. If we found stopwords in tweet then we simply remove those words. Preprocessing technique enhanced accuracy of classification.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

- Classification : In classification, there are mainly two class labels which are spam or not spam. We have collected sms spam dataset from UCI machine learning repository. Later removing stopwords is being done in dataset. Based on the dataset we need to train our Naive bayes algorithm. Naive bayes is supervised learning algorithm so it required trained dataset so we train naive bayes by using SMS Spam dataset. While we we fetch live user tweets from twitter those tweet consider as test input to naive bayes. we are using baging technique for enhancing accuracy of naive bayes.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

$$P(c|X) = P(x_1|c)P(x_2|c):::P(x_n|c)P(c) \quad (2)$$

$P(x_i|c)$ represents when class label is c then how many times x_i is occurred

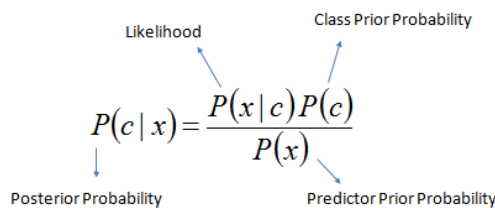
$p(x)$ represents no of times given x word is occurred in dataset.

$p(c)$ - occurrence of given class label in dataset.

- Anomaly detection: Anomaly detection identify anomalous behavior of user by using spam count. In classification module, we will get number of spam tweets of user as well as not spam tweets of user. If spam count is greater than not spam then we consider given user is anomaly.

B. Algorithm

Bayes theorem provides a way of calculating the posterior probability, $P(c/x)$, from $P(c)$, $P(x)$, and $P(x/c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$


$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

$P(c/x)$ is the posterior probability of class (target) given predictor (attribute).

$P(c)$ is the prior probability of class.

$P(x/c)$ is the likelihood which is the probability of predictor given class.

$P(x)$ is the prior probability of predictor.

In ZeroR model there is no predictor, in OneR model we try to find the single best predictor, naive Bayesian includes all predictors using Bayes' rule and the independence assumptions between predictors.

IV. EXPERIMENTAL RESULTS

Sentiments are the observations or attitude representation of sentences in terms of positive or negative. The tweets from twitter account are extracted in the unstructured data form. The unstructured data is being converted into structured form then feature extraction is done. The features of the words are selected and then classification technique is applied on extracted features. Which then lead to classify them into its sentiment polarity that is namely either spam or not spam. Feature words representation based on Nave Bayes classifier is being done.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

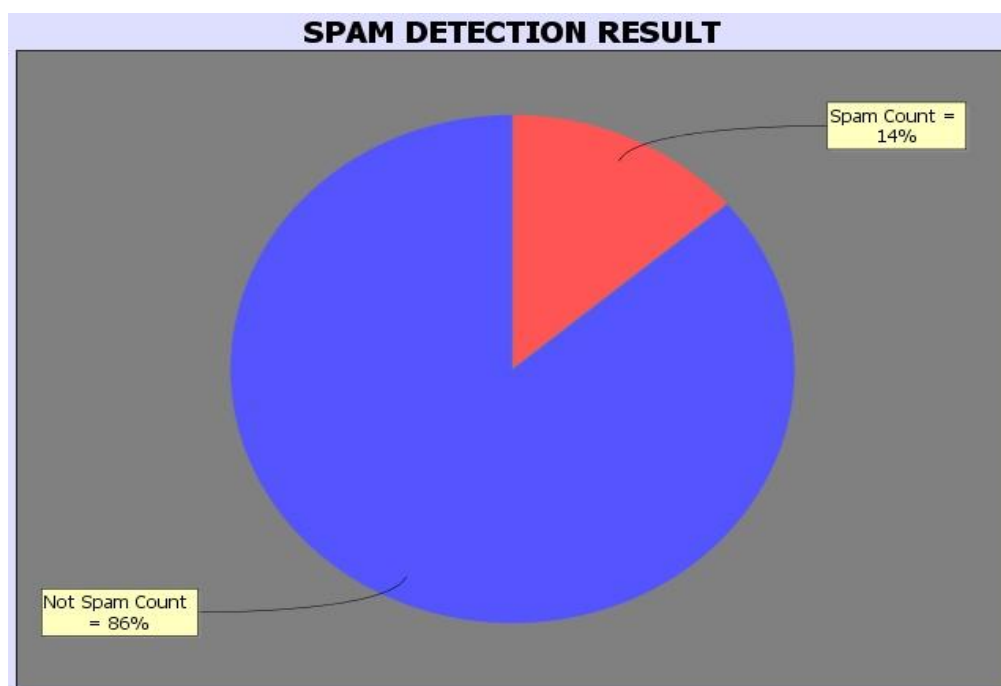


Fig. 2. Graph

After getting counts of tweets with its polarity, it is then shown graphically in fig 2, which is in snapshot form. Hence result and analysis shown graphically.

The examples of stopwords removal and then applying feature extraction we got the polarity of the tweet. Precisely, depending on the polarity the spam count is being detected. Finally can identify whether the user account is spam or not.

Accuracy has been tested with the SVM (Support Vector Machine) supervised learning models with associate learning algorithms. Comparison shown in fig 3 reciprocates the accuracy of naive bayes is higher and hence efficiency increased in the existing system.

The Time graph of our system is also shown in fig 4. Data collection time is only greater as comparatively to the stop-word removal and classification, as the number of tweets present in account are used as input. Outcome from this project is the capability to use application for identify spam on online social network.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

SPAM DETECTION ACCURACY RESULT

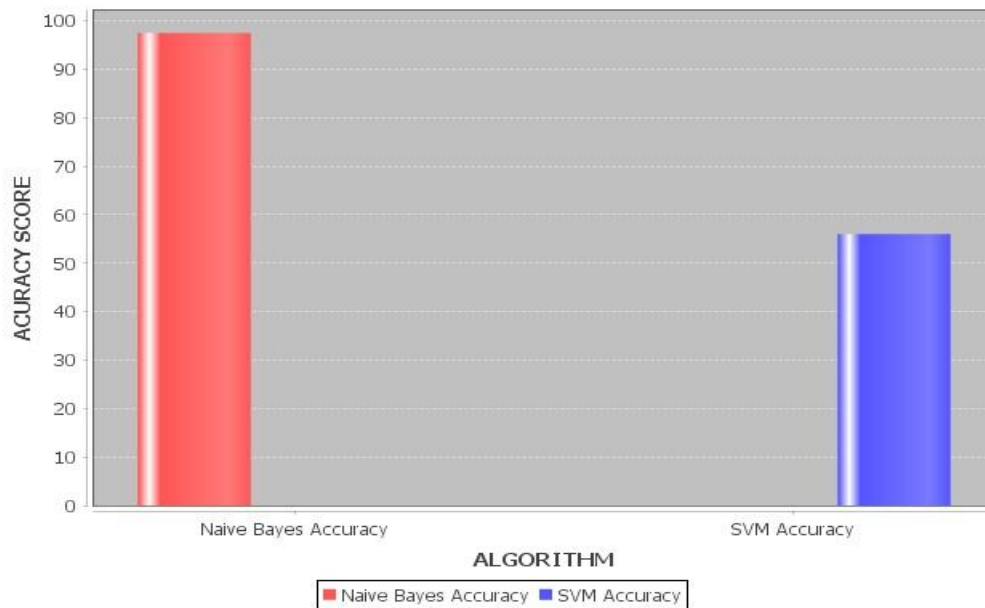
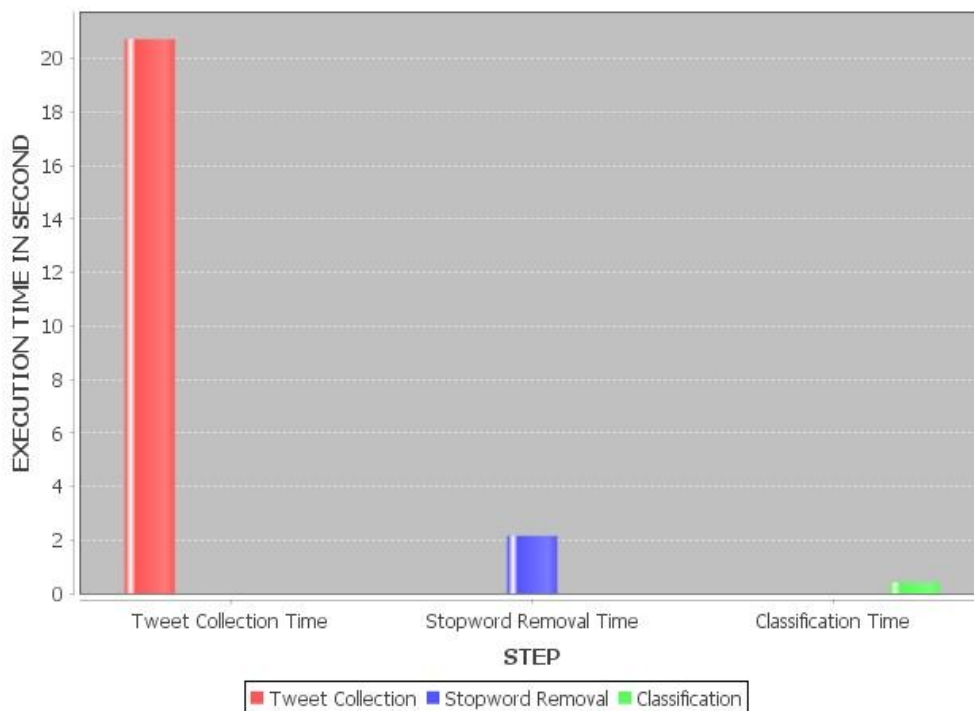


Fig 3. Accuracy

SPAM DETECTION TIME GRAPH



International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

V.CONCLUSION

The zoom of social network over the web usage has intense the anxiety among the authorised user, regarding concerning increasing threat of anomalies gift over the network. during this paper, anomaly detection is being done by mistreatment trained data-set and Naive Bayes rule for the classification. information pre-processing steps embrace the filtering and stop-words removal. Feature choice is being done later. In result we have a tendency to show the whether or not the user account is spam or not. therefore any analysis is finished by comparison the results with SVM rule. In future the system may be accustomed discover anomalies in social networks like Facebook and LinkedIn, any enhancing step may well be to urge authority to get rid of the spam account from the social network.

REFERENCES

- [1] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin and Ching-Yung Lin, TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems." IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 22, NO. 1, JANUARY 2016.
- [2] Hongyu Gao, Yi Yang, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, Alok Choudhary, Spam aint as diverse as it seems: Throttling OSN spam with templates underneath, in Proc. ACSAC, 2014, pp. 7685.
- [3] H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen, and B. Y. Zhao, Detecting and characterizing social spam campaigns, in Proc. IMC, 2010, pp. 3547.
- [4] H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary online spam filtering in social networks, in Proc. NDSS, 2012, pp. 116.
- [5] S. Lee and J. Kim, WarningBird: Detecting suspicious URLs in Twitter stream, in Proc. NDSS, 2012, pp. 113.
- [6] 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. SP, May 2011, pp. 447462.
- [7] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, A geometric framework for unsupervised anomaly detection." In Applications of data mining in computer security, pages 77101. 2002.
- [8] I. Steinwart, D. R. Hush, and C. Scovel, A classification framework for anomaly detection." In Journal of Machine Learning Research, pages 211232, 2005.
- [9] S. T. Teoh, K. L. Ma, S. F. Wu, and X. Zhao, Case study: Interactive visualization for internet security." In Proceedings of the conference on Visualization, pages 505508, 2002.
- [10] C.-F. Tsai, Y.-F. Hsu, C.-Y. Lin, and W.-Y. Lin, Intrusion detection by machine learning: A review." Expert Systems with Applications, 36(10):1199412000, 2009.
- [11] Mladenic, Dunja, and Marko Grobelnik, Feature selection for unbalanced class distribution and naive bayes. CML. Vol. 99. 1999.
- [12] Tang, Bo, Steven Kay, and Haibo He, Toward optimal feature selection in naive Bayes for text categorization. (2016).