iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IX **Month of publication:** September 2020

DOI: https://doi.org/10.22214/ijraset.2020.31677

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

Analysis and Detection of Spam Comments on Social Networking Platforms like YouTube using Machine Learning

Prof. Swati Khodake[1], Sonali Patil[2], Reshma Jadhav[3], Ishwari Mote[4], Dakshata Ramteke[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, JSPM Bhivarabai Sawant Institute of Technology and Research, Wagholi, Pune, Savitribai Phule Pune University, Pune, India*

Abstract: ***The profitability promoted by Google in its well-known video distribution platform YouTube has attracted an increasing number of users. However, such success has also attracted a large number of malicious users, which aim to self-promote their videos or circulate viruses and malware. As we know that YouTube offers limited tools for comment moderation, so spam increases very rapidly and that's why the comment section of the owners is disabled. It is very difficult to established classification methods for automatic spam filtering since the messages are very short and often widespread with slangs, symbols, and abbreviations. In this paper, we have evaluated several top-performance classification techniques for detecting and analyzing spam comments. The statistical analysis of results indicates that, with 99.9% of confidence level, decision trees, logistic regression, Bernoulli Naive Bayes, random forests, linear and Gaussian SVMs are statistically equivalent in maximum rate. Therefore, it is very important to find a way to detect these comments on videos and report them before they are viewed by innocent users.***

Keywords: Spam Filtering, SVM, Naïve Bayes, Machine Learning, YouTube

I. INTRODUCTION

The popularization of wideband around the world has boosted the number of Internet users. With faster connections, video host and sharing services are becoming popular among users. According to a press release of Sandvinel, a company focused on standards-acquiescent network policy control, around 55% of down flow traffic from the United States is due to video platforms like Netflix and YouTube. The availability of resources through internet and the wideband connections allowed the appearance of sophisticated new platforms. In these lines, YouTube is a renowned video content distribution stage with informal community highlights, for example, support for presenting content remarks on giving cooperation between channel proprietors and watchers or endorsers. The success of YouTube is expressed through recent statistics rumored by Google the platform has quite more than one billion users, 300 hours of video are uploaded every minute and it generates billions of views every day or every minute. Around an hour of a creator's views come back from outside their home country and 1/2 YouTube views are on mobile devices.

Recently, YouTube has adopted a monetization system to reward producers, stimulating them to make high-quality original content and increasing the quality of visualizations. After the deployment of this system, the platform was flooded by undesired content, usually of low-quality information known as spam. Among completely different reasonably unsought content, YouTube is facing problems to manage the huge volume of undesired text comments posted by users that aim to self-promote their videos or to disseminate malicious links to steal private data. The spam found on YouTube is directly associated with the engaging profit offered by the substantiation system. According to a press release by Google, more than a million advertisers are using Google ad platforms, the mobile revenue on YouTube is up 100% year over year and the number of hours folk look on YouTube every month is up five hundredth year over year. At the same time, according to Negate, a computer security company, just in the first half of 2013, the volume of social spam increased by 55%. For each spam found on any social network, different two hundred spams square measure found on Facebook and YouTube. The problem became therefore vital that it actuated users to make a petition in 2012, in which they ask YouTube to provide tools to deal with undesired content. In 2013, the YouTube official blog reported efforts taken to deal with undesired comments through various ways like recognition of malicious links, ASCII art detection, and display changes to long comments. However, large number of users still not satisfied with given solutions. In fact, in 2014, the user "PewDiePie, owner of the most subscribed channel on YouTube which has nearly 100 million subscribers disabled comments section on his videos, claiming most of the comments are mainly spam and there's no tool to wear down them. The problem caused by social spam began to be seriously mentioned in 2010, but an earlier work is dated from 2005.

However, unsought comments on YouTube still hurt the platform's community, evidencing such drawback needs attention and analysis. Established techniques for automatic spam filtering have their performance degraded when dealing with YouTube's comments as YouTube has a large dataset with millions of users. This drawback occurs mainly because such messages are usually very short and rife with idioms, slangs, symbols, emoticons, and abbreviations which make even tokenization a challenging task.

II. LITERATURE SURVEY

Spam is nothing but irrelevant content with low quality information send over the internet typically to a large number of users for the purpose of phishing, spreading malware. They are commonly found as images, texts or videos, hindering visualization of interesting things. There are many types of spam such as web spam, blog spam, e-mail spam, and SMS spam. In social networking sites, undesired content is known as social spam. You tube is one of the social networking site ,also having spam comments. To detect and analyses this spam comments or to handle this malicious activity there are some studies and literature to finding the efficient way. There are some existing systems also to handle this activity like SVM, CNN Architecture etc. The most recent method used to handle this spam comments on You tube is by using some algorithms like decision tree, logistic regression, Naive Bayes etc. There are some users which publishes low quality of videos on You tube known as spam videos. There are some studies to find efficient ways to handle this activity through classification methods and feature extraction from meta data. The next common alternative is automatic blocking spammers – users that disseminate spam. There are some comments which also contain links, such types of comments are also detect by this algorithms. you tube spam comment detection and analysis by using machine learning is also useful for other social networking sites. Before this method of spam filtering ,there was some existing methods. In that methods, CNN architecture or SVM is used. Some methods uses Bayesian classifier in which spam filtering is based on network between comments. Some methods detects spam profiles using public features. That methods also have some drawbacks, so all those drawbacks are overcome in this method of spam comments detection.

III. SYSTEM OVERVIEW

The number of spammers can attack social sites of people to spread malicious content also to threaten them. Spammers discard malicious comments on people's post also transfer traffic from one site to another in order to prevent such activities, we implemented such a system.

In the proposed system, five YouTube videos were extracted from four of them used as training sets, and one using for testing. In order to find spam comments, various machine learning algorithms are used, such as Naive Bayes, SVM, decision tree, etc. Here, we create four dictionaries for spam words, stop words, URL links, and remaining words. Finally, based on the accuracy of the model calculated by this ratio, we calculated the ratio of the total number of spam messages to the number of words in the comments.

IV. CLASSIFICATION MODELS

A. Naïve Bayes

In the machine learning algorithm, Naive Bayes is the classification algorithm. Its primarily used for solving text classification problem, which having high dimensional training data sets. It is a probabilistic classifier. It calculates the conditional probability that is the probability of an event based on previous knowledge available. This algorithm is known for its simplicity but also for effectiveness.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

B. C4.5

C4.5 is one of the machine learning algorithms that come from tree family. This algorithm generates trees from a training data set using information entropy. Decision three identifies and classifies information from large sets and returns a coherent output.C4.5 is an extension of the ID3 algorithm. By selecting the attribute of the data at each node algorithm splitting the data samples into a set of subset using the information gain criteria and the highest attribute value will make the decision. These processes are repeated up to the smallest sub-lists.

C. SVM

Support Vector Machine used for the regression but primarily used for the classification. SVM Supervised learning algorithm that looks of data & sorts it into one of the two categories. It is a discriminative classifier that is formally designed by a separate hyper plane. SVM represents the examples as points and mapped in a space so that examples are divided by a clear gap as wide as possible. This is helpful in text and hypertext categorization. SVM is more effective in high dimensional space.

D. Logistic Regression

Logistical regression used as a statistical model for finding the probability of certain class in terms of discrete or categorical results such as yes or no, 1 or 0. It can be binomial, ordinal, or multidimensional. This is used for measuring the relationship between categorical dependent variables one or more independent variables by estimating probability using a logistic function. Also very easy to implement, interpret, and very efficient to train.

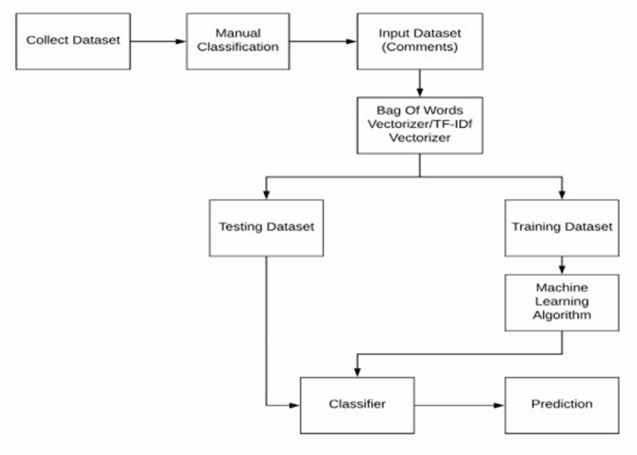


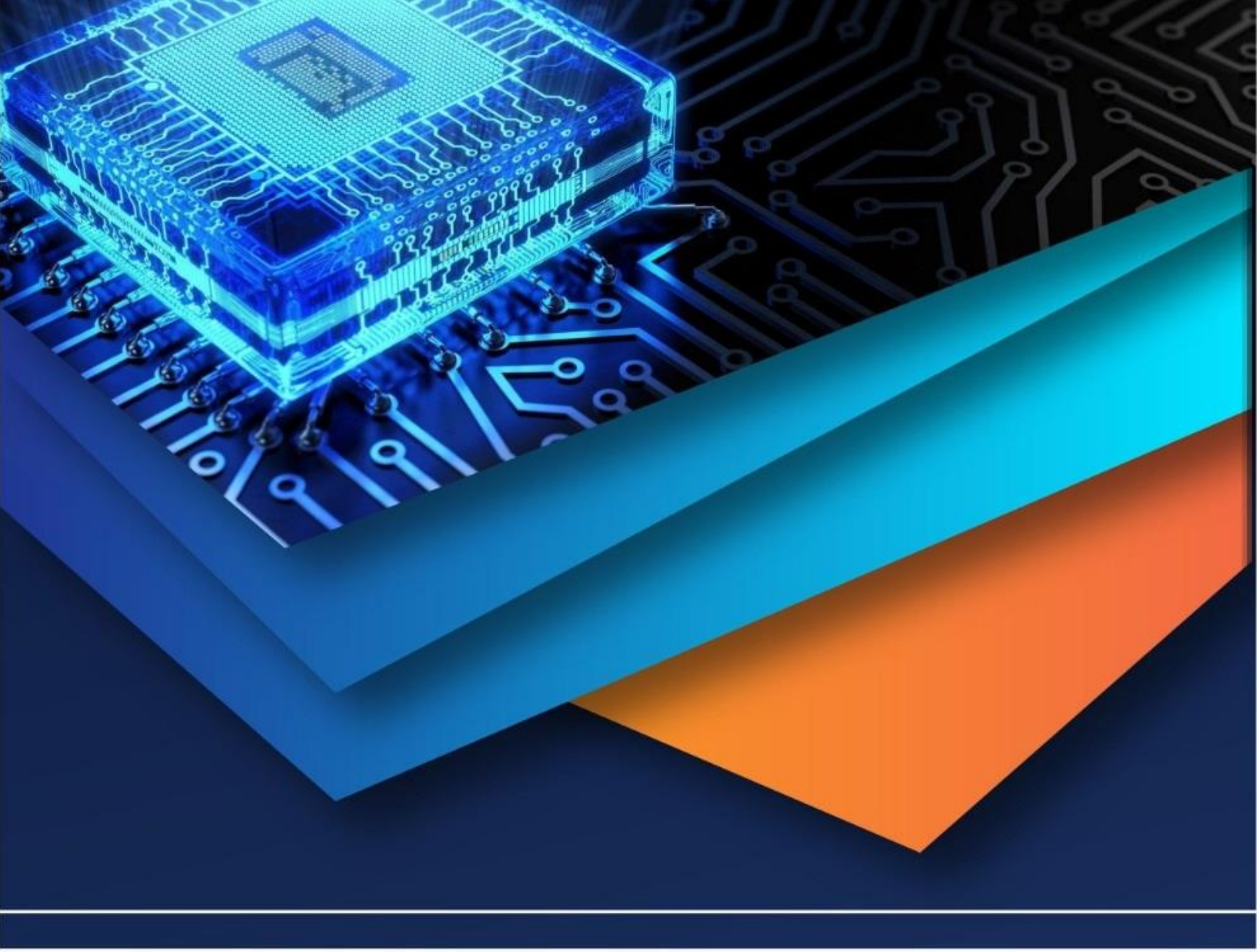
Fig. Proposed system

V. CONCLUSION

Social media networks have become popular and this creates the opportunity for the spammers to publish unwanted comments. Previously, some machine learning algorithms were used for this detection. In the proposed system we also use the advanced machine learning algorithms with advanced features .also compares the efficiency of various algorithms by applying them We construct features based on the features obtained from the user profile and the content that they shared. Based on the experiments conducted, it can be expected that existing classifiers widely used in the data mining community can utilize these functions to detect spammers.

REFERENCES

[1] K. Lee, J. Caverlee, and S. Webb, “Uncovering social spammers: social honeypots+ machine learning,” in Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, pp. 435–442, ACM, 2010
[2] A. H. Wang, “Don’t follow me: Spam detection in Twitter,” in Proc. Int.Conf. Secur. Cryptogr. (SECRYPT), Jul. 2010, pp. 1–10.
[3] Miss.Shukla Twinkle Kailas,Prof.D.B.K shirsagar,”Design of machine learning approach for spam tweet detection”,IEEE,2016
[4] Ala’ M. Al-Zoubi∗, Ja’far Alqatawna, Hossam Faris"Spam Profile Detection in Social Networks Based on Public Features",IEEE,2017
[5] T. Wu, S. Wen, Y. Xiang, and W. Zhou, “Twitter spam detection: Surveyof new approaches and comparative study,” Comput. Secur., vol. 76,pp. 265–284, Jul. 2017.
[6] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou and G. Min, “Statistical Features- Based RealTime Detection of Drifted Twitter Spam”, IEEE Transactions, April 2017, pp.914-925.
[7] Shivangi Gheewala, Rakesh Patel” Machine Learning Based Twitter Spam Account Detection: A Review”, IEEE,2018.
[8] Sreekanth Madisetty, Maunendra Sankar Desarkar"A Neural Network-Based Ensemble Approach for Spam Detection in Twitter",IEEE,2018
[9] Spam tweet detection”,IEEE,2016
[10] Chao chen, Jun Zhang ,Yi Xie and Yang Xiang .”A performance evaluation of machine learning based streaming spam tweets detection”,in IEEE transaction on computational social system,2015,Vol-2 No-3.
[11] A. Gupta and R. Kaushal, “Improving Spam Detection in Online Social Networks”, IEEE, 2015.
[12] M. Verma, Divya, S. Sofat, “Techniques to Detect Spammers in Twitter – A Survey”, International Journal of Computer Applications, January 2014, Vol. 85, No. 10, pp. 27-32.
[13] Ziyan Zhou ,Lei Sun,”Network based spam filter on tweeter”,2014
[14] [2] K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan and S. A. Razak, “Malicious accounts: Dark of the social networks”, Elsevier, 2017, pp. 41-67



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY