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

YOUTUBE SPAM FILTER USING MACHINE LEARNING

Prof. Swati Patil[1], Shrushti Pawar[2], Prathamesh Paware[3], Ashvini Kawade[4], Sahil Gargate[5]

[1] Assistant Professor, Department of Computer Engineering, JSPM's Jaywantrao Sawant College of Engineering, Pune

[2-5] Student, Department of Computer Engineering, JSPM's Jaywantrao Sawant College of Engineering, Pune

Abstract : *The profit promoted by Google in its spick-and-span video distribution platform YouTube has attracted a growing scope of users. However, such success has conjointly attracted malicious users that aim to self-promote their videos or bear viruses and malware. Since YouTube offers restricted tools for comment moderation, the spam volume is shockingly increasing that is leading house owners of known channels to disable the comments section in their videos. Automatic comment spam filtering on YouTube could be a challenge even for established classification ways since the messages square measure terribly short and infrequently rife with slangs, symbols, and **elisions**. During this work, we've evaluated many top-performance classification techniques for such purposes. The applied math analysis of results indicates that with 99.9% of confidence level Bernoulli Naive Bayes, Decision trees, Logistic Regression, Random forests, Linear and Gaussian SVM's square measure statistically equivalent. Therefore, it's important to search out some way to notice these videos and report them before they're viewed by innocent user*

Key Words: Machine learning, Random Forests, Logistic Regression, Bernoulli Naïve Bayes, Decision trees, linear and Gaussian SVMs.

I. INTRODUCTION

In previous years of the pandemic, YouTube, an online video entertaining and also a social media platform is gaining recognition by people in whole world. Due to various types of video content available to watch on this platform people of almost all age groups are attracted to it, which makes YouTube an easy target for spammers. For example, there are educational videos available for us to watch. This comprehensive and attractive environment provided by YouTube creates an opportunity for various spammers to create unrelated content aimed at users. These unsolicited spam comments/messages are aimed to attack users by enticing them into clicking malicious sites which contain malware, phishing, and scams. YouTube has an astonishing feature of "Comments" which is used for users to express their feeling towards video in the form of comments..

II. LITERATURE SURVEY

1. Spam is typically associated with unwanted content with caliber info. they're usually found as pictures, texts, or videos, impeding the mental image of fascinating content. There area unit several pieces of research associated with spam in literature, like internet spam, blog spam, e-mail spam, and SMS spam.
2. On social networking sites, unwanted messages area unit referred to as social media spam, journal comment spam is the most similar state of affairs. However, the most-known strategy to notice a journal spam comment sometimes is to seek out the simplest illustration of language model in post-publication, mistreatment that illustration to filter less connected comments to its original subject. Such a strategy can't be applied on YouTube, since the comments area unit is associated with video content with little or no matter description, thus language models can't be properly mapped from the original publication.
3. YouTube additionally faces malicious users that publish caliber content videos, it's referred to as video spam. There area unit some studies in the literature to seek out economic ways in which to handle this activity through classification strategies and have extraction from data, like title, description, and recognize numbers.
4. Another common way is to automatically block spam – users who disseminate spam. However, unlike spam disseminated in other social networks and email, the spam posted on YouTube is not usually created by bots, but posted by real users aiming self-promotion on popular videos. Therefore, such messages are more difficult to identify due to their similarity to legitimate messages. Automatic spam filtering is useful in other tasks as well. Severyn et al. reported significant improvement of performance in the opinion detection task when spam samples were removed before training a classifier.
5. As reported by Bratko et al., the spam filtering function differs slightly from similar text categorization problems. They claim undesired messages have chronological order and their characteristics may change according to that.

6. It also explains that cross-validation is not recommended, because earlier samples should be used to train the methods, while newer ones should be used to test them. Furthermore, in spam filtering, errors associated with each class should be considered differently, because a blocked legitimate message is worse than unblocked spam.

III. PROBLEM STATEMAENT

To design and develop an ML-based YouTube Spam Filter which is capable of processing data of comments of various YouTube channels and classifying them into spam and legitimate comments based on the dataset on which various machine learnings models have been applied

❖ GOALS AND OBJECTIVES

The main objectives of developing this system are:

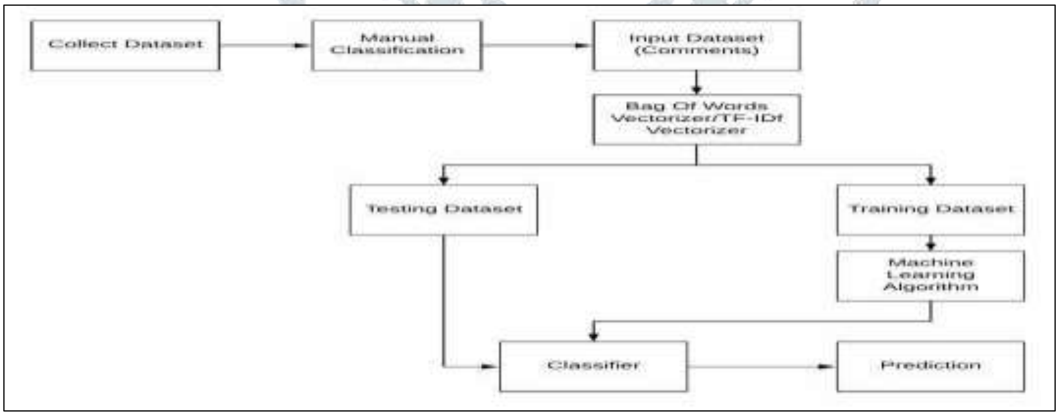
1. Training system on a dataset of most popular YouTube channels to label spam comments.
2. Testing system for the test dataset.
3. Testing system on actual spam data with a presentable user interface.

❖ MOTIVATION

Recently, YouTube used a monetization system to reward producers, to stimulate them to make high-quality original content and increase value to be seen. After the deployment of this system, the platform was flooded by undesired content, usually of low-quality information known as spam. Among different kinds of undesired content, YouTube is experiencing problems managing the huge volume of unwanted text comments posted by users that aim to advertise their videos or to disseminate malicious links to steal private data. The YouTube spam is directly related to the attractive profit provided by the monetization system.

IV. PROPOSED SYSTEM

1. The working is separated into three main stages: Initial, Middle, Last stage.
2. The Initial stage is identified with Data Exploration, Data Cleaning, and Data Transformation.
3. The central stage incorporates data modeling.
4. The final stage comprises data analysis using three models viz. KNN Algorithm, Linear Regression, and SVM.
5. Data exploration is similar to initial data analysis, visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.
6. Data Cleaning is the process to remove what is wrong or wrong records from a recordset, table, or database and refers to recognizing incomplete, incorrect, inaccurate, or unimportant parts of the data and then restoring, altering, or deleting the dirty or coarse data.
7. Data transformation is the process of conversion of data from one format to another, usually from one format of a source system into the required format of a destination system.



4.1 ADVANTAGES

1. When consideration of independent predictions catches on, Of course, the Naive Bayes classifier performs well as compared to other models.
2. Naive Bayes requires a limited amount of training data for the estimation of the test data. So, the training period is less.
3. Naive Bayes is also easy to implement.

4.2 APPLICATIONS

1. Social Network.
2. Spam detection

V. CONCLUSION

Social media networks have become extremely popular and this creates the opportunity for the malicious user to publish unwanted comments. This study has presented a feature set that will be used to detect video spammers available in the YouTube media. The features will be constructed based on the features obtained from the user profile and the content that they shared. Based on the undertaken experiments, it is expected that existing classifiers that were widely used in the data mining community could utilize the features in detecting comment spammers

VI. REFERENCES

[1] Chao Chen, Jun Zhang, Yi Xie, and Yang Xiang, "A overall performance assessment of machine gaining knowledge of-primarily based streaming unsolicited mail tweets detection," in IEEE transaction on the computational social machine, 2015, Vol-2 No-three.

[2] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "aiding the detection of faux money owed in big scale social on-line services," in Proc. Symp. Netw. Syst. Des. put in force. (NSDI), 2012, pp. 197–210.

[3] J. song, S. Lee, and J. Kim, "spam filtering in Twitter the use of sender-receiver dating," in Proc. 14th Int. Conf. recent Adv. Intrusion Detection, 2011, pp. 301–317.

[4] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammer on Twitter," in the 7th Annu. Collab. Electron. Messaging Anti-Abuse unsolicited mail Conf., Redmond, WA, united states of america, 2015.

[5] k. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honey pots + system studying," in. Proc 33rd Int. ACM SIGIR Conf. Res. increase. Inf. Retrieval, 2010, pp 435-442.

[6] Nathan Aston, Jacob Liddle, and Wei Hu*, "Twitter Sentiment in facts Streams with Perceptron," in journal of pc and Communications, 2014, Vol-2 No-eleven

JETIR