Spam Detection for Youtube Comments

M. Nirmaladevi, S. Saraswathi, R. B. Abishri, S. Divya Bharathi, B. Surya Prakash

M. Nirmaladevi, Student, Department of computer science and application, (E-mail: nirmalanisha123@gmail.com)

DR. S. Saraswathi, Professor and Head,Department of computer science and application(E-mail: saraswathis@skasc.ac.in)

R.B. Abishri, Student, Department of computer science and application,(E-mail: abishrirb@gmail.com)

S. Divyabharathi, Student, Department of computer science and application,(E-mail: divyabharathis4481@gmail.com)

B. Surya Prakash, Student, Department of computer science and application,(E-mail: suryaprakashmylife14@gmail.com)

Abstract:

Spamming is the use of messaging system, to send an unsolicited message. YouTube is one of the biggest sites for the user to get information. The best thing about the YouTube is the user can subscribe the channel, like or dislike the video and also giving opinion on the comment section on that video, and YouTube has attracted to increase the number of users. This attracts the spammers by spamming the comments. In this paper the spam comments on YouTube offers limited tools for comment moderation, so that the spam volume is shockingly increasing which leads to the owners to disable the comment section in their video. By using machine learning the comments are detected and prevented. Sklearn, numpy, pandas these are the some of the packages for understanding purpose of machine learning. YouTube spam comment has potential to spread malware through comment fields, which will exploit vulnerabilities in the user's machines. There are many approaches to detect Spam such as Artificial intelligence (AI), Machine Learning (ML), cryptography and others. A you tube has an automatic spam filter; this is the challenge even for the classification method. YouTube classifier classifies the comments by using classification methods as very short, symbols, abbreviations, links.

Keywords— Framework detection, spam detection approaches, data collection, pre-processing, feature selection and extraction, classification

I. INTRODUCTION

YouTube is a video content for famous publication platform with social network, and which supports for posting text comments this provides the interaction between the channel owner (producers) and the viewers. Recently, you tube has embrace a monetization system to channel owner (producers) to make high quality content and increasing the viewers in amount of visualization. By defining the spam comments as comments which have propaganda purpose which is not related to the content to which the comments are posted. In order to detect the spam comments, several techniques can be used in the spam detection. These techniques were divided into two types: Prevention technique, Detection technique. Machine learning algorithms can be used to build detection models. The prevention technique includes registering the users of the website, CAPTCHA, limiting the comments number for a specific post, control the comments number for the users or the users IP address, blocking IP addresses and limiting the time

for commenting. The set of features are assessing for detecting spams:

- Hand-engineered features(Hef)i
- Keyword features(Kf)j
- Text features(Tf)k

$$\sum i = \sum j + \sum k$$

Let "i" be considered as an (Hef) Hand-engineered features, "j" be considered as an (Kf) Keyword features and "k" be considered as an (Tf) Text features. The detection technique is used to classify the comments into spam comments (sc) and ham comments (hc).

$$\sum y = \sum s + \sum h$$

Let "y" be consider as an (YC) You tube Comments, "s" be consider as an (sc) spam comments and "h" be

consider as an (hc) ham comments. In which YouTube comments (yc) will have both spam comments (sc) and ham comments (hc) which was loaded by the you tube users. YouTube has a prospective to unfurl malware through comment.

According to statistical More than 800 million people uses the YouTube in every month, they used to watch 3 million hours of video in a month. Online videos are considered as passive consumption like a social exchange. According to a YouTube platform, it has more than 1 billion users in a week, 500 hours of video are unloaded in every 60 second and it produces more than billion views in one single day. Around 70% of viewers are from the outside country and more views only by using the mobile.

II. METHODOLOGY
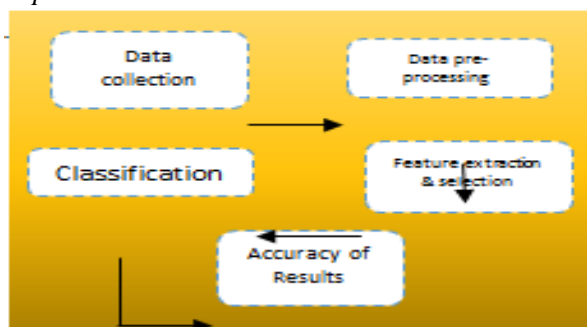
A. Spam Detection Framework

Figure 1: You Tube spam detection framework

There are five steps in the framework detection. Some of the five steps are Data collection, Data processing, Feature Selection and feature extraction, Classification and comparison of results accuracy. These are some of the methodology for the detection framework and which helps us to provide the good result accuracy.

- *Data collection*

Data collection is the main procedure need to be done first in this methodology detection framework. Data collection is done for the reference of machine learning which produces the prediction of future. Datasets download and collected from the UCI machine learning repository. Finally, the datasets contain 5 downloaded video from YouTube through API. The comments are taken from PSY, Katy Perry, LMFAO, Eminem, and Shakira. The total number of spam and ham comments are taken in PSY is 351, Katy Perry contains 351, LMFAO contains 439, Eminem contains 449, and Shakira contains 371.

- *Pre-processing*

Pre-processing procedure is used for the raw datasets which will be execute the data cleaning such as tokenization, stop words removal and stemming are performed in this Pre-processing technique. Clean datasets are mainly used for the next procedure feature of selection and extraction.

- *Feature selection and extraction*

Feature selection and extraction are the procedure which is extracted from the Pre-processing technique. This is the process before a classification class. Some of the suitable features are derived from the datasets.

- *Classification*

Classification is a procedure which is mainly used for the training and testing process. Then out of 100, 60% is for training and another 40% is used for testing. The datasets needs are trained based on machine learning techniques.

- *Accuracy of results*

The result will be produced by 3 terms Precision, Re-call and F-measure.

B. Spam Detection approaches

You tube are not excluded from malicious user, in which often found to expose in spamming and promotional activities. Some of the approaches to detect a spam cryptography, Machine learning and others. These are the detection done by the you tube Detection of inappropriate content on YouTube, Detection of malicious user behaviour. The YouTube spam detection can have an existing study in Manwar and Alberto by using the (SVM) Support vector machine as a classifier in classification. Manwar started the SVM classification in binary-two class. Class are denoted by 0 and 1.

e	Comment	User behaviour
0	0	Safe uploaders, safe uploaders
0	1	Safe uploaders, unsafe uploaders
1	0	Unsafe uploaders, safe uploaders
1	1	Unsafe uploaders, unsafe uploaders

Table 1: Possible combination of user behaviour

$$K = (\sum l + \sum m) + (\sum o + \sum p)$$

Let "k" be the user behaviour, "l" be the safe uploaders, "m" be the safe content, "o" be the unsafe uploaders, "p" be the unsafe content You tube has a safe content and the unsafe content which are published by safe uploaders and unsafe uploaders or in a form of inappropriate words in comments. Let user

behaviour consider here by "1" represents an unsafe uploaders and "0" represents a safe uploader.

To detect the you tube spam comments by using machine learning we need enormous data, so we collect the data from the various sources. And store it in the particular table for the reference to predict the future results. Some of the simple example for the data collection machine learning.

COMMENT_ID	AUTHOR	DATE	CONTENT	CLASS
LZQPQhLy	Julius NM	2013-11-07T06:20:48	Huh, anywaycheck	1
z13lfzdo5vm	ferleckferles	2013-11-27T21:39:24	Subscribe to my channel ?	1
z13bgdvy	Zielimeek21	2013-11-28T21:49:00	I'm only checking	0
z12ntlcqht	Owen Lai	2013-12-01T04:51:52	just checking the views?	0
z121zxaxsq	Uploadaddict	2013-11-28T18:12:12	subscribe comment?	1
z13kxpqqssa	Tasha Lucius	2014-01-19T13:25:5	2 billion.. views	0

Table 2: Data collections for machining learning.

## III. RESULTS AND DISCUSSION

A. Framework Detection

In detection frame work there are several phases by using machine learning technique such as data collection, feature selection, classification and detection.

The datasets are collected from the social media such as you tube, email, Facebook, twitter and Instagram. The UCI will collect those comments and form a dataset according to the social media categories. By using the collected data, the machine check the given comments are spam or ham. The raw data was already classified as spam and ham.

Raw Data	Spam	Ham
URL links	32.8%	95.1%
Words	67.2%	4.9%

Table 3: Datasets from social media
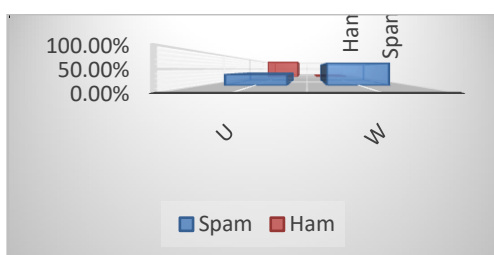
Graph 1: Datasets from social media

"Table3" represents the datasets which are collected and used in the experiments are reported in this paper, and amount of samples in spam and ham are given, total number of spam and ham comments are also given in this table. Each sample represents the text comments which can be posted in the comment section for each selected video

Data sets	Spam	Ham	Total
Psy	175	176	351
Katy Perry	175	176	351
LMFAO	233	206	439
Eminem	239	210	449
Shakira	170	201	371

Table 4: Datasets for You Tube spam comments

$$A = \sum p + \sum kp + \sum lm + \sum em + \sum sh$$

Let "A" be consider as Data sets, "p" be consider as "Psy", "kp" be consider as Katy Perry, "lm" be consider as "LMFAO", "em" be consider as "Eminem" and "sh" be consider as "Shakira". Data sets contains "p", "kp"," lm"," em", "sh" by total 1961 data. In which "p" and "kp" contains 351 data, "lm" contains 439 data, "em" contains 449 data, "sh" contains 371 data.

B. Comparison of results

The result will be produced by 3 terms Precision, Re-call and F-measure.

$P = t / (f + t)$

Let "P" be consider as an "Precision", "t" be consider as an "true statement", "f" be consider as an "false statement".

$R = t / (f + t)$

Let "R" be consider as an "Re-call", "t" be consider as an "true statement", "f" be consider as an "false statement".

$F = 2 * R * P / (R + P)$

Let "F" be consider as an "F- measure", "P" be consider as an "Precision", "R" be consider as an "Re-call".
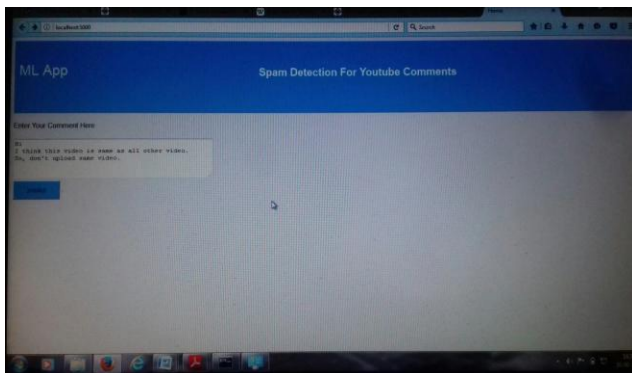

Figure 2: command prompt
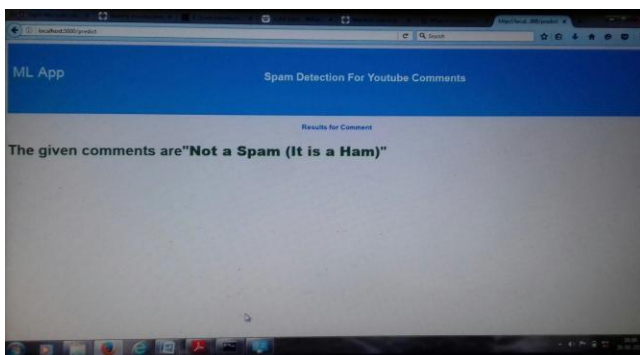

Figure 3: Comment page

Figure 4: Result page

IV. CONCLUSION

The goal of this paper is to improve the ham comments in high accuracy and to avoid the unwanted spam comments on the You tube. The result of this paper provides the baseline for the people who interested in the you tube spam comments and improve the results for the future comparison. The YouTube spam comments are collected from social network data. This data will be going to test with data mining tool for the comparison of result accuracy.

REFERENCE

1. Aqliima Aziz1, CikFeresaMohd Foozy2, Palaniappan Shamala3, Zurinah Suradi4 YouTube spam comment detection using support vector machine and K–nearest neighbour. (ACT) Faculty of Computer Science and Information Technology, (UTHM).
2. K Subba Reddy; 2 Dr E Srinivasa Reddy a Survey on Spam Detection Methodologies in Social Networking Sites, Research Scholar, Computer Science and EngineeringAcharyaNagarjuna University, Guntur, AP, India, Principal, Anucet, Computer Science and EngineeringAcharyaNagarjuna University, Guntur, Ap, India
3. RishabhKaushal_, SrishtySaha_, PayalBajajy and PonnurangamKumaraguru_ KidsTube: Detection, Characterization and Analysis of Child Unsafe Content & Promoters on YouTube_ Cybersecurity Education and Research Centre (CERC)Indraprastha Institute of Information Technology, Delhi.
4. Tube Spam: Comment Spam Filtering on YouTube T´ulio C. Alberto, Johannes V. Lochter, Tiago A. AlmeidaDepartment of Computer ScienceFederal University of S˜ao Carlos – UFSCar
5. How Useful Are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings Stefan Siersdorfer, SergiuChelaru,Wolfgang Nejdl L3S Research Center
6. Scheltus, P., Dorner, V., &Lehner, F. (2013). Leave a Comment! An In-
7. Depth Analysis of User Comments on YouTube. Wirtschaftsinformatik, 42.
8. Manwar, S. R., Lambhate, P., &Patil, J. (2017). Classification Methods for Spam Detectio in Online Social Network.
9. Alberto, T. C., Lochter, J. V., & Almeida, T. A. (2015, December). Tube spam: Comment spam filtering on you tube. In Machine Learning and Applications (ICMLA), 2015 IEEE 14th International Conference