



SURVEY: YOUTUBE SPAM COMMENTS DETECTION AND DELETION

Darshan Bhavsar, Abhishek Chandekar, Prof .Chetana Patil, Mihir Chaudhari, Stephen Dcruz
Department of Computer Science
Dhole Patil College of Engineering
Pune, India

Abstract— The quality of online social platforms has risen, and spammers have developed a number of strategies to trick consumers into clicking on harmful links. This is accomplished by posting spam to various social media networks' comment sections. This paper suggests a method for identifying spam comments on YouTube, which have grown significantly recently. YouTube operates a spam blocking system, although it consistently fails to do so effectively so our program tries to tackle the problem by not only detecting but also deleting the spam comments by using Machine Learning.

Keywords— YouTube, Spam Comments, Spam Detection, Machine Learning

I. INTRODUCTION

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning is important because it gives enterprises a view of trends in customer behaviour and business operational patterns, as well as supports the development of new products.

Spam is the unnecessary information that needs to be removed both to prevent wasting the users' time and to preserve the platform's reputation for quality and reliability. Whereas Ham is a crucial and necessary ingredient for varied objectives. The biggest video-sharing website in the world, YouTube, was established in 2005, and Google acquired it in 2006. on YouTube evolved significantly as a venue for video content, with the video has recently become more prevalent online.

Currently, more than 4.5 million videos totalling 400 hours are uploaded watched on YouTube every minute. Recently, there has been a massive increase of spam on YouTube where fake impersonator accounts leave spam/ scam replies to hundreds of users on a creator's videos. For some god-forsaken reason, YouTube offers no way to delete all comments by a specific user at once, meaning you must delete them one by one BY HAND. YouTube offers a functionality to ban a user, but it does NOT delete previous comments. YouTube offers a functionality to ban a user, but it does NOT delete previous comments.

II. RESEARCH

Research on identifying spam users and content focuses on a variety of areas. Numerous research looked into website spam (e.g., portal sites and blogs). Spammers target YouTube with low-quality content or advertisements as it becomes more and more popular as a platform for sharing videos. The number of spammers harming the YouTube community is growing, making it interesting to conduct study on how to identify them. Therefore, we carry out the research on identifying spam Comments on YouTube Platform.

Applying the same strategy (i.e., language modelling) to YouTube comments is unsuccessful since the data's attributes are different. Less textual descriptions and information are represented by features of YouTube comments. They don't really relate to the information in the video. So, in order to find spam on YouTube, a different strategy must be utilised. Some research use classification algorithms to identify spam videos or a collection of comments and categorise them as real or spam.

III. PROPOSED SYSTEM

The proposed system is basically python program which aims to detect and delete the spam comments on a particular YouTube channel. This system makes use of some supervised learning models used in machine learning. The system gains access to the comments using the YouTube API by help of Google Cloud Platform.

This system aims to achieve the features such as:

- 1) Automatically Detects multiple spammer techniques
- 2) Scan comments by entering the spammers channel ID
- 3) Scan Usernames for non ASCII special characters

There are different scanning modes in the system such as scan a single video, scan recent comment across entire channel (all videos), scan a community post

It tries to achieve automatic deletion of all found spam comments. The system will include option to report spam comments or Hold for View. Another feature that can be added in the system is recovery mode through which we can re-instate previously deleted comments.

Spam comments can also be viewed and analysed how accurate the particular spam comments are detected as spam or not.

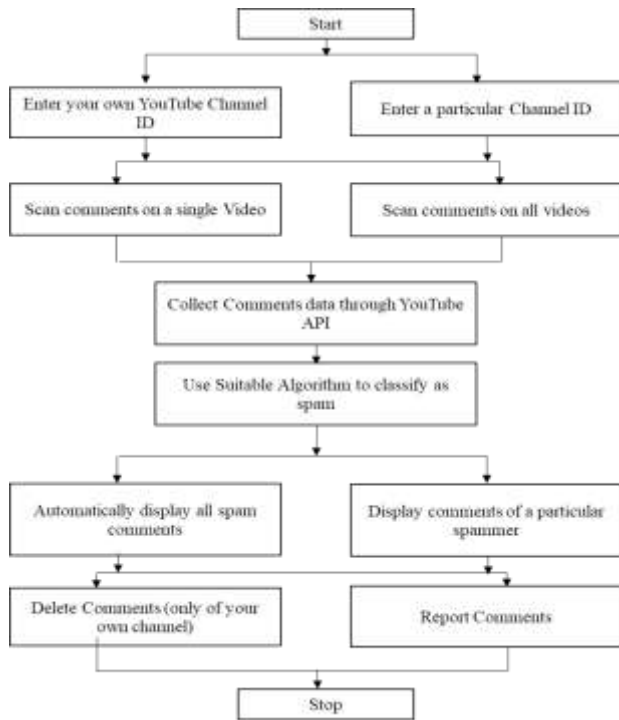


Fig 1: System Diagram

IV. MACHINE LEARNING ALGORITHMS

1) Support Vector Machine:

Data is divided into groups of comparable class values using the Support Vector Machine approach, which also uses surface boundaries to build hyperplane boundaries between points in multidimensional space. In other words, the objective of SVM is to produce a fairly homogeneous split on both sides of a flat boundary known as a hyperplane that divides space. As one of the most widely used machine learning algorithms, SVM is a versatile machine learning model because it can be applied to a variety of tasks, including linear and nonlinear classification and regression.

The following describes how dimensional SVM works. Finding the maximum margin hyperplane straight line among the numerous dimensions is the task of the SVM algorithm.

The maximum margin is used to divide straight lines that divide the two classes. The maximum margin enhances the likelihood that the point will stay on the right side of the boundary even if there is random noise, and the straightest line that separates with the largest margin will be the most generalised for future data. In other words, the points closest to the maximum margin hyperplane MMH in each class are what Support Vectors exactly signify. SVM selects the extreme vector to find the hyperplane, and these vectors are known as support vectors. Even with a large number of features, SVM provides a compact way to store classification models, and it is easier than using neural networks.

2) Logistic Regression:

Algorithm for classifying data called logistic regression is typically employed when there are several categories to classify. When the dependent variable is categorised as a binary response variable, logistic regression is very useful. For example, it can be used to determine if a disease exists or not based on risk variables like the prevalence of diabetes caused by obesity or the response variable. Additionally, it is employed in statistical analyses of mortality or survival when postoperative survival rates are represented by probabilities with values ranging from 0 to 1.

3) Naïve Bias:

A supervised learning technique called Naive Bayes uses the idea of Bayes to classification issues. Prior to using the classifier, training must be carried out using a training vector that determines the likelihood that the outcome will be seen based on the evidence shown by the feature values. Additionally, it categorises things by determining which class new data should go into and predicting it with the best degree of certainty. The benefits include good handling of noise and missing data as well as high storage and processing efficiency. This is so that it may incorporate both straightforward, quick, and supported feature vectors.

Additionally, the training needs only the number of layers and characteristics specified during training a case, although it performs exceptionally well for huge cases. In particular, it is well known for its great performance in practise since the estimation probability for prediction may be simply determined. However, because each attribute of the datasets is equally significant and independent, they are occasionally unsuitable for practical applications.

V. CONCLUSION

Thus after carrying out the necessary research we try to create a program that will help the YouTube community to identify the spam comments on their channels and hence help them to detect and delete using our tool.

Also our program will include a unique feature that enables us to scan and view the spam comments not only our channel but also the channels of other users which we wish to scan by gaining access to their channel IDs. Our program also proposes several methods or options using which the users can detect the spam comments as per their choice. Therefore by creating the above mentioned program/tool we aim to reduce spammers on the YouTube platform, thereby helping the YouTube Community from being victim to spammers.

VI. REFERENCES

- [1]. S. Aiyar and N. P. Shetty, "N-gram assisted YouTube spam comment detection," *Proc. Computer. Sci.*, vol. 132, pp. 174—182, Jan. 2018, doi:10.1016/j.procs.2018.05.181.



- [2]. Alberto, J. V. Lochter, and T. A. Almeida, "TubeSpam: Comment spam filtering on Youtube," in Proc. IEEE 14th Int. Conf. Mach. Learn.ppl. (ICML4), Dec. 2015, pp. 138-143, doi: 10.1109/ICMLA.2015.37.
- [3]. S. Sharmin and Z. Zaman, "Spam detection in social media employing machine learning tool for text mining," in Proc. 13th Int. Conf. Signal-image Technol. Internet-Based Syst. (SITIS), Dec. 2017, pp. 137—142, doi:10.1109/SITIS.2017.32.
- [4]. A. O. Abdullah, M. A. Ali, M. Karabatak, and A. Sengur, "A comparative alysis of common Youtube comment spam filtering techniques," in Proc.th Int. Symp. Digit. Forensic Secur. (ISDFS), Mar. 2018, pp. 1—5, doi:10.1109/ISDFS.2018.8355315.
- [5]. S. Jain and D. M. Patel. (2019). Analyzing User Comments of Learning Videos From YouTube Using Machine Learning.
- [6]. A. Aziz, C. F. M. Foozy, P. Shamala, and Z. Suradi, "Youtube spam comment detection using support vector machine and k-nearest neighbor," Tech. Rep., 2018, doi: 10.11591/ijeecs.v12.i2.pp607-611.
- [7]. L. song, R. Y K. Lau, R. C.-W. Kwok, K. Mirkovskl, and w Dou, "Who are the spoilers in social media marketing? Incremental learning of latent semantics for social spam detection," Electron. Commerce Res., vol. 17, no. 1, pp. 51-81, Mar. 2017, doi: 10.100%10660-016-9244-5.
- [8]. R Abinaya ,Bertilla Niveda E ,P Naveen "Spam Detection On Social Media Platforms" IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020 doi: 978-1-7281-7223-1/20/ICSSS
- [9]. Hayoung Oh "A YouTube Spam Comments Detection Scheme Using Cascaded Ensemble Machine Learning Model" doi: 10.1109/ACCESS.2021.3121508
- [10]. A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. Joseph, and J. D. Tygar, "Robust detection of comment spam using entropy rate," in Proc.th ACM Workshop Secur. Artif Intell. (AISec), 2012, pp. 59—70, doi: 10.1145/2381896.2381907.
- [11]. A. Madden, I. Ruthven, and D. McMenemy, "A classification scheme for content analyses of YouTube video comments," J. Documentation, vol. 69, no. 5, pp. 693-714, Sep. 2013, doi: 10.1108/JD-06-2012-0078.
- [12]. E. Poche, N. Jha, G. Williams, J. Staten, M. Vesper, and A. Mahmoud, "Analyzing user comments on YouTube coding tutorial videos," in Proc. IEEE/ACM 25th Int. conf Program Comprehension (ICPC), May 2017, pp. 196-206, doi: 10.1109/ICPC.2017.26.
- [13]. R. K. Das, S. S. Dash, K. Das, and M. Panda, "Detection of spam in YouTube comments using different classifiers," in Advanced Computing and Intelligent Engineering, 2020, pp. 201—214, doi: 10.1007%2F978- 981-15-1081-6 17.
- [14]. N. M. Samsudin, C. F. B. Mohd Foozy, N. Alias, P. Shamala, N. E Othman, and W. I. S. Wan Din, "YouTube spam detection framework using Naïve Bayes and logistic regression," Indonesian J. Electr. Eng. Comput. Sci., vol. 14, no. 3, p. 1508, Jun. 2019, doi: 10.11591/ijeecs.v14.i3.pp1508-1517.
- [15]. N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Predilection decoded: Spam review detection techniques: A systematic literature review," Appl. Sci., vol. 9, no. 5, p. 987, Mar. 2019, doi:10.3390/app9050987.
- [16]. P. Bansal. (2019). Detection of Offensive YouTube Comments, a Performance Comparison of Deep Learning Approaches. [Online]. Available: <https://core.ac.uk/reader/301313034>