IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204325 |

Cascaded Ensemble Deep Learning Model to Detect Youtube Spam Comments

Dr. D.J. Samatha Naidu[1], K.Venkata Ramya[2], P.Mahesh Babu[3]

Principal, Dept. of MCA, Annamacharya PG College of Studies, Rajampet, India

Assistant Professor, Dept. of MCA., Annamacharya PG College of Computer Studies, Rajampet, India

PG Student, Dept. of MCA., Annamacharya PG College of Computer Studies, Rajampet, India

ABSTRACT: This paper proposes a technique to detect spam comments on YouTube, which have recently seen tremendous growth. YouTube is running its own spam blocking system but continues to fail to block them properly. Therefore, we examined related studies on YouTube spam comment screening and conducted classification experiments with six different machine learning techniques (Decision tree, Logistic regression, Bernoulli Naïve Bayes, Random Forest, Support vector machine with linear kernel, Support vector machine with Gaussian kernel) and two ensemble models (Ensemble with hard voting, Ensemble with soft voting) combining these techniques in the comment data from popular music videos - Psy, Katy Perry, LMFAO, Eminem and Shakira.

KEYWORDS: spam comments, deep learning, screening,

I. INTRODUCTION

YouTube, the world's largest video sharing site, was founded in 2005 and acquired by Google in 2006. YouTube has grown tremendously as a video content platform, with the recent shift in online content to video. At present, more than 400 hours of video are uploaded and 4.5 million videos are watched every minute on YouTube. It is easy for users to watch and upload videos without any restrictions. This great accessibility has increased the number of personal media, and some of them have become online influencers.

YouTube creators can monetize if they have more than 1,000 subscribers and 4,000 hours of watch time for the last 12 months. Accordingly, spam comments are being created to promote their channels or videos in popular videos. Some creators closed the comment function due to aggression such as political comments, abusive speech, or derogatory comments not related to their videos.

YouTube has its own spam filtering system, though there are still spam comments that are not being caught. In this paper, we review related studies on YouTube spam comments and propose the Cascaded Ensemble Machine Learning Model aware YouTube Spam Comments Detection Scheme to improve the performance of the model. In previous studies, various machine learning techniques were applied to each dataset to detect spam comments and compare their performance. Therefore, in this paper, we propose an ensemble machine learning method that combines the results of several models to produce the final result.

This paper is organized as follows: In Section 2, we review related work. Section 3 describes the system model and proposed techniques, and Section 4 describes the experiments and results. Then we conclude in Section 5.

II. RELATED WORK

Most prior work on more general OM has been carried out on more standardized forms of text, such as consumer reviews or newswire. The most commonly used datasets include: the MPQA corpus of news documents (Wilson et al., 2005), web customer review data (Hu and Liu, 2004), Amazon review data (Blitzer et al., 2007), the JDPA 1The corpus and the annotation guidelines are publicly available at: http://projects.disi.unitn. it/iKernels/projects/sentube/ corpus of blogs (Kessler et al., 2010), etc. The aforementioned corpora are, however, only partially suitable for developing models on social media, since the informal text poses additional challenges for Information Extraction and Natural Language Processing. Similar to Twitter, most YouTube comments are very short, the language is informal with numerous accidental and deliberate errors and grammatical inconsistencies, which makes previous corpora less suitable to train models for OM on YouTube. A recent study focuses on sentiment analysis for Twitter (Pak and Paroubek,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204325 |

2010), however, their corpus was compiled automatically by searching for emoticons expressing positive and negative sentiment only.

Siersdorfer et al. (2010) focus on exploiting user ratings (counts of 'thumbs up/down' as flagged by other users) of YouTube video comments to train classifiers to predict the community acceptance of new comments. Hence, their goal is different: predicting comment ratings, rather than predicting the sentiment expressed in a YouTube comment or its information content. Exploiting the information from user ratings is a feature that we have not exploited thus far, but we believe that it is a valuable feature to use in future work.

Most of the previous work on supervised sentiment analysis use feature vectors to encode documents. While a few successful attempts have been made to use more involved linguistic analysis for opinion mining, such as dependency trees with latent nodes (Tackstr ¨ om and McDonald, ¨ 2011) and syntactic parse trees with vectorized nodes (Socher et al., 2011), recently, a comprehensive study by Wang and Manning (2012) showed that a simple model using bigrams and SVMs performs on par with more complex models.

In contrast, we show that adding structural features from syntactic trees is particularly useful for the cross-domain setting. They help to build a system that is more robust across domains. Therefore, rather than trying to build a specialized system for every new target domain, as it has been done in most prior work on domain adaptation (Blitzer et al., 2007; Daume, 2007), the domain adapta- ´ tion problem boils down to finding a more robust system (Søgaard and Johannsen, 2012; Plank and Moschitti, 2013). This is in line with recent advances in parsing the web (Petrov and McDonald, 2012), where participants where asked to build a single system able to cope with different yet related domains.

III. PROPOSED ALGORITHM

A. EXPERIMENT METHOD AND ENVIRONMENT

Our proposed method is based on comparative research [9] which is a representative study on YouTube spam comment detection. Our method applied six machine learning techniques (i.e., CART (Decision Tree), LR (Logistic Regression), NB-B (Bernoulli Naïve Bayes), RF (Random Forest), SVM-L (Support vector machine with linear kernel), and SVM-R (Support vector machine with Gaussian kernel)) to improve the performance of the Cascaded Ensemble Machine Learning Model aware YouTube Spam Comments Detection Scheme. These performed well in [9] and were significant with 99.9% confidence. We propose an ensemble model combining them and evaluate the performance. The experimental environment used version 3.7.1 of Python and version 0.20.1 of the Cicely Library on Jupiter notebooks [13]–[15].

B. EXPERIMENTAL OVERVIEW OF THE PROPOSED TECHNIQUE

As shown in Figure 1, we collect 1,983 comments and distinguish 1,369 (70% of total) into training data and 587 (30%) into test data. To classify the spam data, we remove stop words such as articles (i.e., the, a, an) and pronouns (e.g., I, you, it). Additionally, in [9], only BoW vectorization was performed. In this paper, TF-IDF vectorization preprocessing is used to solve the issue that BoW may not find significant meaning in a sentence because it appears frequently in other sentences. References [12] suggests that there is no single technique that performs well on all datasets. The ensemble model achieved good performance in [9]. We carry on the experiment with multiple techniques to find the best classification algorithm, using six machine learning algorithms (i.e., CART, LR, NB-B, RF, SVM-L, SVM-R). We use two ensemble models, ESM-H (Ensemble with hard voting) and ESM-S (Ensemble with soft voting), to train and test our dataset. They predict and evaluate the class**.**

C. DATASETS

We use open datasets, which can be downloaded from [16]. They consist of comment data on five popular music videos provided in [9]. They contain YouTube ID, comment author, date, comment content, and labeled class (0: Ham or 1: Spam). We only use comment content and labeled class. Each training and testing of the five data sets as shown in Table.1 can result in overfitting, where the five classifiers perform well only on that data and do not apply well to comment data in other videos. Therefore, in this paper, to generalize the result, we include all five video's datasets. As shown in Fig. 1, we employ 1,983 comments with 1,369 comments (70%) for training and 587 comments (30%) for testing.

International Journal of Innovative Research in Computer and Communication Engineering

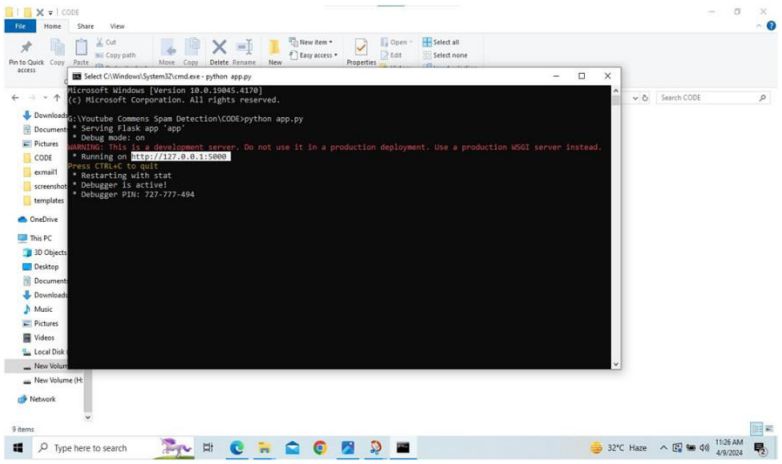
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

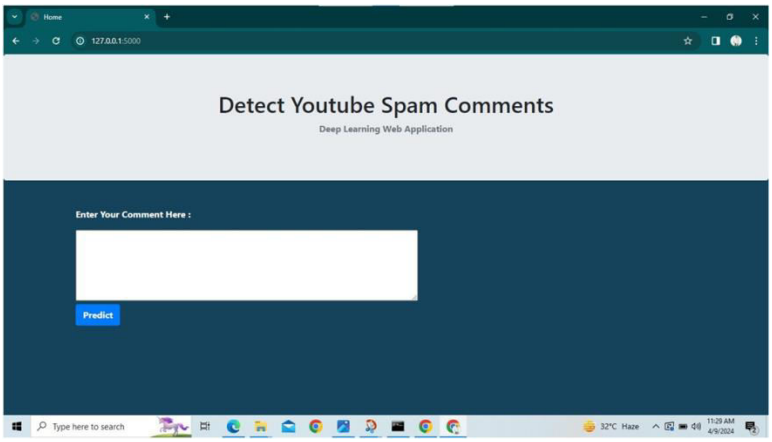
| DOI: 10.15680/IJIRCCE.2024.1204325 |

IV. RESULT ANALYSIS

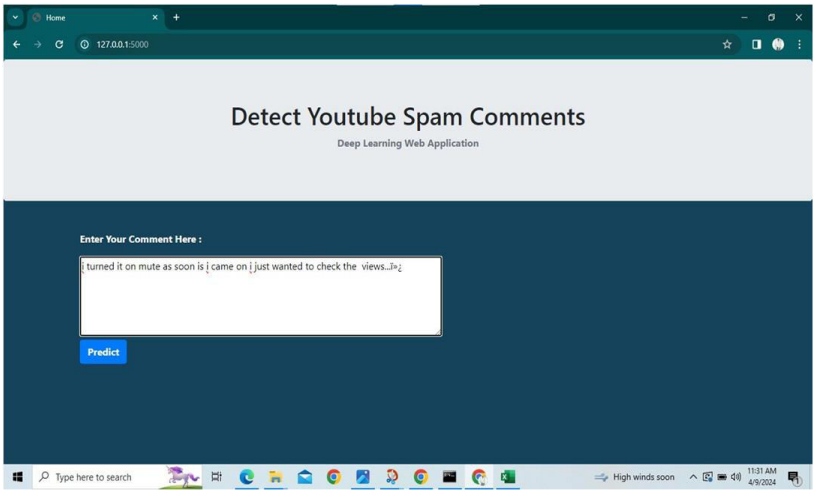
SCREEN 1:GENERATING LINK IN CMD



SCREEN 2:HOME PAGE



SCREEN 3:ENTER A COMMENT



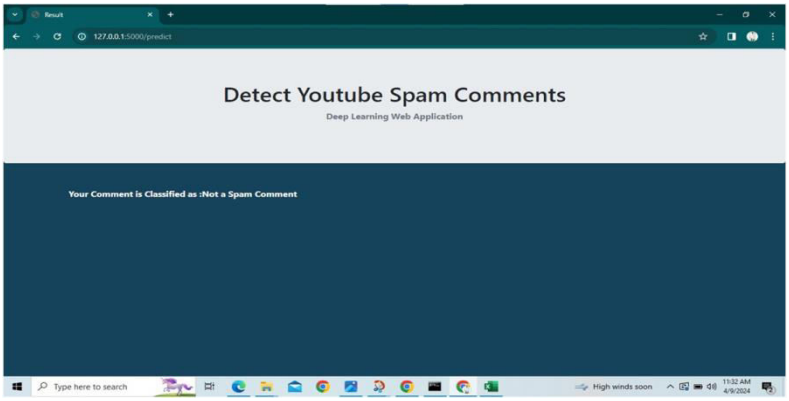
International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204325 |

SCREEN 4:RESULT PAGE



V. CONCLUSION

For classifying the YouTube comments as spam and not spam (ham) there are various techniques used. This approach has been tested with real-time YouTube comments and given an overall outcome which is 18% more accurate than the existing approach. As YouTube API is open platform to all users, it might change the behavior of spammers over the period of time. In real world, YouTube spam feature will not be constant it keeps on changing an precipitous way..

REFERENCES

[1] S. Aiyar and N. P. Shetty, ‘‘N-gram assisted Youtube spam comment detection,’’ Proc. Comput. Sci., vol. 132, pp. 174–182, Jan. 2018, doi: 10.1016/j.procs.2018.05.181.
[2] A. Kantchelian, J. Ma, L. Huang, S. Afroz, A. Joseph, and J. D. Tygar, ‘‘Robust detection of comment spam using entropy rate,’’ in Proc. 5th ACM Workshop Secur. Artif. Intell. (AISec), 2012, pp. 59–70, doi: 10.1145/2381896.2381907.
[3] 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.
[4] A. Severyn, A. Moschitti, O. Uryupina, B. Plank, and K. Filippova, ‘‘Opinion mining on Youtube,’’ in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Long Papers), vol. 1, 2014, pp. 1–10, doi: 10.3115/v1/P14- 1118.
[5] M. Z. Asghar, S. Ahmad, A. Marwat, and F. M. Kundi, ‘‘Sentiment analysis on Youtube: A brief survey,’’ 2015, arXiv:1511.09142. [Online]. Available: http://arxiv.org/abs/1511.09142
[6] T. C. Alberto, J. V. Lochter, and T. A. Almeida, ‘‘TubeSpam: Comment spam filtering on Youtube,’’ in Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA), Dec. 2015, pp. 138–143, doi: 10.1109/ICMLA.2015.37.
[7] A. U. R. Khan, M. Khan, and M. B. Khan, ‘‘Naïve multi-label classification of Youtube comments using comparative opinion mining,’’ Proc. Comput. Sci., vol. 82, pp. 57–64, Jan. 2016, doi: 10.1016/j.procs.2016.04.009.
[8] J. Savigny and A. Purwarianti, ‘‘Emotion classification on Youtube comments using word embedding,’’ in Proc. Int. Conf. Adv. Informat., Concepts, Theory, Appl. (ICAICTA), Aug. 2017, pp. 1–5, doi: 10.1109/ICAICTA.2017.8090986.
[9] S. Sharmin and Z. Zaman, ‘‘Spam detection in social media employing machine learning tool for text mining,’’ in Proc. 13th Int. Conf. SignalImage Technol. Internet-Based Syst. (SITIS), Dec. 2017, pp. 137–142, doi: 10.1109/SITIS.2017.32.
[10] A. O. Abdullah, M. A. Ali, M. Karabatak, and A. Sengur, ‘‘A comparative analysis of common Youtube comment spam filtering techniques,’’ in Proc. 6th Int. Symp. Digit. Forensic Secur. (ISDFS), Mar. 2018, pp. 1–5, doi: 10.1109/ISDFS.2018.8355315.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details