NLP BASED ADVANCED METHOD OF DETECTING HARMFUL VIDEOS

Chapter I: Introduction

Online video promotion benefits greatly from the appeal and reputation of commercial, social, and academic realms. The most popular places for people to publish their own opinions and engage in debates are video platforms, such as various web portals. As technology advances, people may come across unethical videos containing politically incorrect content, as well as illegal sexual, criminal, or terrorism-related content that is publicly available and influences viewers all over the world on a psychological and moral level (Moon et al., 2021).

To find unnecessary video, a cutting-edge approach that combines machine learning and natural language processing is applied. The solution also includes a framework for analytics databases that can recognise, approve, or reject unnecessary and immoral video content as well as help with statistical correctness (Sehgal et al., 2020). Text may be recovered from video footage using the video detection system in two steps: first, video is converted to MPEG-1 audio layer 3 (MP3), and then MP3 is converted to WAV. The data gathered can be examined and prepared using the text portion of natural language processing (Jahan, 2020).

This method involves converting the video MP4 data to plain text data using a function from the Python Advance Library. Using verbal video record data, the methodology aids in the examination of the identification of unlawful, antisocial, pointless, unfinished, and malevolent videos. By analysing data sets, this model can be used to decide which movies should be approved or disapproved for future action.

Aim & Objectives

The aim of the research is to develop a video recognition system that can identify hazardous videos that are inappropriate for our day to day lives.

The main objective of the research is that,

1. To convert the selected video file to audio file and ensuring that the processed audio is free of any noisy data.
2. To convert the audio files into text files such that the written transcripts of the content is visible.
3. To understand the written transcripts into machine readable formats using Natural Language Processing.
4. To classify the transcripts into harmful and ordinary content in order to filter the videos with inappropriate content.

Chapter II: Literature Review

Constricted video streams have become more popular online over the past ten years as a result of business strategies that lock down video content behind geofences. Because of the practise, network administration now has to deal with more difficulties. The challenge is that classification of encrypted traffic is challenging due to the inaccessibility of traffic data. Prioritization and monitoring tasks get harder to do (Shi et al., 2021). More convincing and successful fake videos have been produced as a result of the rapid development of social media and deep generative networks. The dissemination of false information can result in catastrophe under urgent circumstances (Yesugade et al., 2020).

As technology develops, we occasionally all must deal with unethical videos that are based on politically incorrect as well as unauthorised content of sexual, criminal, or support for terrorism and the content publishing openly that have an impact on psychologically and morally on the users around the world. For the same, the author has employed both the Naive Bayes algorithm and the classifier method logistic regression. Therefore, the Naive Bayes accuracy value is 82%, and the Multinomial Logistic Regression accuracy value is 79% (Moon et al., 2021). Using deep learning-based classifiers, the author proposed a classification technique for locating the origin of heterogeneous network traffic flow in video streaming. The feature condenses the encrypted network traffic into a manageable text-like representation with little need for manual feature engineering. 95.7% of predictions are correct when employing deep learning. (Shi et al., 2021).

It is possible to distinguish between hazardous and typical data content, and while converting a video file to an audio file, no background noise is added. Additionally, the audio file is extracted so that the text can be used to detect any malicious content. As can be shown, deep learning algorithms perform more accurately than other algorithms. Therefore, deep learning algorithm is applied in the suggested work to obtain more effective and highly accurate results.

Chapter III: Methodology

Many people have given up on speaking their minds and seeking out opposing viewpoints due to the prospect of internet harassment and abuse. Due to platforms' failure to effectively promote public conversations, many communities have either severely restricted or completely eliminated user comments.

In order to solve these issues Challenge of Classifying Toxic Comments (Kaggle.com, 2021) dataset is used. This dataset makes it easier to spot unpleasant and insulting comments.

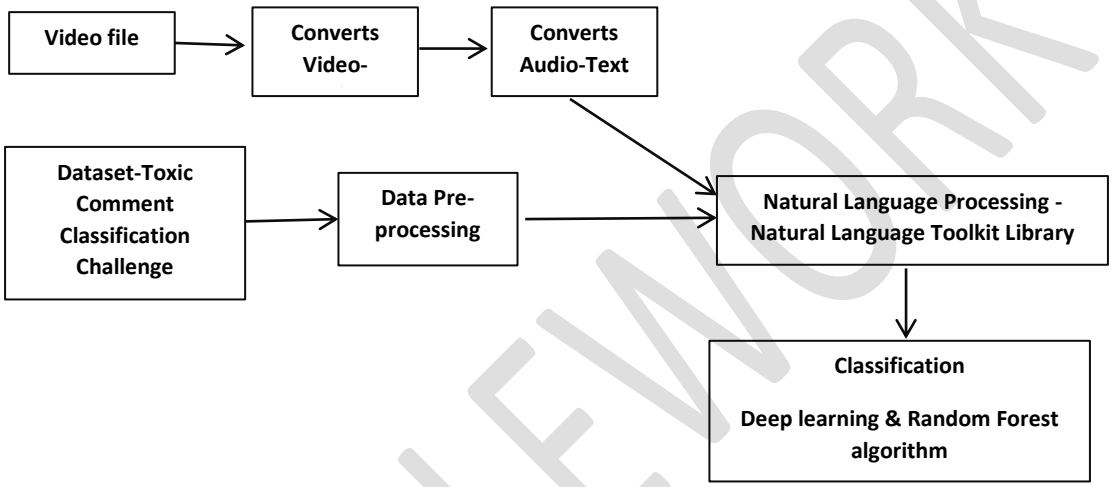
Data pre-processing is done once the dataset has been defined. As part of this process, extraneous tags and markers, background noise, stop words, syntactic analysis, vocabulary development, and annotation of potentially hazardous terms are removed.

On the other hand, the video file is changing from video to MPEG-1 audio layer 3 (MP3) and subsequently from MP3 to WAV to create an audio file. A multiplier is needed to boost the audio, which had previously been at noticeably lower levels because of background noise. Due to the dynamic nature of speech and noise signals, it is difficult to determine a constant multiplication factor (MF) for the incoming signal. The environment's signal and noise levels must therefore be examined. Based on the various characteristics of the speech and noise signals, the voice signal will be amplified using a dynamically changeable gain. The video file is being converted to text format after being converted to audio format.

The text file and preprocessed data are sent to natural language processing (NLP) in the following stage, where NLP makes use of the Natural Language Toolkit package (NTLK, 2020). It focuses on filtering through text for useful information and building data models based on the findings. also transforms free-text sentences into features that are structured. The Deep Learning algorithm is used in combination with the Random Forest method to classify data since it includes numerous layers that may be utilised to analyse the process at various levels. The test dataset was produced for training by extracting textual data from video material. The categorised text makes it easier to distinguish between dangerous and common content in transcripts. The aforementioned algorithm, which produces outcomes that are effective and have more accuracy, is used to carry out this operation.

Implementation Flow

Figure 1:



References

Jahan, M.S. (2020). Cyber Bullying Identification and Tackling Using Natural Language Processing Techniques. *FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING*. [Online]. pp. 1–75. Available from: http://jultika.oulu.fi/files/nbnfioulu-202008222869.pdf.

Kaggle.com (2021). *Toxic Comment Classification Challenge Identify and classify toxic online comments*. [Online]. 2021. Available from: https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/overview.

Moon, N.N., Salehin, I., Parvin, M., Hasan, M.M., Talha, I.M., Debnath, S.C., Nur, F.N. & Saifuzzaman, M. (2021). Natural language processing based advanced method of unnecessary video detection. *International Journal of Electrical and Computer Engineering (IJECE)*. [Online]. 11 (6). pp. 5411. Available from: http://ijece.iaescore.com/index.php/IJECE/article/view/24026.

NTLK (2020). *Natural Language Toolkit*. [Online]. 2020. Available from: https://www.nltk.org/.

Sehgal, S., Sharma, J. & Chaudhary, N. (2020). Generating Image Captions based on Deep Learning and Natural language Processing. In: *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. [Online]. June 2020, IEEE, pp. 165–169. Available from: https://ieeexplore.ieee.org/document/9197977/.

Shi, Y., Feng, D., Cheng, Y. & Biswas, S. (2021). A natural language-inspired multilabel video streaming source identification method based on deep neural networks. *Signal, Image and Video Processing*. [Online]. 15 (6). pp. 1161–1168. Available from: https://link.springer.com/10.1007/s11760-020-01844-8.

Yesugade, T., Kokate, S., Patil, S., Varma, R. & Pawar, S. (2020). Fake News and Deep Fake Detection. *Journal of Critical Reviews*. [Online]. 7 (19). pp. 3876–3884. Available from: http://www.jcreview.com/fulltext/197-1597751056.pdf.