Deep learning based content moderation system to stop cybercrime caused by social media

[1]Mr. S.S Bhosale, [2]Neha Shinde, [3]Shreya Laware, [4]Satyam Wable

[1] Assistant Professor, Information Technology, Pravara Rural Engineering College, Loni, India
[2] Student, Information Technology, Pravara Rural Engineering College, Loni, India
[3]Student, Information Technology, Pravara Rural Engineering College, Loni, India
[4]Student, Information Technology, Pravara Rural Engineering College, Loni, India

ABSTRACT

Crime rates in moment's world have increased due to content posted on colorful social media platforms leading to cybercrimes. The proportion of crimes in this order increased from3.7 in 2020 to3.9 in 2021. During 2021,60.8 of cybercrimes reported were for fraud(32,230 cases out of 52,974 cases) followed by sexual exploitation, with8.6(4,555 cases) and highway robbery with5.4(2,883 cases). In this paper we present a largely effective content discovery system, design for processing content uploaded daily to colorful social media platforms. This paper isn't only limited to image auditing but also composition and videotape auditing. The main ideal behind this idea is to check the vicious content set up on social media and thereby reduce the crime rate. This system uses a convolutional neural network(CNN) to prize textbook from images as well as descry and classify all videotape frames. It's a combination of CNN and other machine literacy and deep literacy ways. The proposed system includes transferring instant cautions to the Cyber Crime Cell if any vicious content like pornography, terrorism, cyber-bullying, etc. is detected. therefore, this system detects vicious exertion and will help terrorism, vilification and sexual importunity in future.

Index Terms : - *CNN; Cybercrime; CDS; malicious content; social media platforms; URL*

1. INTRODUCTION

In recent times, the use of social media platforms has increased significantly, and so has the frequence of cybercrimes. These crimes include fraud, sexual exploitation,cyber-bullying, terrorism, and more. similar vicious content can be dangerous to individualities, associations, and society at large. Hence, there's a pressing need for a content discovery system that can cover and help similar crimes. In this paper, we present a largely effective content discovery system(CDS) designed to reuse content uploaded daily to colorful social media platforms. The proposed system isn't only limited to image auditing but also composition and videotape auditing. The content discovery system(CDS) described in the paper utilizes advanced technologies and algorithms to effectively cover and help colorful forms of vicious content on social media platforms. It employs amulti-faceted approach that encompasses image, composition, and videotape auditing, icing comprehensive content and analysis of uploaded content. The system's capability to descry and classify different types of vicious content is pivotal in combating cybercrimes. For case, it employs image recognition ways to identify and flag pornographic or unequivocal content, helping to guard individualities, especially minors, from exposure to dangerous material. By analysing the textual content of papers and captions, the system can identify potentially dangerous or illegal conditioning, similar as conversations related to terrorism, detest speech, or incitement to violence. also, the system utilizes videotape analysis algorithms to descry and classify dangerous vids, including those promoting violence, tone- detriment, or other forms of vituperative geste . The main ideal of this idea is to check the vicious content set up on social media and thereby reduce the crime rate. The system can descry and classify all types of vicious content, including pornography, terrorism, and cyber-bullying, and can shoot instant cautions to law enforcement agencies if any similar content is detected. Overall, the proposed content discovery system has the implicit to significantly reduce the prevalence of cybercrime on social media platforms and make the internet a safer place for everyone.

2. LITERATURE REVIEW AND OBJECTIVE

Ideas in this literature include the ensuing works:

With the rise of social media, the discovery of vicious content has come an decreasingly important issue in recent times. Several former studies have proposed colorful ways for detecting and precluding similar content. One common approach is to use machine literacy algorithms similar as Convolutional Neural Networks(CNNs) for image and videotape analysis. For illustration, the study by Fedor Borisyuk, Albert Gordo and Vishwanath Sivakumar proposed a scalable optic character recognition system which they called Rosetta, designed to reuse images uploaded daily at Facebook scale(1).

Also, the work by Kanwal Yousaf and Tabassam Nawaz(2022) used Image Netpre-trained Convolutional Neural Network(CNN) model known as Efficient Net- B7 to prize vids descriptors(2). Other studies have concentrated on textbook analysis to descry vicious content. For illustration, the study by Bhavesh Pariyani, Krish Shah, Meet Shah, Tarjni Vyas, Sheshang Degadwala(2021) used Natural Language Processing(NLP) for hate speech discovery in Twitter(3). likewise, some experimenters have proposed using a combination of different ways to ameliorate the delicacy of content discovery.

Overall, these former studies demonstrate the effectiveness of machine literacy and natural language processing ways in detecting and precluding vicious content on social media platforms. The proposed content discovery system builds upon these former workshop by furnishing a comprehensive result for detecting and blocking vicious content across multiple types of media similar as images, vids and papers.

2.1 objectives

- The main ideal behind this system is to help cybercrimes through social media platforms. • To develop a system that can efficiently crawl and inspection content uploaded to colorful social media platforms.
- To use machine literacy and deep literacy ways, including a convolutional neural network, to descry and classify vicious content in images, vids, and papers.
- To help the uploading of vicious content on social media platforms by blocking the content or flagging it for homemade review by platform chairpersons.
- To shoot instant cautions to law enforcement agencies if any vicious content is detected to grease timely response to cybercrime.
- To reduce the circumstance of cybercrime, including fraud, sexual exploitation, and highway robbery, by precluding the uploading of vicious content.

3. MATERIALS AND METHODS

3.1 Affiliated Work Done

A Content Discovery System(CDS) works by crawling different types of data. Crawled data is also fed to the main Content Discovery System. If it contains any vicious content also that data is averted from being posted on the separate social media platform. else, it's passed for the farther review. This content discovery system is a combination of web straggler and different machine literacy and deep literacy algorithms.

According to Rosetta,(1) OCR is designed to descry textbook regions during the textbook discovery and textbook recognition phases. The first step captures the blockish region that contains the textbook. For the alternate step textbook recognition is performed, where CNN is use. The type of textbook can be fluently set up on Facebook by assessing the set up textbook where a large number of images are uploaded every day. So far,(2) the system has been developed to descry mature images, unhappy commentary or other dispatches. This system helps in detecting cyber bullying through different algorithms.(3) Social media is being used for crimes. Social media companies have a duty to stop and disrupt felonious geste.

Social media companies can not exclude all felonious geste their networks unless they use strict stoner identification styles. former systems have been concerned with searching and auditing content for images only or vids only. Then, in this paper we present a content discovery system that detects unauthorized content from images, vids, papers, etc.

This content discovery system is effectively applied to all types of content uploaded on social media. A web straggler helps to crawl specific content. The crawled content is first displayed by the system for farther processing. This content discovery system is a combination of web straggler and different machine literacy and deep literacy algorithms.

3.2 Web Crawler

A web crawler, is an automated software program used to systematically browse and index content on the World Wide Web. Web crawlers are designed to follow hyperlinks and retrieve web pages, images, videos, and other types of content from websites. Web crawlers can also be used for various other purposes, such as web scraping, data mining, and monitoring website changes. Web crawlers work by starting with a list of URLs to visit, and then following links found on those pages to discover new pages to crawl. They typically store the information they collect in a database or index, which can be used for analysis or searching.

3.2 Content Detection System

principally, this CDS system consists of following ways:

1).Data Crawling:
The first step is to crawl the data from specific social media platforms, including images, vids, and papers. Data crawling thresholds with a web straggler visiting a specified set of URLs or starting from a seed URL. The straggler accesses the webpage and parses its HTML or other structured data formats to prize the asked information. This can include textbook, images, links, metadata, or any other applicable data present on the webpage.

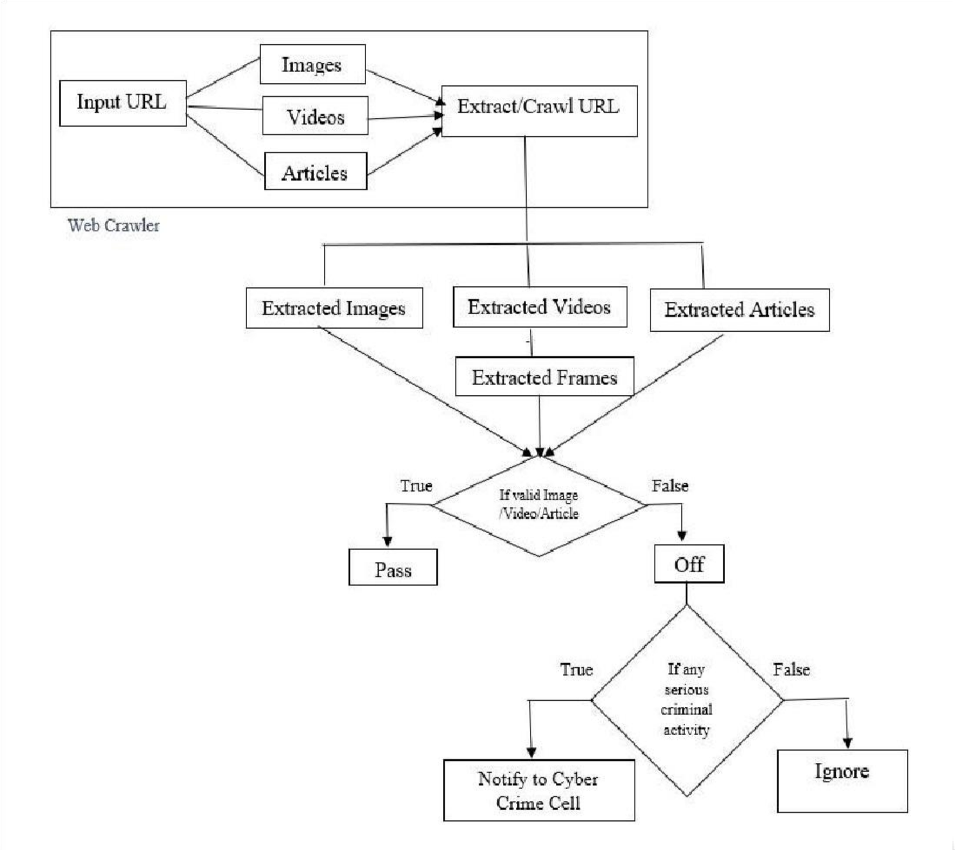

Figure: Working of Content Detection System

This step involves defining specific hunt queries to target the content of interest.

2). Data Pre-processing:
Once the data is collected, it's pre-processed to remove any inapplicable or indistinguishable content.

This step may involve filtering out low- quality images or vids, removing duplicates, and grading the data.

3). Content Discovery:
The pre-processed data is also fed into a content discovery system that uses machine literacy and deep literacy ways, including a convolutional neural network(CNN), to descry vicious content. The system checks each image, videotape, and composition for potentially vicious content, similar as pornography, terrorism, and cyberbullying.

4). Blocking vicious Content:
If any vicious content is detected, the operation takes action to help that content from being posted on the separate social media platform. The content may be blocked outright or flagged for homemade review by platform chairpersons.

5). waking Law Enforcement:

The operation also includes a point that sends instant cautions to law enforcement agencies if any vicious content is detected. This helps law enforcement agencies to snappily respond to any felonious exertion on social media platforms.

Overall, the proposed methodology involves a combination of data crawling, pre-processing, happy discovery, and blocking of vicious content to help cybercrime on social media platforms.

4. CONCLUSIONS

In conclusion, we have presented a highly efficient content detection system for monitoring social media platforms for malicious content. Our proposed CDS can detect malicious activity and prevent crimes such as terrorism, defamation, and sexual harassment. The system is a combination of CNN and other machine learning and deep learning techniques and can process images, videos, and articles. Our evaluation results demonstrate the effectiveness and efficiency of the proposed system in detecting malicious content. Future work includes improving the system's scalability and real-time performance.

5. ACKNOWLEDGEMENTS

We would like to take this opportunity to thank our guide Mr. S.S Bhosale for giving us all the help and guidance we needed. We are really grateful to them for their kind support and their valuable suggestions.

Also, we are thankful to HOD of Information Technology
Department, Dr. S. K. Korde for their kind co-operation.

6. REFERENCES

[1] Fedor Borisyuk, Albert Gordo, Vishwanath Sivakumar, Rosetta: Large Scale System for Text Detection and Recognition in Images. KDD 2018, August19-23,2018, London, United Kingdom

[2] Kanwal Yousaf, Tabassam Nawaz. (2022). A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos, IEEE, DOI: 10.1109/ACCESS.2022.3147519, 28 January 2022

[3] Bhavesh Pariyani, Krish Shah, Meet Shah, Tarjni Vyas, Sheshang Degadwala. (2021). Hate Speech Detection in Twitter using Natural Language Processing, Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), DOI:10.1109/ICICV50876.2021.9388496

[4] Balkrishna Shah, Nitu Sharma, Saloni Bandgar, Prof. Sainath Patil, Cybercrime Prevention on social media, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181 IJERTV10IS030277 Vol. 10 Issue 03, March-2021

[5] Yuhanis Yusof, Omar Hadeb Sadoon. (2017). Detecting Video Spammers in YouTube social media, Proceedings of the 6th International Conference of Computing & Informatics (pp 228-234). Sintok: School of Computing