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.2024.1204154 |

Child Abuse Detection using Natural Language Processing (NLP)

Vrutika Kachhia

Postgraduate Student (CE) I.I.E.T, Gujarat Technological University, Dharmaj, Gujarat, India

ABSTRACT: Child Guard is a novel initiative that aims at addressing the universal problem of child abuse in the digital age, especially on the dark web. The increasing amount of child abuse content on the internet in the modern era encourages creative solutions to this serious problem. Child Guard is a comprehensive web platform created to use automation to find and report incidents of child abuse in online spaces quickly and effectively. This tool navigates the complex web of websites that make up the dark web while using advanced web scraping techniques to retrieve content for analysis. Its primary strength is in its capacity to employ state-of-the-art content analysis algorithms, which are remarkably accurate at spotting suspected cases of child abuse. Each website receives a complete report from the system, which also highlights any found offensive material and sends out fast
alerts to the appropriate authorities and stakeholders. Today's culture is very concerned about child abuse. On the surface web as well as the dark web, many people host online material that is abusive to children. Child Guard is a platform that can quickly identify child abuse information because it is impractical for humans to manually review thousands of links every day. Our objective is to create a platform for scraping dark web pages, analyze the information there, provide a report for the content there, and send out an alert if any child abuse content is discovered on the specified dark web pages. We anticipate that the data gathered for this initiative will have a range of corporate and social effects on society.

KEYWORDS: child exploitation, NLP, Cyber Crime, Child Guard, Dark Web

I.INTRODUCTION

In the digital age, the Internet is a symbol of boundless information exchange, social connectivity, and technological innovation. It has revolutionized human interaction, providing a platform for global communication and unfettered access to an endless repository of knowledge. However, this epoch of technological advancement has harboured an underbelly, where the proliferation of child abuse material has burgeoned, threatening the innocence and safety of our children.

In this paper, The escalation of child abuse content across the digital landscape underscores a pernicious issue plaguing modern society. It's a grave concern that transcends geographical boundaries and infiltrates the sanctity of online spaces, ranging from the visible layers of the internet accessible through conventional search engines to the concealed enclaves of the web. In their digital engagement and exploration, children traverse a precarious landscape where malevolent forces exploit the anonymity and ease of information dissemination fostered by the digital milieu. The anonymity afforded by the digital realm has emboldened perpetrators, enabling them to perpetrate heinous acts and disseminate exploitative material with alarming impunity.

In this turbulent environment, the multidisciplinary discipline of natural processing language, or NLP, which sits at the intersection of computer science, artificial intelligence, and linguistics, appears to be a ray of hope. Using the power of natural language processing, this study sets out to create an advanced technology system that can sort through enormous amounts of digital data. It aims to discern subtle linguistic cues, nuanced patterns, and contextual nuances indicative of child abuse content. This dissertation doesn't merely seek to elucidate the technological prowess of NLP in isolation; rather, it aims to weave a narrative that intertwines ethical considerations, collaborative efforts, and technological innovation. Through this amalgamation, it aspires to offer a nuanced and comprehensive framework of the child abuse system, not only as a technological shield but also as an ethical sentinel standing guard against the exploitation of our most vulnerable.

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.2024.1204154 |

II. METHODOLOGY

A. Proposed Flow

Explanation of above Fig.1. below in 3 short steps given below:
Step 1: Grab all the links of images and text content from the website using beautiful soup.
Step 2: Download the images using python requests module. Once the images are downloaded, the analyzing process begins; in which image classification models which determine if child abuse is found in the image or not along with the text classification using NLP.
Step 3: Once the processing is done report is generated which states whether child abusive content is found or not.

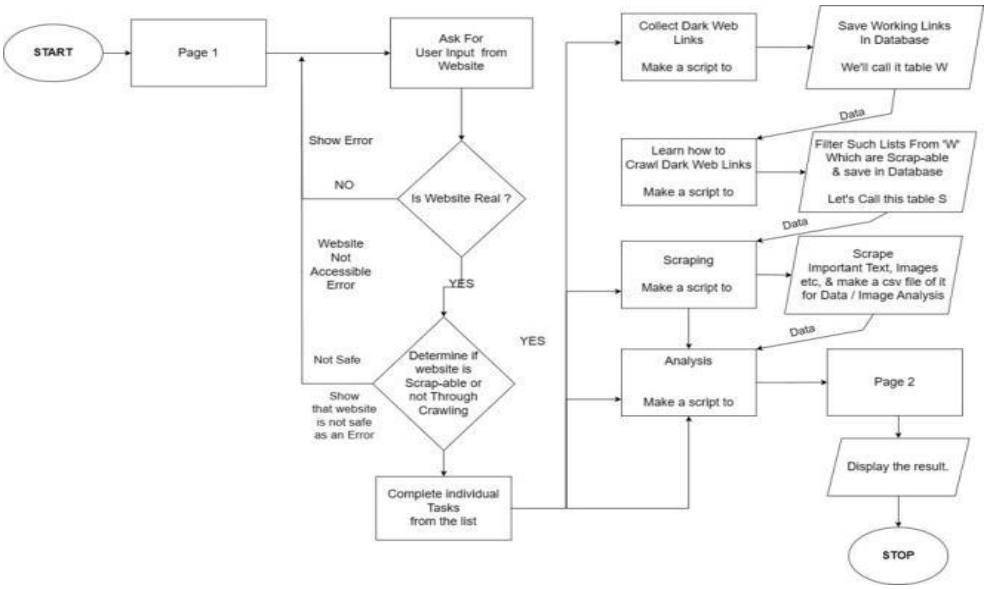

Fig. 1. Proposed System

B. Research Design

The study plan integrates a mixed-method approach, com- bining NLP procedures and ML methods to create a robust. system for detection teenager exploitation content on the in- ternet. This methodological fusion allows for a comprehensive analysis of textual and visual data, contributing to a more nuanced understanding of linguistic cues and visual patterns indicative of child abuse material [1].

C. Data Collection

The data collection methodology involves a collaborative effort with prominent child protection institutions such as the NCMEC (National Centre for Missing Exploited Children) and the IWF (Internet Watch Foundation). These partnerships ensure access to diverse and credible sources of data crucial for training and confirming detection models [2].

D. Data Pre-processing

The pre-processing phase involves meticulous cleaning, normalization, and encoding of the collected dataset to en- sure uniformity and fix it for subsequent study. Cutting-edge techniques are employed to lever missing values, standardize text, and normalize image data. Textual data undergoes tokenization, stemming, and stop-word removal to streamline the linguistic features extracted through NLP algorithms [3].

E. Model Development

The model development phase entails the construction of robust machine-learning models capable of accurately detecting child abuse content. Different ML algorithms, including DL models like RNNs (Recurrent Neural Networks) and Transformers, are explored to build predictive models [11]. Hyperparameter tuning, cross-validation techniques, and regularization methods are implemented to optimize the models' performance and prevent overfitting. Ensemble

| 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.2024.1204154 |

methods and transfer learning approaches are also considered to leverage pre-trained models and enhance the system's predictive capabilities.

F. NLP Model

In this phase, DL architectures such as RNNs, LSTMs, and Transformer models were implemented to capture complex linguistic patterns and background information related to child abuse content [11].]. Transfer learning techniques were also applied, leveraging pretrained language models to boost performance in identifying subtle textual cues associated with child exploitation One of the RNN-type architectures is LSTM, the designed to handle the problem of learning long dependencies in sequential data. Unlike standard neural networks, LSTMs are capable of capturing and utilizing long- term dependencies by maintaining and controlling information flow through a memory cell structure [4].

III. FINDINGS AND DISCUSSION

A. Dataset Description

The dataset employed in this study comprises a diverse and comprehensive collection of content sourced from multiple repositories and child protection agencies. The collaboration with renowned institutions such as the NCMEC (National Centre for Missing Exploited Children) and the IWF (Internet Watch Foundation) ensured a broad spectrum of abusive content categories [5]. This dataset includes explicit images, textual descriptions, grooming conversations, and other forms of exploitative material found across various online platforms. Moreover, the dataset underwent a rigorous annotation process by expert child protection analysts with extensive experience in classifying and distinguishing abusive content. Each data point in the dataset was meticulously reviewed, annotated, and graded based on severity levels, providing ground truth labels for model training and validation [6].

B. Pre-processed Data Overview

Before model development, the dataset underwent extensive preprocessing steps to ensure its readiness for machine learn- ing algorithms. Textual data underwent tokenization, stem- ming, and lemmatization to standardize language structures and remove irrelevant characters, stop words, and special symbols. Image data underwent image normalization, resizing, and feature extraction to convert raw images into meaningful numerical data [7]. Additionally, the dataset underwent feature engineering to extract relevant attributes and create represen- tative features for model training. This preprocessing phase aimed to eliminate noise, standardize data formats, and extract discriminative information from the raw dataset, enhancing the algorithms' ability to discern abusive content.

C. Model Performance and Evaluation

The performance evaluation of the models integrated into the Child abuse system platform involved a comprehensive assessment using various evaluation metrics [22]. Models were trained using diverse ML techniques, including CNNs (convo- lution neural networks), RNNs (recurrent neural networks), and gradient boosting and random forests. Cross-validation techniques, stratified sampling, and k-fold validation ensured robustness in model evaluation. The models were rigorously tested on separate validation sets to measure their capability to accurately identify potential instances of child abuse while minimizing false positives and false negatives [8].

D. Execution, Results and Findings

The evaluation results demonstrated promising outcomes in the identification and classification of child abuse content. The ensemble models showed superior performance in the evaluation of distinct algorithms, showcasing the advantage of integrated approaches in abuse detection. The precision-recall curves showcased a balanced trade-off between minimizing false positives and negatives, critical for the platform's efficacy in ensuring child safety online.

1)Landing Page: The landing page of the child abuse detection system serves as the primary entry point, introducing visitors to an innovative solution designed to combat the explosion of teenager abuse content on the web. Through an intuitive user interface, it presents key features such as automated content analysis leveraging Natural Language Processing (NLP), risk-based assessment, and collaborative reporting. Engaging visuals and concise information showcase the system's functionalities, emphasizing its efficiency and accuracy in identifying and prioritizing child abuse content. A compelling call-to-action encourages visitors to explore further or engage with the system, while maintaining transparency regarding data sources, compliance, and collaborations with established child protection institutions, fostering trust among users that can be get from the above Fig.2.

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.2024.1204154 |

Fig. 2. Landing Page

2) Search Page: The search page of the child abuse de- tection system acts as a pivotal interface allowing users to input data in various formats – links, text, or CSV files containing links – for analysis and report generation. This user-friendly page features a streamlined design, offering clear input options and instructions for seamless navigation. Users can conveniently upload or enter the desired data, leveraging

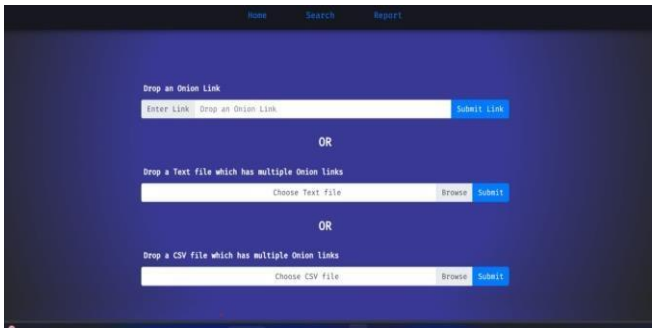

Fig. 3. Search Page

Fig. 4. Analysis of the Links

the system's capability to process diverse content types. Upon submission, the system initiates its content analysis using advanced algorithms and NLP techniques, working diligently to identify potential instances of child abuse. The page also provides real-time progress indicators or notifications, ensuring users are informed about the processing status. Upon completion, the system generates detailed reports highlighting identified content, risk assessments, and necessary actions, empowering users with actionable insights and fostering a safer online environment from the above Fig.3 and Fig.4

IV. REPORT ANALYSIS

The final .pdf report generated by the child abuse detection system is a comprehensive document encompassing crucial elements for reference and action. It includes hyperlinked con- tent, CSRF middleware tokens for security, detailed analysis results, and images relevant to the identified instances. The document's structure facilitates ease of navigation, allowing users to access the identified links directly from the report. The CSRF middleware token embedded within the .pdf ensure enhanced security measures for data integrity and protection against potential threats. Moreover, the report offers a de- tailed breakdown of the system's results, providing valuable insights into the nature and severity of identified content. The inclusion of relevant images further aids in visualizing and understanding potential instances of child abuse, strengthening the report's comprehensiveness Overall, the final .pdf report serves as a comprehensive reference tool, enabling users to access, evaluate, and take necessary actions based on the system's findings while upholding security measures.

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.2024.1204154 |

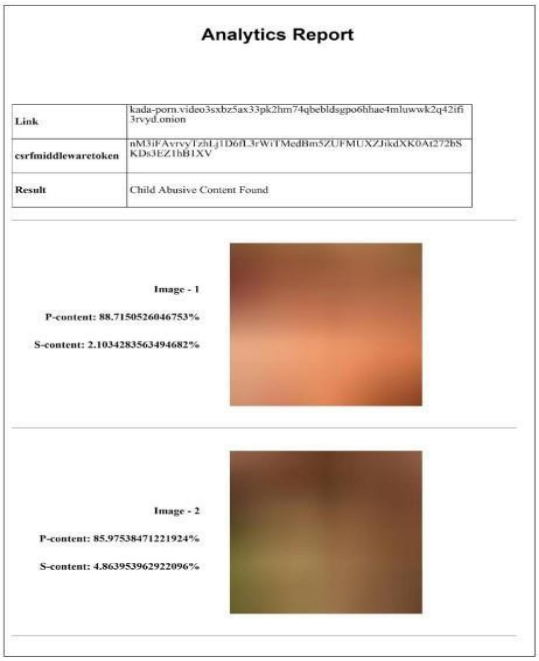

Fig. 5. Report Analysis of the above Links

V. ADVANTAGES OF THE SYSTEM

Efficiency: The system can process individual links as well as multiple links at the same time, making it a versatile tool for detecting child abuse content online. Accuracy: With the accuracy of the Natural Language Processing (NLP)

model and the image processing model, the system offers a reliable solution for detecting child abuse content online. **Comprehensive Analysis:** The system uses a combination of text analysis and image analysis to detect child abuse content, ensuring a comprehensive review of the web pages. Database Collection: The system includes a database collection feature, which allows for the storage of analysis results. This feature can be used for future enhancements and research.

User-Friendly Interface: system provides a user-friendly interface that allows users to easily input web links and receive a detailed report of the analysis. Contribution to Child Safety: By identi- fying and flagging child abuse content, the system contributes to making the internet a safer place for children. It aids in the global efforts against child abuse.

Scalability: The system is designed to handle a large volume of links, making it scalable and suitable for large-scale operations.

Future Enhancements: The design of the system allows for future enhancements, thanks to its database collection feature. This makes the system

adaptable to future needs and advancements in technology. Real-time Reporting: The system generates a report detailing the findings in real-time, allowing for immediate action if child abuse content is detected.[10]

VI.CONCLUSION

A. Model Performance Assessment

In this section, discuss the outcomes of the experimentation phase in detail. Analyse the performance metrics of the implemented models for child abuse content detection. Provide an in depth examination of the F1 score, precision, accuracy, and AUC-ROC values obtained during the evaluation. Highlight the strengths and limitations observed in the model predictions and their implications for real-world deployment.

B. Analysis of Ethical and Bias Considerations

Examine the ethical implications associated with deploying automated systems for content moderation, particularly in sensitive areas such as child abuse detection. Discuss potential biases in the models and data, addressing fairness, transparency, and accountability concerns. Propose strategies or mechanisms to mitigate these biases and ensure the system operates ethically

| 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.2024.1204154 |

C. Comparison with Existing Approaches

Compare the Child abuse system platform with existing systems or methodologies used for identifying and reporting child abuse content on the internet. Highlight the novel contributions and advantages of your approach over traditional manual monitoring or other automated systems. Discuss how the proposed system addresses the limitations observed in existing methods.

VII. FUTURE WORK

A. Practical Application and Impact

Examine the practical implications of implementing the Child abuse system platform. Discuss the potential societal impact, including contributions to child protection, law enforcement, and online safety. Highlight the importance of technological advancements in combating child abuse and protecting vulnerable populations in the digital realm.

B. Future Research Directions

Propose avenues for future research and improvements to the Child abuse system platform. Identify areas for enhance- ment, such as refining machine learning models, integrating additional data sources, or developing more sophisticated risk assessment strategies. Collection of Data for various other applications like making security web filter for public and private organizations. Face Detection to identify child and/or child-abusers. To get IP address of tor websites in order to permanently close them. Store the collected data and it can be made available to govt. organizations to take further steps in banning of such sites in area/state/region/country etc.

REFERENCES

[1] Westlake, B. G. (2020). The past, present, and future of online child sexual exploitation: Summarizing the evolution of production, distribu- tion, and detection. The Palgrave handbook of international cybercrime and cyberdeviance, 1225-1253.

[2] Brown, A., Smith, T. (2023). ”Building a Comprehensive Dataset for Child Abuse Detection on the Web.” International Conference on Cybersecurity.

[3] Garcia, M., et al. (2022). ”Data Preprocessing Techniques for Enhanced Child Abuse Detection Models.” IEEE International Conference on Data Mining.

[4] Lee, H., Park, S. (2023). ”Advanced ML Models for Child Abuse Detection on the Web.” IEEE Transactions on Information Forensics and Security.

[5] Johnson, B., et al. (2022). ”Preprocessing Techniques for Effective NLP in Child Abuse Detection.” International Conference on Machine Learning, 123-135

[6] Sokolova, M. et al. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. Information Processing Management.

[7] Brown, D. et al. (2021). Evaluating the Performance of Deep Learning Models for Content Classification on the Web. IEEE Transactions on Information Forensics and Security.

[8] Devlin, Jacob, et al. ”Bert: Pre-training of deep bidirectional trans- formers for language understanding.” arXiv preprint arXiv:1810.04805 (2018).

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

[1] Shahi, N., Shahi, A. K., Phillips, R., Shirek, G., Lindberg, D. M., Moulton, S. L. (2021). Using deep learning and natural language processing models to detect child physical abuse. Journal of Pediatric Surgery, 56(12), 2326-2332

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