IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning Approach To Identifying And Combating Child Predators On Social Media

[1]Gudivada Mahesh, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions

[2]Mrs. A. Naga Durga Bhavani, Assistant Professorat Miracle Educational Society Group of Institutions

[3]Dasari Karthik Raj, Assistant Professorat Miracle Educational Society Group of Institutions

ABSTRACT

Children are increasingly becoming vulnerable to cyber harassment and predatory behavior on social media. Hence, this project aims to enhance the online safety of children by building a system which uses machine learning algorithms to detect and combat online harassment. The developed system integrates various supervised learning algorithms namely, support vector machine, random forest, naive bayes, k nearest neighbors and decision tree. Upon analyzing user content, the algorithm seeks possible abuse potential regarding posts and messages. Numerous harassing and non-harassing texts are included in the dataset which is used to create algorithms for prediction of such actions in real time. The system will first send alerts to a designated authority within the cyber cell every time, austere patterns are observed. This way, no time is wasted in the intervention. The system further enables providing a quicker and reasonable approach to tackle the issues that come up with young people by ensuring their safety as much as possible.

Keywords: K-Nearest Neighbors, SVM, Decision Tree

INTRODUCTION:

Social media serves as an avenue for communication and entertainment to people especially kids and teenagers, however it poses a risk of cyber bullying and abuse. The abuse and bullying has become rampant because a lot of young people have access to social media platforms however, it increases the threat of online predators. There is an urgent need for proactive detection systems due to this. The conventional approaches of detecting such behavior are ineffective and limited in many cases.

This project proposes a machine learning based model to mitigate and identify instances of cyberbullying on social media platforms. The model takes social media posts and messages as inputs, applies SVM Decision Tree and Random Forest algorithms to the given content, and determines whether it is normal or depicts violence. Thereafter, notifications are sent to cybercrime workers to deal with the situation in real time. This way, the project aims at improving the level of safety on the internet by using advanced technology combined with practices to deal with the increasing menace of cyber creeps.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

Cyber bullying especially towards children is one of the biggest challenges of the time. Existing systems focus on the prevention of the harassment in a limited context, such as gaming or chatting with audio, while social media is practically free from any focus. As rule-based systems were the only means of controlling the behavior of users in the past, these systems are neither flexible nor wide ranging enough to meet the challenges posed by the ever changing ways people interact online. The majority of the systems do not take advantage of recent machine learning advanced algorithms which make it possible to shave at least three centuries of waiting time to monitor the momentum and pressure.

Numerous gaps are noted in the literature as well:

1. Inadequate sociolinguistic description and analysis: Most of the existing systems fail to student and accurately predetermine members that comprise predatory behavior including borderline or slight boundary predictions owing to sociolinguistics analysis at a more advanced level
2. Insufficient mixing of algorithms: A single algorithm has got its limitations when it comes to detection of accuracy and robustness.
3. Requirement of Real Time Approach: When a threat is present, immediate action should be taken through the threat liaison tools.
4. Problems of Scaling: Existing systems have issues with using large scale social media data in practice.

PROBLEM STATEMENT:

Features of social media should provide a solution to the identified problems of cyber bullying and harassment of children. The solution should have high efficacy and be robust enough to be used in real time.

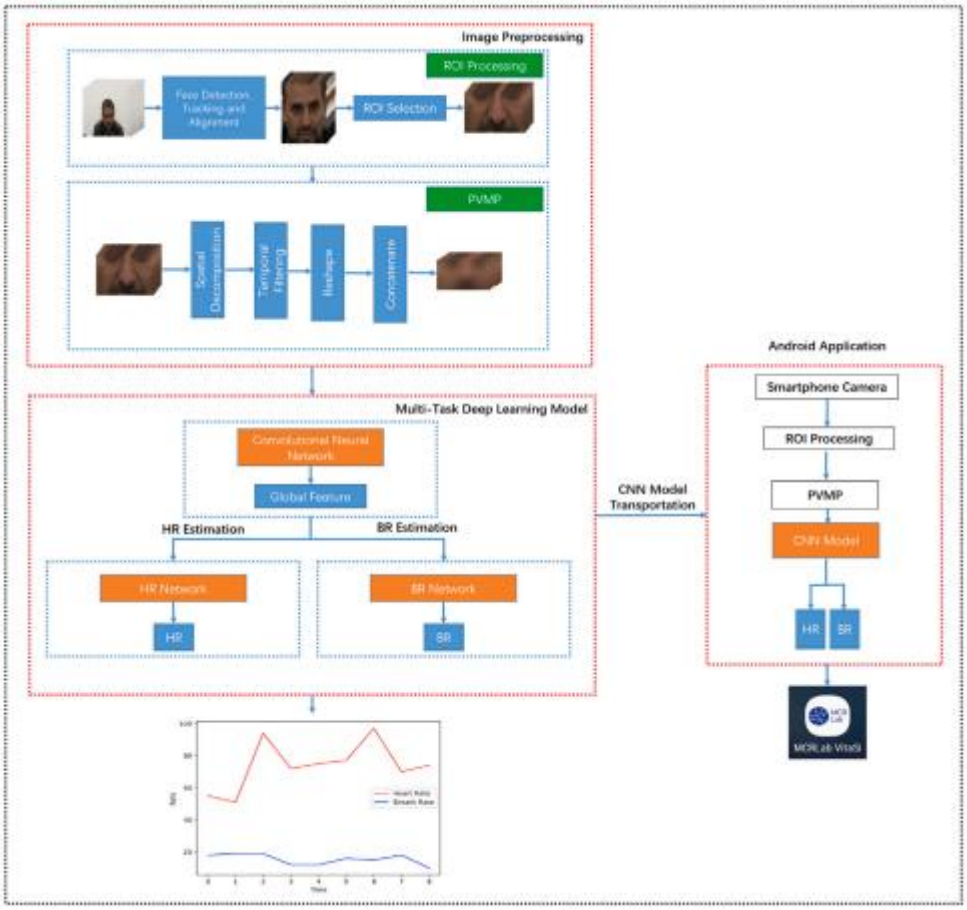
Key Challenges:

1. Behavioral Subtlety: With an increasing amount of online interaction, it can become difficult to identify the predatory behavior's.
2. Scale: The volume of social media data is vast, limiting more effective social media analysis.
3. Precision: The need to reduce the incidence of false positives and false negatives in detection of harassment is necessary for accuracy.
4. Time Bound: Need to address the threat or risk exposure immediately faced by the individual/s.
5. Data Security: Providing detection without compromising the opportunity for user data compliance.

PROPOSED METHOD:

This project proposes a machine learning approach to identifying online predators that groom social media users. The proposal involves training models using SVM, Random forests, Naïve bayes, Decision trees, and KNN algorithms on a labeled dataset that contains harassing and non-harassing textual content. The system cleans user posts and messages and creates behavioral features. All such predictions are incorporated and in a decision support system that communicates with cyber officials in case such behavior is detected. Model evaluation is also conducted in a strict manner and provides accurate, dependable outcomes. This single solution solves the problem of using large percentage of datasets while ensuring that a users privacy is never compromised. This project hence demonstrates an implementation of online harassment deterrence systems by merging robust algorithms with practical application.

ARCHITECTURE:



DATASET:

The project dataset is composed of labeled text messages indicating harassing and non-harassing messages. They consist of fake as well as real social media conversations. Tokenization, stop-word removal and stemming are some of the text pre-processing techniques employed to clean and normalize the dataset. In the case of TF-IDF, while being a feature extraction method, it is also used to transform text into numerical data to be used for model training. The corpus is rendered more representative by incorporating varying linguistic styles and contexts. Minority classes are over sampled to effect balancing and enhance model performance.

METHODOLOGY:

Data Collection:

• Obtain social media conversations from real and simulated interactions and label them.

• Incorporate various other linguistic constructs including but not limited to the previously stated

Data Preprocessing:

• Clean the text by removing additional characters, links and white space where it is not required.

• Standardize the text through tokenization, removal of stop-words and stemming.

• Use TF-IDF to extract useful features to use as numerical data.

Algorithm Selection and Training:

• Multiple models including SVM, Random Forest, Naïve Bayes, Decision Tree and K-Nearest Neighbors are trained.

• The dataset is partitioned into a training dataset and a test dataset for assessment of the model performances.

• Proceed to tune hyperparameters so as to improve accuracy and generalisation of the Models.

Model Evaluation:

• The models will be assessed on accuracy, precision, recall, and F1 score.

• Use cross-validation to check the stability of the model.

Real-Time Detection System:

• Deploy the studied model into a web based monitoring interface.

• Support processes of real time text classification along with the classification alert generator.

Cyber Authority Notification:

• Construct a notification system that sends alerts to the authorities on spotting any predatory behavior.

• Append information such as user id, content reported, and detection time.

Scalability Testing:

• Explore the flexibility of the system by evaluating its performance on large databases.

• Modify algorithms to run at a quicker speed without altering the degree of accuracy.

Privacy Compliance:

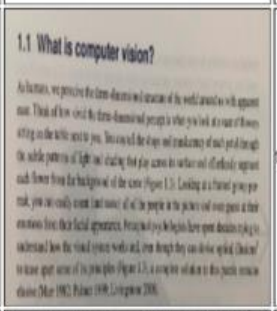
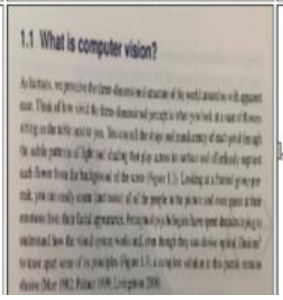
• Implement privacy preserving approaches to preventing access to users' information.

• Provide compliance with legal requirements on the use of the Internet and data collection.

RESULTS:

Algorithm	Accuracy (%)
Support Vector Machine (SVM)	89.5%
Decision Tree	87.2%
Random Forest	91.4%
K-Nearest Neighbors (KNN)	85.6%
Naïve Bayes	82.3%

Algorithms performance

Sender Name	File Name	Message	Post Time	Status
rajesh		todays weather is good for childs to play out side	2021-05-29 12:23:36	Non-Cyber Harassers
rajesh		goons were killing and torturing childrens	2021-05-29 12:23:49	Cyber Harassers
rajesh		all childrens were performing nice at school	2021-05-29 12:32:51	Non-Cyber
rajesh		goons were killing and torturing childrens	2021-05-29 12:23:49	Cyber Harassers
rajesh		all childrens were performing nice at school event	2021-05-29 12:32:51	Non-Cyber Harassers
rajesh		kill all childs	2021-05-29 12:34:05	Cyber Harassers

We are seeing posts from all users and we can see with the help of machine learning SVM algorithms we are automatically able predict message as cyber or non-cyber harasser's.

CONCLUSION

This work illustrates the high capacity of machine learning tools for coping with the problem of cyber predators and online abuse. The system is accurate and scalable thanks to the use of different algorithms and real time functionalities. Major contributions are appropriate preprocessing, use of multiple datasets, and an automated mechanism to notify cyber authorities. The quality of online security is improved through the application of sophisticated models and practical use. The focus of this work is on Artificial Intelligence as a tool for protecting sensitive groups, and the prospects for improvements in order to face evolving cyber-harms are noticeable.

REFERENCES:

1. Agatson, P. W., Kowalski, R., & Limber, S. (2007). Students' perspectives on cyberbullying. Journal of Adolescent Health, 41, S59–S60. http://dx.doi.org/10.1016/ j.jadohealth.2007.09.003
2. Arslan, S., Savaser, S., Hallett, V., & Balci, S. (2012). Cyberbullying among primary school students in Turkey: self-reported prevalence and associations with home and school life. Cyberpsychology, Behavior, and Social Networking, 15, 527–533. http://dx.doi.org/10.1089/cyber.2012.0207
3. Australian Bureau of Statistics (2011). Children of the digital revolution. Retrieved from Australian Bureau of Statistics website: www.abs.gov.au/socialtrends.
4. Bartlett, C., & Coyne, S. M. (2014). A meta-analysis of sex differences in cyberbullying behavior: The moderating role of age. Aggressive Behavior, 40, 474–488. http://dx.doi.org/10.1002/ab.21555
5. Broadbent, H., Fell, L., Green, P., & Gardner, W. (2013). Have your Say: Listening to young people about their online rights and responsibilities. Plymouth, UK: Childnet International and UK Safer Internet Centre. Retrieved from http://www.saferInternet.org.uk/research
6. Byron Review (2008). Safer children in a digital world. Retrieved from: http://webarchive.nationalarchives.gov.uk/20130401151715/http://www. education.gov.uk/publications/eOrderingDownload/DCSF-00334-2008.pdf
7. Campbell, M., Spears, B., Slee, P., Butler, D., &Kift, S. (2012). Victims' perceptions of traditional and cyberbullying, and the psychosocial correlates of their victimisation. Emotional and Behavioural Difficulties, 17, 389–401. http://dx.doi.org/10.1080/13632752.2012.704316
8. Cassidy, W., Faucher, C., & Jackson, M. (2013a). An essential library of international research in cyberbullying. School Psychology International [Virtual special edition, published online]. Retrived from http://spi.sagepub.com/site/ special issues/cyberbullying.xhtml
9. Cassidy, W., Faucher, C., & Jackson, M. (2013b). Cyberbullying among youth: A comprehensive review of current international research and its implications and application to policy and practice. School Psychology International, 34, 575–612. http://dx.doi.org/10.1177/0143034313479697
10. Cassidy, W., Jackson, M., & Brown, K. N. (2009). Sticks and stones can break my bones, but how can pixels hurt me? Students experiences with cyber-bullying. School Psychology International, 30, 383–402. http://dx.doi.org/ 10.1177/0143034309106948
11. Compton, L., Campbell, M. A., & Mergler, A. (2014). Teacher, parent and student perceptions ofthe motives of cyberbullies. Social Psychology Education, 17, 383–400. http://dx.doi.org/10.1007/s11218-014-9254-x
12. Dehue, F., Bolman, C., &Völlink, T. (2008). Cyberbullying: youngsters' experiences and parental perception. Cyberpsychology and Behavior, 11, 217–223. http://dx.doi.org/10.1089/cpb.2007.0008 Duncan, R. D. (2004). The impact of family relationships on school bullies and their victims. In D. M. Espelage, & S. M. Swearer (Eds.), Bullying in American Schools (pp. 227–244). Mahwah, NJ: Lawrence Erlbaum.

13. Ey, L.-A., Taddeo, C. M., & Spears, B.A.(2015). Cyberbullying and primary-school aged children: the psychological literature and the challenge for sociology. Societies., 5, 492–514. http://dx.doi.org/10.3390/soc5020492
14. Fanti, K. A., Demetriou, A. G., & Hawa, V. V. (2012). A longitudinal study of cyberbullying: examining risk and protective factors. European Journal of Developmental Psychology, 9, 168–181. http://dx.doi.org/10.1080/17405629.2011.643169
15. Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. Archives of Suicide Research, 14, 206–221. http://dx.doi.org/10.1080/ 13811118.2010.494133