iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40832

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Online Child Predator Detection Using ML (Research Paper)

Prof. Swati Gade[1], Utkarsha Bhadre[2], Tabbasum Shaikh[3], Mahesh Bangar[4], Sarika Waghmode[5]
[1]*Professor, Computer Department, PDEA's College of Engineering, Manjari, Pune*
[2, 3, 4, 5]*Student, Computer Department, PDEA's College of Engineering, Manjari, Pune*

Abstract: ***Proficient analysts should all the more completely comprehend the risks of online sexual sales and manners by which to shield youth from sexual stalkers who utilize the Internet. Albeit the Internet has numerous positive angles, quite possibly the most malicious aspect is its expected use for online sexual predation. The Internet addresses a medium that permits sexual stalkers admittance to innumerable kids in a moderately mysterious climate. The primary goal of our task is to distinguish kid hunter base on remarks and post of web-based media account and send hunter record to digital cell admin. A late public review demonstrated that around one out of five youth are requested for sex over the Internet yearly (Finkelhor, Mitchell, and Wolak, 2000; Mitchell, Finkelhor, and Wolak, 2001). This task report presents our present advancement to empower the formation of the framework. Thus, with the created framework, youngster hunter accounts recognition any report to administrator for additional activity.***

Index Terms: ML, dataset, Training Module, Predator.

I. INTRODUCTION

The online exposure of children to pedophiles is one of the fastest growing issues on social media. As of March 2014, the National Society for the Prevention of Cruelty to Children (NSPCC), reported that i) 12% of 11-16 year olds in the UK have received unwanted sexual messages; and ii) 8% of 11-16 year olds in the UK have received requests to send or respond to a sexual message [16]. The detection of children cyber-sexual-offenders is therefore a critical issue which needs to be addressed. Children in their teens have started to use social media as their main means of communication [20].Moreover a recent study of cognition, adolescents and mobile phones(SCAMP) has revealed that 70% of 11-12 year olds in the UK now own a mobile phone rising to 90% by age 14 [28]. While social media outlets (e.g., chat-rooms, images and videosharing sites, microblogs) serve as contact points for paedophile (predators)to potentially exploit children (victims),the automatic detection of children abuse on the Web is still anopen question. A common attack from paedophiles is the so-called online child grooming, where adults engage with minors via social media outlets to eventually exchange sexually explicit content. Such grooming consists of building a trust-relationship with a minor, which finally leads into convincing a child to meet themin person [19]

Previous research on detecting cyber paedophilia online, including the efforts of the first international sexual predator identification competition (PAN'12)[11], has focused on the automatic identification of predators in chat-room logs. However little has beendone on understanding predators behaviour patterns at the various stages of online child grooming, which include Deceptive Trust Development, Grooming, and Seeking for Physical Approach (Section 2). Characterising such stages is a critical issue since most of the sexually abused children have been driven to voluntarily agreeto physically approach the predator [36]. This suggests that understanding the different strategies apredator uses tomanipulate children behaviour could help in educatingchildren on how to react when expose to such situations. Moreover the early detection of such stages could facilitate the detection of malicious conversations on the Web. We believe that a deeper characterisation of predator behaviour patterns in such stages could aid in the development of more robust surveil-lance systems which could potentially reduce the number of abused children. This pa-per advances the state of the art on predator detection by proposing a more fine-grained characterisation of predators' behaviour in each of the online child grooming stages[21]. The main contributions of this paper can be summarised as follows:(1) We propose an approach to automatically identify grooming stages in an online conversation based on multiple features: i) lexical; ii) syntactical; iii) sentiment;iv) content; v)psycho-linguistic; and vi) discourse patterns.(2) We generate classification models for each stage, using single and multiple features. Our findings demonstrate that the use of Label discourse pattern features alone canachieve on average a gain in precision (P) of 4.63% over lexical features. While the use of combined features in classifiers consistently boost performance in P with again of 7.6% in all grooming stages.(3). We present a feature analysis to identify the most discriminative features that characterise each online grooming stage. The rest of the paper is organised as follows: Section 2 introduces Olson's theory of luring communication which characterises predator's child grooming stages. Sec-tion3 presents related work regarding detection of online predator-victim conversations as well as previous work in online child grooming. Section 4 presents the set of features selected to characterised the language used by predators.

©IJRASET: All Rights are Reserved

Section 5 introduces our methodology for characterising and identifying grooming stages. Results and discussion are presented in sections 6 and 7. Conclusions are presented in Section 8

II. LITERATURE SURVEY

Grooming is the process where a predator builds trust with a child with the intention of sexual abuse [1]. This usually includes lowering the child's inhibitions to sexual content. The word cyber is normally used to describe something that involves computers and networks. Thus, cyber grooming is when a predator is grooming a child over the Internet. Sexual predators come in contact with potential victims on social media and chat rooms.

Various researchers have focused on detecting grooming chat conversations using author profiling techniques related to pattern recognition and textual analysis. One of the problems with analysis of chat data is that the manner in which people write in a chat is very informal, contains grammatical and typographical errors, and includes emoticons and other non-relevant information. Besides that, each chat message in itself is only very short and there is interaction between the chatters in a chatroom. Pendar [2] used automatic text categorization techniques to identify sexual predators, using Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN) as classifiers. Pendar used 701 different predatory conversations from the Perverted Justice (PJ) website (see Section III for a description), and classified chatters as predator or victim based on all the messages of a chatter in a full conversation. McGee et al. [3] took an approach where single messages were labeled as predatory or benign and they used 33 PJ conversations to test the performance of their system. They managed to correctly identify message on average 68.11% of the time. Inches and Crestani [4] gave an overview of the task and results of the PAN-2012 competition (see Section III for more details). The main goal of this competition was to identify predators in Authorized licensed use limited to: University of Canberra. Downloaded on June 07,2020 at 23:19:45 UTC from IEEE Xplore. Restrictions apply. a chat. The best results were obtained by Villatoro-Telloet al. [5] who used Neural Networks (NN) and SVM in a 2- stage approach. Their optimal result was achieved when using NN with a binary weighting scheme in both stages. They achieved a recall of 0.7874 and a precision of 0.9804. The submission of Eriksson and Karlgren [6] reached 5th place in the PAN-2012 competition, but this submission had the highest recall (0.8937) of all submission, and their precision was 0.8566. In their approach, they model the characteristics of the conversation, in terms of the vocabulary of a particular chatter compared to other chatters, the length of a conversation and the number of participants in a conversation. In 2016, Ebrahimi et al. [7] were the first to apply Convolutional Neural Networks (CNN) for predator detection. They also used the PAN-2012 data set for their analysis and their best results were a precision of 0.9157 and a recall of 0.7241. These results would have earned them a third place in the PAN2012 competition. Some research was performed on full PJ conversations. Pandey et al.

III. PROPOSED METHOD

A. *System Design*

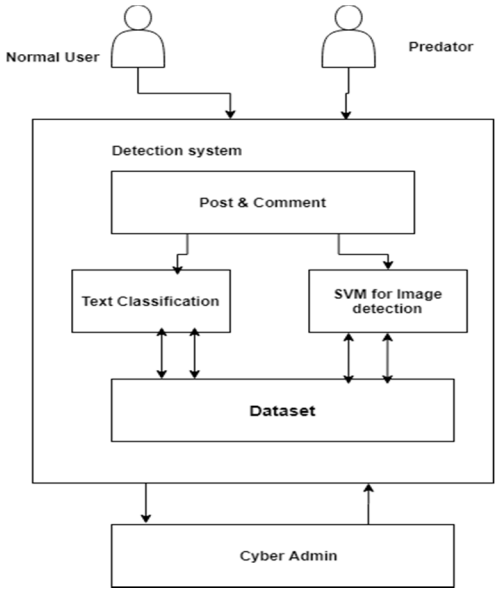


Fig 1. System Architecture

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

B. Implementation

1) We propose system for child predator detectionsystem.
2) We implementing 3 Modules for detection system.
3) User
4) Training Module
5) Cyber System
6) Function of System-:
7) User-: In this project we will show two types of user. First normal user another type showing predator behavior .
8) Training Module-: In training Module we using dataset for text classification and SVM algorithm for image detection. After Training Module will send predator report to cyber admin.
9) Cyber System-: Checking all predator report and taking action according to that report.

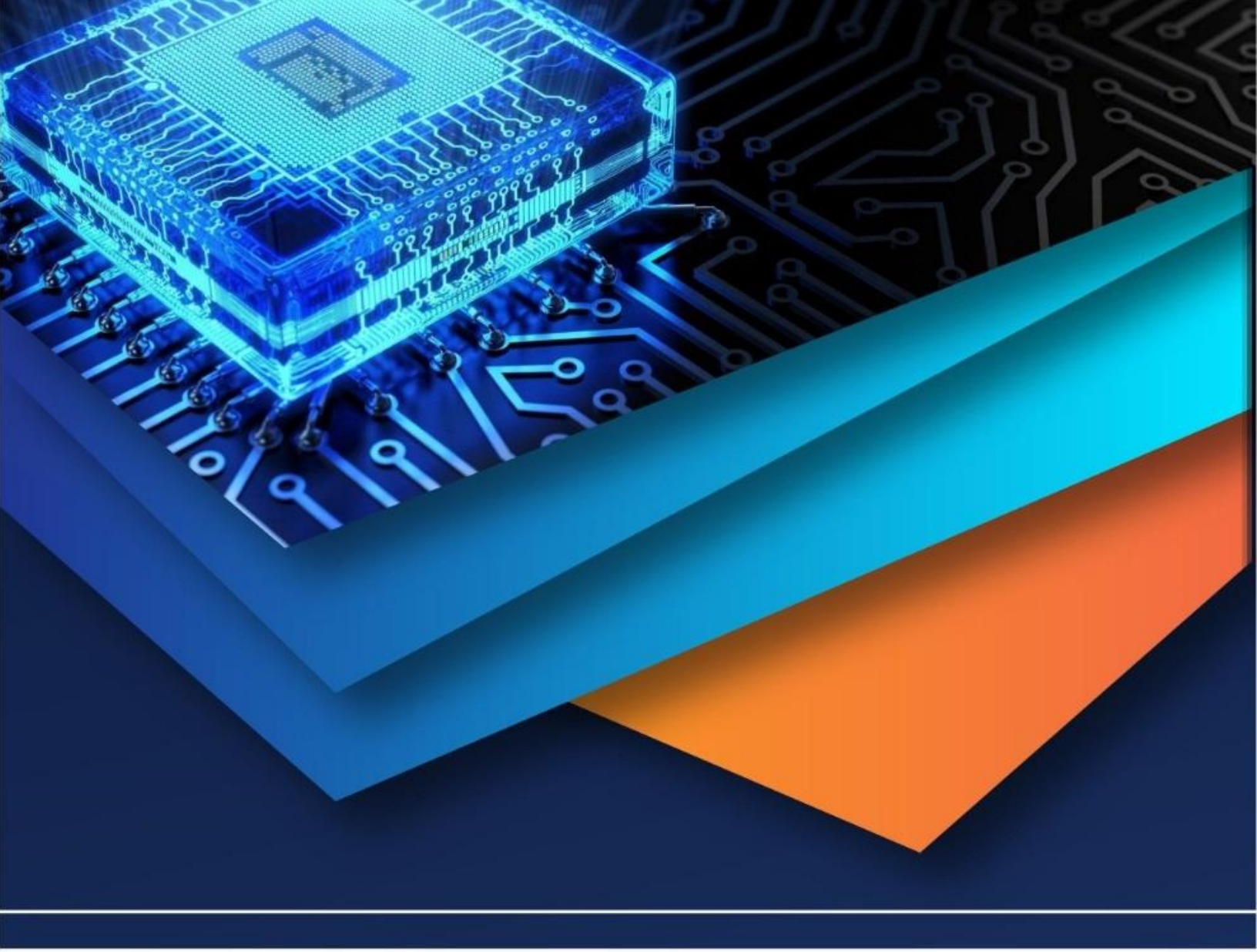
IV. CONCLUSION

The cost to children and society of sexual perpetration is too great to overlook the hazards of online solicitation. The aim of the groomer is to builda relationship with a child inorder to gain access to that child. When grooming takes place it is common that an adult groomer is pretending to be a child withcommon hobbies or interests to build a relationship that includes trust with the child. In this project we detect predator of child for child safety. And send report to cyber admin for action.

REFERENCES

[1] H. Ngejane, G. Mabuza-Hocquet, J. H. P. Eloff, and S.Lefophane, “Mitigating online sexual grooming cybercrime on social media using machine learning: A desktop survey,” in 2018 InternationalConference on Advances in Big Data, Computing and Data CommunicationSystems (icABCD), Aug 2018, pp. 1–6.
[2] N. Pendar, “Toward spotting thepedophile telling victim from predator in text chats,” in InternationalConference on Semantic Computing (ICSC 2007), Sep. 2007, pp. 235–241.
[3] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride,and E. Jakubowski, “Learning to identify internet sexual predation,” International Journal of Electronic Commerce, vol. 15, no. 3, pp. 103–122, 2011.
[4] G. Inches and F. Crestani, “Overview of the international sexual predator identification competition at PAN-2012,” in CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20,2012, 2012.
[5] E. Villatoro-Tello, A. Juarez-Gonz ´alez, H. J. Escalante, M. Montes-y- ´ Gomez, and L. V. Pineda, “A two-step approach for effective detection ´ of misbehaving users in chats,” in CLEF 2012 Evaluation Labs and Workshop, Online Working Notes, Rome, Italy, September 17-20, 2012, 2012.
[6] G. Eriksson and J. Karlgren, “Features for modelling characteristics of conversations,” in CLEF 2012 Evaluation Labs and Workshop, OnlineWorking Notes, Rome, Italy, September 17-20, 2012, 2012.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)