Detection and Prevention measures for Cyberbullying and Online Grooming

Aishwarya Upadhyay
Department of Computer Engineering
SAE, Savitribai Phule Pune University
INDIA

Akshay Chaudhari
Department of Computer Engineering
SAE, Savitribai Phule Pune University
INDIA

Arunesh
Department of Computer Engineering
SAE, Savitribai Phule Pune University
INDIA

Sarita Ghale
Department of Computer Engineering
SAE, Savitribai Phule Pune University
INDIA

Prof. S. S. Pawar
Department of Computer Engineering
SAOE, SP Pune University, Pune
INDIA

***Abstract*— From the day internet came into existence, the era of social networking sprouted. In the beginning, no one may have thought internet would be a host of numerous amazing services like the social networking. Today we can say that online applications and social networking websites have become a non separable part of one's life. Many people from diverse age groups spend hours daily on such websites. Despite the fact that people are emotionally connected together through social media, these facilities bring along big threats with them such as cyber attacks, which includes online grooming and cyberbullying.**

The primary objective of this project is to detect such cases over the social media and implement some preventive measures so as to avoid such incidents. Using machine learning and data mining techniques we can achieve our goal. The watchdog application could find out the unexpected, unpleasant and unwanted content that could be abusive-disrespectful text or nude images or unpermitted actions and help user removing it or preventing it from being spread all over. This will help the public to assure safety and nonviolence.

Keywords—Cyberbullying, Online Grooming, Watchdog, Machine Learning, Cyber Security.

I. INTRODUCTION

Social networking sites are being widely used today for multiple purposes like entertainment, networking, etc. Social networking sites are a stop for multiple reasons to billions of people today. All the social media platforms require the consent of all the participating people. However, Facebook is the most widely used platform to establish a relationship of any form. Social Networking Portal like Facebook attracts more several illegal activities. As the world knows, it is being used by billions of people. So for better understanding, a social networking portal similar to Facebook is developed by our team. Grooming is a process of preparing children for illegal sexual encounters by adults.

Cyberbullying[1] is a way of threatening or intimidating a person either through messages or by posting objectionable contents on the internet. This activity has caused a lot of damage to users of social media networks, especially amongst the youth. Several cases have been reported which show that Cyberbullying has caused teenagers to commit suicide.

Online grooming[2] is an activity of portraying yourself to be someone you are not and exploiting the victim for your purposes, often sexual. Cyber-groomers use internet platforms that are frequently accessed by the young people. Statistically 48 % of secondary school children have communicated to people they do not know when using social media, 58% of parents are concerned about the threats posed by strangers online, On average, 24% of high school students have been the victim of cyberbullying, 95% of the schools talk to children about not giving their personal information to anyone online[3].

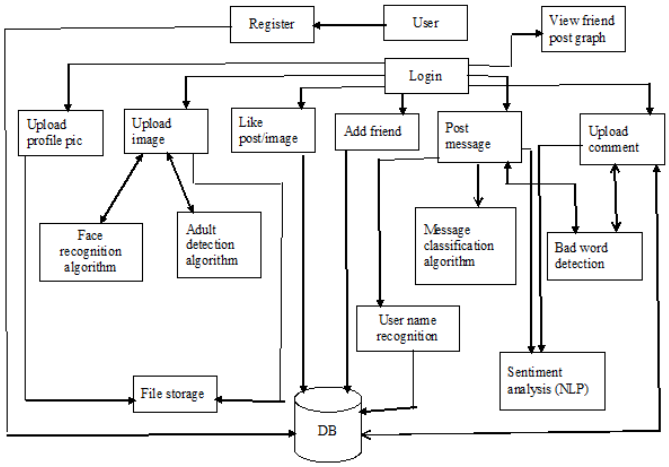
A typical example of online grooming is when adults and children exchange sexual text and media using social media and internet chat platforms. Adults are able to draw the line and are aware of the threats that are posed by these social media platforms, however, out of curiosity children, adolescents and teenager often fall prey to such attacks.

II. Literature survey

Cyber security is the major concern for all the security giants in the world. Recently there have been many people who have come up with various research works to make this web network a more secure place for the people who consider internet to be their workplace. Cyberbullying and online

grooming have been a very hot topic for research in the research communities.

A paper by Jeff David & al. "A Facebook Profile-Based TV Recommender System" gives us a detailed way of extracting information from the facebook profiles[4]. Then we can use it for our purposes, and machine learning domain has also given us different methods to use this data on a mass level to process them[4].



Other studies illustrate that persons aged between 12 to 17 years[5] mostly become victims of cyber-attacks on social networks. The consequences of physical or mental abuse are dramatic and include social isolation, depression and suicide. The main objective of proposed system is to classify the post. It gives probable effect of the original post so it's easier to understand social effect of any post on Facebook.

Ashcroft & al. gives a detailed info[6] on detection of A Step Towards Detecting Online Grooming and Identifying Adults Pretending to be Children. They've used effective ways to successfully detect the attackers. They're using step by step approach which includes user profiling, age detection, groomer identification, and Detecting Fake Kids as well. They've used datasets for experiments and upon the observations conducted by them for training and testing they're able to identify the groomers. In this work we have focused on a process that first identifies a person as an adult or a child based on the writing style. The second step in the process is to determine if a child is a fake child or not. Our results shows that it is possible to separate children and adults with good accuracy if they write formal text.

Website Perverted Justice[7] gives us real time cases that have been reported around the world there we get access to the chat logs that have been used by the police to identify attackers who does online grooming on innocent victims.

III. PROPOSED SYSTEM

Till now the papers and research works that we've referred to have been working on some singular parts of our project. We will be working collaboratively on all the parts viz. extraction of data, detection of online grooming, detection of bullying activities, and hence protecting the users from becoming a victim. Further addition to our project is that we would be taking actions against the perpetrators so that the people who come with an intention of committing crime would not be given many chances. Also we are going to notify users if someone have uploaded their images on the portal. The ones who are found to be violating the rules will permanently be banned from the portal after a few attempts.

A. Working:

To develop a system that classifies[8] Facebook posts with the help of NLP. Extract texts from posts, images of social media as Facebook to know how people feel about different posts and do sentiment analysis[9]. Classifying the post gives a probable effect of the original post, so it's easier to understand the social effect of any post on Facebook or any such social media. So we focus mainly on Facebook statuses, which we can view as opinions of users or their reaction on the content we want to analyze.

Fig No.1

Social Networking Portal similar to Facebook is developed and limitations of Facebook are removed by keeping watch on all types of activities performed on this Portal. Classification of user posts into different categories such as Politics, Education, Entertainment, Sports etc. is made based on analysis of post messages in this Social Network Portal. Any user can view other Facebook user's interest based on categories so that user can get to know about the interest of his/her friend.

The interest of any friend can be seen by a user of this network from categories graph data. The image uploaded by a user has to go through Adult Image Detection Algorithm so that image can be categorized into Adult or Safe categories. Adult images are blocked from coming into Social Network using adult detection based on Image Processing.

User posts, comments are analyzed to classify into Crime, Riots, and Worst or Vulgar categories using Text Mining from Dataset and Online API's. Examples of such words that are in the list are adult, bang, boyfriend, age, alone, clean, clothing, cute, date, dating, depressed, erotic, experience, hotel, hole, illegal, intercourse, intimate, home, lucky, loving, meeting, phone problem, relationship, sexy, shag, shaved, wet, whore, and young. A foreign word could be words borrowed from other languages as well as misspelled words, abbreviations or slang words.

Illegal posts must be checked using Sentiment Analysis based on text. Sentiment analysis is done on user posts and comments using NLP (Natural Language Processing) algorithms to detect user sentiments.

Here we are using two datasets where the User Messages are classified using **Bad Words Dataset** and **Sensitive Words Dataset** and user's bad count is recorded into the database. If any user exceeds the limit of bad posts or adult images then he/she is automatically banned from using this social networking portal. And hence, this portal is more secure and reliable than Facebook.

If a person posts name of any person then he/she is notified. If a user uploads a picture of any other user, then he/she gets notified about the photo uploading event. Then the user can allow or deny the process of uploading. While uploading picture of any friend, approval of that user is needed otherwise that picture will not be uploaded on portal.

Watch on user activities is much more in this social network than in any other social network in the world.

B. Purpose:

The main purpose of this system is to avoid adult content and accessing Facebook posts from the developers account. Also, here we aim at providing safety to the user so that no one could use his name and personal content like images in a wrong manner. User can upload Images, send messages, and post Comments to posts as well as make friends and upload profile pictures, etc. User Posts will be classified into different categories such as Politics, History, Education, Entertainment, Sports, etc.

C. Scope:

This system is used to classify user posts and comments which are further analyzed to classify into Crime, Riots, Worst or Vulgar categories. Sentiment analysis is done on user posts and comments using NLP (Natural Language Processing) algorithm to detect user sentiments. So this system proves to be very useful in immediately blocking the content which is not according to the moral behaviour or is offensive or doesn't seem ethical.

IV. ALGORITHM / METHODOLOGIES DETAILS:

A. Adult Image Detection Algorithm:

This area deals with the discipline of image and video analysis. Progress within these two areas are crucial to efficiently detect[10] violent or pornographic content. This Algorithm detects whether image is an adult image or not using Skin Tone Pixels detection and threshold value for number of Skin tone pixels allowed. Any user who uploads an adult image is warned about posting such images.

Every user has few chances before their account is blocked. If the user is found to be posting such content, there is a provision for the entry in the database which checks if the account shall be allowed to continue or not. If this entry exceeds after permissible amount attempts then, it results in the permanent failure of the account.

B. Irrelevant Posts Detection Algorithm:

Traditional text analytics framework, consists of a Text Preprocessing, Text Representation, and Knowledge Discovery phase. The Text Preprocessing phase makes the input text more consistent using techniques such as stop word removal and stemming. Stop words and function words are words that expresses a grammatical or structural relationship with other words in a sentence. These words have little or no meaningful content and therefore the topics that are discussed in the text that is analyzed is not taken into consideration. Stop words and function words are also known as grammatical words. We also count the occurrences of characters such as letters of the alphabet, various punctuation signs and numbers that are present in the text.

We have added a list of grooming and sexual words. This list contains words also from previous [11],[12], and [13], synonyms to these words and words that have appeared in conversations about sex and grooming.

Stop word removal refers to the procedure of removing general and meaningless words from analyzed text. Potential words can be found in relevant lists. *Stemming in turn* refers to reducing inflected or derived words to their base or root form.

This algorithm extracts keywords from text files containing different categories data and user messages are scanned to find whether they contain those keywords to classify messages into Crime/Worst/Riots category and to find the sentiment of a message.

C. NLP Algorithm:

In this algorithm, four categories of sentiment are defined. The analysis of comments and the amount of positive or negative reactions combined with sentiment analysis through different text mining modules give a rough overview of a user's social status. User Post is parsed into Tree Structure and then words of post are analyzed from tree structure to find the sentiment of the post.

In addition to image analysis, text analysis[14] of the textual content of shared images can provide helpful information in order to improve the classification accuracy. Another problem analysts are confronted with, is the distinction between pornographic and child[9] pornographic images, the distinction between legal and illegal content.

The main perspective of the system is accessing the Facebook posts from the user's account. User can upload Images, Messages, and Comments to posts as well as make friendship and upload profile pictures, etc.

V. CONCLUSION AND FUTURE WORK

In this paper we have highlighted an overall idea of how to implement a system aiming to protect the teenagers from online grooming and cyberbullying[9] attacks. If we are able to successfully detect such posts which are not suitable for adolescents or teenagers, we can very effectively deal with the crimes that are committed using these platforms. We would carry on further more research and would finalize how many number of times the users should be allowed to post such things and what would be the way to detect if anyone has posted something mistakenly.

When analyzing the most important features that were used in our classification of adults pretending to be children, foreign words was one of the features that was important. This is an interesting observation and future work could investigate the vocabulary variation that is used in chat rooms to obtain more relevant features.

Nevertheless, we have to concede that our current state of research focuses on the first step of the agenda and is mainly dedicated to the collecting process of different measurements which are necessary for discussion with professional experts with profound psychological or social science backgrounds. A strong collaboration between various disciplines is required to ensure appropriate results.

We further intend to add more modules to this application which could add more security features to the existing system.

In the near future, we plan to publish modules which have been proved successful. Realization of this Facebook watchdog application would provide additional security features to the current Facebook system and would help reducing the crimes that are committed through this platform.

ACKNOWLEDGMENT

We are more than happy to thank Prof. S. S. Pawar and Prof. Kiran Avhad for such a valuable guidance. Also pay our sincere thanks to all those who motivated and guided us along the whole path and gave us their valuable suggestions.

REFERENCES

[1] Homa Hosseinmardi, Richard Han, Qin Lv, Amir Ghasemianlangroodi and Shivakant Mishra "Towards Understanding Cyberbullying Behavior in a Semi-Anonymous Social Network" in ACM International Conference on Advances in Social Networks Analysis and Mining, 2014.

[2] Michael Ashcroft, Lisa Kaati and Maxime Meyer "A Step Towards Detecting Online Grooming- Identifying Adults Pretending to be Children" in 2015 European Intelligence and Security Informatics Conference 978-1-4799-8657-6/1

[3] "Onlineroomingwhatyoushouldknow", http://www.internetmatters.org/online-grooming

[4] Jeff David, Samir Bajaj, Cherif Jazra "A Facebook Profile based TV recommender system" CS229, Stanford University.

[5] Marlies Temper, Rainer Poisel, Simon Tjoa - "Facebook Watchdog: A Research Agenda for Detecting Online Grooming and Bullying activities"

[6] Michael Ashcroft, Lisa Kaati and Maxime Meyer "A Step Towards Detecting Online Grooming- Identifying Adults Pretending to be Children" in 2015 European Intelligence and Security Informatics Conference 978-1-4799-8657-6/15.

[7] "Perverted-justice.com - the largest and best anti-predator organization online," http://www.perverted-justice.com

[8] Jalel Akaichi, Zeineb Dhouioui, and Maria José López-Huertas Pérez "Text mining facebook status updates for sentiment classification" in System Theory, Control and Computing (ICSTCC), 2013 17th International Conference, 2013

[9] NaliniPriya. G and Asswini. M (2015) "A Dynamic Cognitive System For Automatic Detection And Prevention Of Cyber-Bullying Attacks", ARPN Journal of Engineering and Applied Sciences ©2006-2015 Asian Research Publishing Network (ARPN). VOL. 10, NO. 10, JUNE 2015.

[10] "Protective shield shield for social networks to defend cyberbullying and online grooming attacks" in Proceedings of 40 th IRF International Conference, Pune, India, ISBN: 978-93-85832-16-1, 2015.

[11] A. Gupta, P. Kumaraguru, and A. Sureka, "Characterizing pedophile conversations on the internet using online grooming," CoRR, vol. abs/1208.4324, 2012

[12] R. O'Connell, "A typology of child cyber sexploitation and online grooming practices," Cyberspace Research Unit, Tech. Rep., 2003.

[13] A.Kontostathis,"Chatcoder:Toward the tracking and categorization of internet predators," in Proc. of text mining workshop 2009 held in conjunction with the the 9th SIAM international conference on data mining., 2009.

[14] Julie Kane Ahkter, Steven Soria "Sentiment Analysis: Facebook Status Messages" Final Project CS224N, Stanford University.