Team Name: Bring a change

*Topic: Detection of Cyber Harassers on Social Media*

*Team Members:*

1.Gaurav Sharma

2.Pragiti Bagul

3.Abhishek Doshi

4.Abhinav

5.Deep Vora

6.Sanil Pandhare

*Index*

1.Abstract

2.Analysis of Problem

3.Design and Approach

4.Methodologies used

5.Metrics

6.Results and Conclusion

1) ABSTRACT

Problem Statement:

Online predators try to gradually seduce their targets through attention, affection, kindness, and even gifts, and often devote considerable time, money and energy to this effort. They are aware of the latest music and hobbies likely to interest kids. They listen to and sympathize with kids' problems. They also try to ease young people's inhibitions by gradually introducing sexual content into their conversations or by showing them sexually explicit material.

Purpose

Our web application scans text-based conversations and rates them on the probability that a user might be trying to sexually exploit children. It evaluates and rates conversation, characteristics and assigns an overall probability rating that helps determine whether the user is a predator or not.

Solution and Explanation:

The solution will detect suspect profiles based on child grooming behavior patterns followers, hate speech provokers, stalking and bullying mentality profiles and explicit content explorers (postings, comments) on social media platforms and other websites.

This will work on the basis of a predictive Machine Learning model which predicts whether the person under evaluation of certain researched parameters will be a cyber harasser or not.

This predictive model alongside the answers from the user to certain relatively less weighted questions will help determine accurately whether the online speech of the person classifies him as a cyber harasser.

2) Analysis of Problem:

What strategies do online sexual predators use?

Contrary to the widespread belief that online predators “trick” kids, research shows they rarely lie about their age or their motives. Their tactic is not one of deception but of seduction: they will shower a youth with attention, sympathy, affection and kindness, in order to persuade him or her that they love and understand them. The majority of adolescents who accept invitations to meet in person do so knowing that they will be engaging in sex. For 73 per cent of these youth, this will become a recurring sexual relationship.

Which age group are most at risk?

When it comes to online sexual exploitation, some youth are more at risk than others. Research indicates that 13- to 15-year-old girls are most vulnerable, particularly those who voluntarily place themselves in risky situations- by engaging in online discussions with strangers, flirting and talking about sex online, and by publicly posting personal and intimate information in Web environments such as social networking sites. It’s important to remember that young people who are most at risk online also tend to be those who are most at risk offline: they include youth who engage in harmful or risk-taking behavior in the real world, gay or questioning sexuality (males), youth who are experiencing physical or sexual abuse, youth who are experiencing mental health difficulties and youth who have relationship difficulties with parents or caregivers.

How must this problem be tackled?

Based on criteria found through research, cyber harassers and sexual predators can be found and recognized through thorough analysis of these which are identified as that of these people.

3) Design and Approach:

In this we have dynamically taken chats from the user and have distributed into 25 different features to train our machine. The dataset has been taken from a GitHub repository and the sample text of identifying sexual predators have been taken from <https://pan.webis.de> of the topic “Sexual Predator Identification”.

After taking the dataset we have applied machine learning to accurately predict whether the target person is a sexual predator or not.

In machine learning we have used logistic regression model and used a dataset of around three lakh and tested it on 97,000 test cases and achieved an accuracy of about 98.8%.

The chat shall be given by the user on the website we have created which also contains other helpful information for the user. After giving the input the user will be further be asked a few questions for the user’s opinion and by combining all the data, an estimate of how much the online person may be a sexual predator is determined.

4) Methodologies used:

The main web app was developed on HTML, CSS and Java Script and the chat history and the responses by the user will be stored in SQL database and by comparing the database to the model by the machine learning program the computer shall predict with more than 95% accuracy whether the person whom the user is chatting is a Sexual Predator or not.

The machine learning program was made using logistic regression model involving NumPy and Pandas.

5) METRIC:

Space Occupied:

Accuracy: 99.8%

6) Future Scope:

* We can add image based and phrase-based detection in the future.
* If the same person has been reported previously on our app, he would be classified as a predator or a threat to the society.
* We can gain access to private conversations from the police and alert a potential predator before hand.

7) Results and Conclusion:

Through the use of a predictive Machine Learning model and by implementing a web application which uses logistic regression.