Termination of Cyber-Sexual Harassment and Abuse with Teenagers using Artificial Intelligence

Ratnark Gandhi

Data Scientist, BTM Financial, 153-54, Spaze I Tech Park, Sector 49, Gurgaon-122018, India

Corresponding author: ratanxcd@gmail.com

ABSTRACT

With the rapid pace of lives transforming by the means of innovative technologies being introduced into our lives, the rate of people willing to abuse technologies is increasing as well. The motive of method being introduced by me in this paper is to monitor textual conversation of underage adolescents on social media platforms and creating alerts (if found a case of sexual harassment/abuse) through an application. In the proposed method I will be concentrating more on the algorithm used in an application for the classification of textual messages (to provide new insights towards chances of tackling cybercrime) rather than the development of application itself.

Keywords: Adolescents, sexual harassment, cybercrime, abuse

Artificial Intelligence

Artificial Intelligence is a large branch of computer science that deals with the creation of smart machines capable of performing tasks that usually require human intelligence. AI is an interdisciplinary science with different approaches, but in nearly every field of the tech industry, developments in machine learning and deep learning are causing a paradigm change.

Why Teenagers?

As a human, teenagers are still evolving, but they don't know what their full identity is. It is very easy for the world to let them know what they should be and how they should behave as an individual, without researching it and forming their own opinion on it, adolescents are likely to adopt ideas. It is easy for a person to put ideas in the head of a child, since they may lack assertiveness that helps them to understand what they should say no to and yes to, making it very difficult to interpret mind games. That's the main reason behind teenagers being this paper's target.

What is sexual harassment?

According to EEOC, Sexual Harassment can be defined as unwelcome sexual advances, requests for sexual favors, and other verbal or physical conduct of a sexual nature when:

❒ Submission to such conduct is made either explicitly or implicitly a term or condition of an individual›s employment.

❒ Submission to or rejection of such conduct by an individual is used as a basis for employment decisions affecting such individual.

❒ Such conduct has the purpose or effect of unreasonably interfering with an individual›s work performance or creating an intimidating, hostile, or offensive working environment.

Unwelcome Behaviour is the critical word. Unwelcome does not mean "involuntary." A victim may consent or agree to certain conduct and

actively participate in it even though it is offensive and objectionable. Therefore, sexual conduct is unwelcome whenever the person subjected to it considers it unwelcome. Whether the person, in fact, welcomed a request for a date, sex-oriented comment, or joke depends on all the circumstances.

Cyber Sexual Harassment

Sexual Harassment can be categorized into various forms, one of which and the main focus of this paper is Cyber Sexual Harassment. When the victim of Sexual Harassment is targeted through the means of any Online Social Platform such as Facebook, Instagram, Telegram, E-Mails, etc case falls under the umbrella of Cyber Sexual Harassment. Means of harassment through the internet could be:

❐ Abusive Texts

❐ Inappropriate Images

❐ Sexual Threats

❐ Offensive voice messagesare based on an individual's traits.

❐ Erotic stickers and gif(s)

❐ Links comprising of sexual content

Effects of Cyber Sexual Harassment

Being harassed sexually can impact an individual in several ways, affecting physical well-being as well as causing mental traumas and breakdown.

Some of the consequential reactions are listed below [2]:

Physiological Reactions

1. Headaches
2. Lethargy
3. Phobias
4. Panic Reactions
5. Sleep Disturbance
6. Nightmares
7. Weight Loss

Psychological Reactions

1. Depression, anxiety, shock, denial
2. Anger, fear, frustration, irritability
3. Insecurity, embarrassment, feelings of betrayal
4. Confusion, feelings of being powerless
5. Shame, self-consciousness, low self-esteem
6. Guilt, self-blame, isolation

Statistics Involved

❐ There are more than 42 million survivors of sexual abuse in America. *(National Association of Adult Survivors of Child Abuse)*

❐ 1 in 3 girls is sexually abused before the age of 18. *(The Advocacy Centre)*

❐ 1 in 5 boys is sexually abused before the age of 18. *(The Advocacy Centre)*

❐ 1 in 5 children is solicited sexually while on the Internet before the age of 18. *(National Children's Alliance: Nationwide Child Abuse Statistics)*

❐ 30% of sexual abuse is never reported. *(Child Sex Abuse Prevention and Protection Centre)*

❐ Nearly 70% of all reported sexual assaults (including assaults on adults) occur to children age 17 and under. *(Children's Advocacy Centre)*

❐ 90% of child sexual abuse victims know the perpetrator in some way. *(U.S. Department of Justice)*

❐ Approximately 20% of the victims of sexual abuse are under age eight. *(Broward County)*

❐ 95% of sexual abuse is preventable through education. *(Child Molestation Research and Prevention Institute)*

❐ 38% of the sexual abusers of boys are female. *(Broward County)*

❐ There is worse lasting emotional damage when a child's sexual abuse started before the age of six and lasted for several years. Among child and teen victims of sexual abuse, there is a 42 percent increased chance of suicidal thoughts during adolescence. *(American Counselling Association)*

❐ "More than 90% of individuals with a developmental delay or disability will be sexually assaulted at least once in their lifetime." *(Valenti-Heim, D.m Schwartz L.)*

❐ "There are nearly half a million registered sex offenders in the U.S. – 80,000 to 100,000 of them

are missing." *(The National Centre for Missing and Exploited Children)*

❑ "A typical pedophile will commit 117 sexual crimes in a lifetime."

Around the Internet

In the 2016 Annual Report from the Internet Watch Foundation (IWF),18 whose remit is to remove Child Sexual Abuse Material hosted anywhere in the world, including non-photographic content hosted in the United Kingdom, of the 105,420 reports processed in 2016, 57,162 were received from public sources withthe remainder identified through analysts actively searching the open Internet using a combination of analyst searching and bespoke web crawlers. Of these, 57,335 URLs contained Child Sexual Abuse Content, twenty- eight percent of these reports were confirmed as containing serious cases regarding the same.

INHOPE (a global network of hotlines from 49 members whose remit is to deal with illegal content online and remove Child Sexual Abuse Material from the Internet),20 reported that they received 9,357,240 reports in 2016, with 8,474,713 confirmed as containing Child Sexual Abuse Material. Cybertip. ca, the hotline hosted by the Canadian Centre for Child Protection processed 40,251 reports in 2016-17, 49% of which were forwarded to law enforcement, child welfare, and (or) INHOPE, or notice was sent to an Electronic Service Provider to report Child Sexual Abuse Material hosted by a service provider in Canada or the United States.

During 2015, in addition to receiving reports of online Child Sexual Abuse, 85% of hotlines in 2015 also accepted other types of reports, which included for example racism/hate speech (69%), adult pornography (64%), bullying (62%) and self-harm/suicide (44%).

However, in this global study of hotlines, approximately one-third of those surveyed indicated that Child Sexual Abusereports made up the majority of their workload. The National Centre for Missing and Exploited Children (NCMEC), serves as the United States of America's clearinghouse for Child Sexual Abuse Material through the Cyber Tip-line. This provides an online mechanism for members of the public and electronic service providers (ESPs) to report incidents of suspected child sexual exploitation. This includes Child Sexual Abuse Material,sexual exploitation of children in travel and tourism, online enticement, trafficking of children for sexual purposes, child sexual molestation, misleading domain names or words, and unsolicited obscene material sent to a child. In the report to the United States House of Representatives Subcommittee in March 2017, it was noted that over recent years, the volume of Cyber Tip-line reports received by NCMEC had increased from over 1.1 million reports in 2014 to more than 4.4 million reports in 2015, to more than 8.2 million reports in 2016.

Detection of Cyber Sexual Harassment

The above reports show that 20% of all teenagers experience Cyber Sexual Harassment through social media before reaching 18, and this number might be much bigger in absence of proper resources to detect such crimes. We live in a world where everyone is free to share their views through social media, but some lawbreakers take advantage of the liberty provided by these platforms to reach out to anyone, they start targeting innocent children. My proposed architecture focuses on tackling and monitoring of textual messages often received by children by unknown/known people who tend to harm children because of their sexual urges.

Proposed Architecture

1. Real-time monitoring of texts being exchanged across social media platforms, options to include, or exclude specific platforms are always under the control of user's guardian.

2. A list of applications to be monitored is created internally according to options chosen by the user along with the guardian. Whenever the chosen application starts monitoring starts automatically.

3. Whenever a text message is tagged as Sexually Inappropriate by algorithm, it starts searching for sexually offensive words (collection of 453 such words are with me already) in text messages sent by the opposite person. As soon as any word matches from a list of abusive words with me, a warning trigger is generated and sent across different platforms to the local guardian.
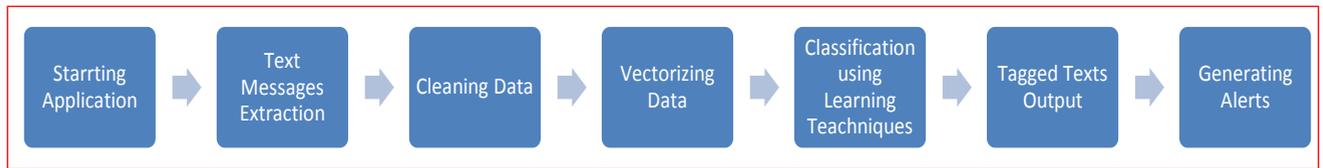
Fig. 1: Process architecture

User Data Protection

Since the user data in criminal cases like sexual harassment is very sensitive, hence as a developer of this application I'm bound to keep the total privacy of user's data. I know that if I failed to do so it will have disastrous consequences for my application from both legal and PR standpoint. I will take few necessary steps to safeguard my application against security breaches:

No Access to Sensitive Data: No one from the team of developers/executives (if any) should've access to the client's data without being reported for any technical issue by the user and only if the user agrees to share data with the helping executives. There should be a centralized system only through which executives would be able to help the user.

Dedicated Server: Using shared servers to cutdown the costs is normal practice among running businesses, however, this comes with a security risk, as these servers share risk with other websites. While dedicated servers are expensive, they also provide an additional layer of protection for both you and your client.

Proper Firewalls and Antivirus Protection: Basic security measures to safeguard data include the installation of firewalls and antivirus protection. While firewalls would help to prevent unauthorized access, antivirus software helps prevent, detect, and remove harmful programs from your computer and hence reducing the vulnerabilities of the whole system.

Periodic Security Updates: Security hacks do exploit known security holes, so in order to prevent exploitation of security holes, developers should provide regular updates covering those holes and pushing them to users.

METHODOLOGY

Sample Set

In this case, finding the accurate dataset for this specific task was kind of impossible, so I prepared my dataset by mixing:-Sexual Harassment Conversation between victims and their criminals from around 8320 cases, extracted directly from stories told by multiple victims which were provided by Safecity India. Along with normal text conversations containing 9360 texts provided by Unicamp with a list of 453 sexually offensive words taken from GitHub.

Training and Testing Set: After mixing up both Abusive as well as Normal text messages, I got 17680text messages with me. Splitting it as 80:20 gave me with 14144 tagged text messages for training the algorithm and 3536 for testing.

Research Methodology

Once an application starts which belongs to the list of apps to be monitored, real-time extraction of text starts after the following steps used for detection:

Reading and Conversion of raw and semi-structured text into structured format: When the data is extracted directly from text messages, it's always in a semi-structured format. To operate and extract details from text messages I converted semi-structured texts into a structured format.

Cleaning Data: I have structured text with me now,but structured texts still contain unwanted or irrelevant content for detection, removing them increases efficiency in reading and extracting meanings from texts as well as decreasing execution time for algorithms.

Steps involved in cleaning are:

- ❐ **Removing Punctuation:** I used the inbuilt PUNCTUATION method which returns all sets of punctuation to me, imported from the STRING library.

- ❐ **Tokenization:** It is a way of separating a piece of text into smaller units called tokens. Tokens produced are then used to prepare a vocabulary, vocabulary refers to the set of unique tokens in the corpus.

- ❐ **Removing Stop Words:** Stopwords are the most

common words in any natural language adding extra and irrelevant words for processing. Words like "is", "in", "for"," where", etc are considered as stopwords. After the removal of stopwords, I am left with only meaningful and relevant words for text processing.

❑ **Lemmatising/Stemming:** It is the process of reducing inflected (or sometimes derived) words to their word stem or root. Lemmatising is the process of grouping together the inflected forms of a word so they can be analyzed as a single term, identified by the word's lemma. For example, words like "typing", "typed" will reduce to the word "type". Leaving me with only the necessary words to understand the meaning of its usage.

Vectorizing Data: Process of encoding text as integers to create feature vectors, where feature vector is an n-dimensional vector of numerical features that represents some object. Hence, in my context,I took individual text messages and converted them into vector numerals. Normal texts were converted into numerals so that, python and machine learning algorithms can understand the data.

Learning Techniques Used

Naive Bayes: Naive Bayes constitutes multiple algorithms that I can use while doing text classification. The member of Naïve Bayes which I used is Multinomial Naive Bayes (MNB). The main reason behind using this algorithm is I got really good results when data available with me was not much and computational resources were scarce.

Support Vector Machines: SVM is one of the algorithms being used majorly for text classification. It is similar to Naive Bayes in the manner that it does not need much training data to start providing accurate results. In short, SVM takes care of drawing a line of a hyperplane that divides a space into two subspaces,one subspace that contains the vectors that belong to a group and another subspace that contains vectors that do not belong to that group. These vectors represent the training texts and a group is a tag given to text messages of similar type.

Output

After classifiers are trained on the training dataset, I tested those classifiers on the test set giving an output of text tags as Sexually Inappropriate or Normal.

Table 1: Confusion Matrix for MBT

N = 3536	Positive	Negative
Positive	1893	254
Negative	181	1208

Table 2: Confusion Matrix for SVM

| N = 3536	Positive	Negative
Positive	1932	153
Negative	127	1324
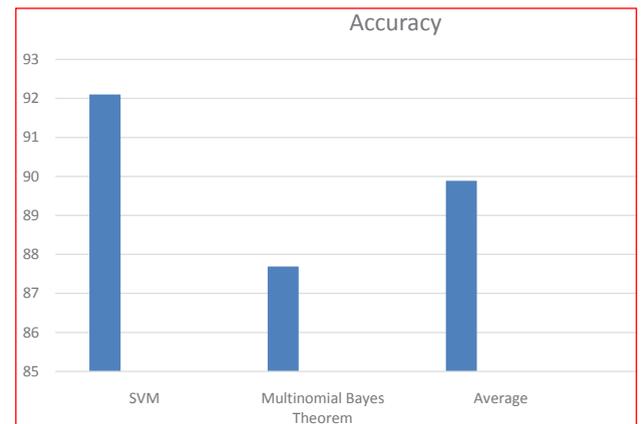
Fig. 2: Average and Individual Algorithm's Accuracy Plot

Alert Generation

Once a text message is tagged as sexually abusive, the presence of abusive words (present in list taken from GitHub) in text messages is tested and the presence of such words leads to the generation of warning trigger sent to local guardians.

Tools Used

The programming and preprocessing part on the collected data set was done on SPYDER IDE, accessed from ANACONDA with the help of few open-source and pre-installed packages available on spyder including PANDAS, SCIKIT-LEARN, etc.

Similar Cyber Crimes and Statistics Involved

Around 9.2% of 630 adolescents surveyed in the Delhi-National Capital Region had experienced cyber-bullying and half of them had not reported it to teachers, guardians, or the social media companies concerned, a recent study by Child Rights and You, a non-governmental organization, found.

Vulnerability rose with internet use: 22.4% of respondents, aged 13-18 years, who used the internet for longer than three hours a day were vulnerable to online bullying, while up to 28% of respondents, who used the internet for more than four hours a day, faced cyber-bullying.

CONCLUSION

The proposed NLP based Emotion Recognition using Natural Language Processing approach works to recognize sexual harassment and abuse efficiently and effectively. Algorithms used in learning techniques gave an aggregate accuracy of 90.25% even with a limited amount of data available. Also,I believe that with the expansion of the training dataset provided for learning techniques, accuracy will improve substantially.

REFERENCES

ECPAT website (https://www.ecpat.org/news/online-child-sexual-abuse-material-the-facts/) provides unknown facts on online child sexual abuse materials present worldwide.

Race Against Abuse of Children Everywhere (RAACE) website (https://www.raace.org/join-the-raace) has built their mission to prevent hidden epidemic of child sexual abuse.

Rhea Maheshwari, India Spend, 13 Mar 2020, accessed 3 September 2020, Link: https://scroll.in/article/956085/in-one-year-alone-cyberbullying-of-indian-women-and-teenagers-rose-by-36

The U.S. Equal Employment Opportunity Commission (EEOC) website (https://www.eeoc.gov/overview) mission is to prevent and remedy unlawful employment discrimination and advance equal opportunity for all in the work place.

University of South Florida (USF) website (https://www.usf.edu/student-affairs/victim-advocacy/types-of-crimes/sexual-harassment.aspx)has a separate department for crime info and support, which provides adequate definition and examples of sexual harassment.