iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** X **Month of publication:** October 2021

DOI: https://doi.org/10.22214/ijraset.2021.38496

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: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

Proactively Discouraging Cyberbullying Activities

Puneetha KR[1], Reena Rodrigues[2], Shreenidhi Shetty[3], Pooja H[4]
[1, 2, 3, 4]*Computer Science and Engineering Department, St Joseph Engineering College, Mangaluru*

Abstract: ***Research into cyberbullying detection has increased in recent years, due in part to the proliferation of cyberbullying across social media and its detrimental effect on young people. Cyber bullying is one of the most common problems faced by the internet users making internet a vulnerable space hence there has to be some detection that is needed on the social media platforms. Detecting the bullies online at the earliest makes sure that these platforms are safer for the user and internet indeed becomes a platform to share information and use it for other leisure activities. Even though there has been some research going on implementing detection and prevention of cyber bullying, it is not completely feasible due to certain limitations imposed. In this paper lexicon-based approach of the NLTK sentiwordnetis used to differentiate the positive and negative words and produce results. These words are given negative and positive values greater than or less than zero for positive and negative words respectively. Lexicon based systems utilize word lists and use the presence of words within the lists to detect cyberbullying. Lemmatization is used to find the root word. This paper essentially maps out the state-of-the-art in cyberbullying detection research and serves as a resource for researchers to determine where to best direct their future research efforts in this field.***
Keywords: Abuse and crime involving computers, natural language processing, sentiment analysis, social networking

I. INTRODUCTION

Bullying is defined as intentional aggression carried out repeatedly by one individual or a group of individuals towards a person who is unable to easily defend him or herself. Cyberbullying is, defined as “an aggressive, intentional act carried out by a group or individual using electronic forms of contact, repeatedly or over time against a victim that cannot easily defend him or herself”. Cyberbullying has been found to be quite prevalent on social media with as many as 54% of young people reportedly cyberbullied on Facebook[1].

Web 2.0 has profoundly affected contact and relationships in today's society. While most internet use is harmless and the advantages of digital communication are obvious, the online environment of privacy and anonymity makes people vulnerable, with cyberbullying being one of the major threats. Bullying is not a recent phenomenon and cyberbullying happened as soon as digital media became the main means of communication. As social media users exponentially increased, cyberbullying has emerged as a form of bullying through electronic messages. Social networks provide a rich environment in which bullies can use these networks as vulnerable to attacks on victims. Given the effects of cyberbullying on victims, it is necessary to find appropriate actions to detect and prevent it. Machine learning can be helpful in detecting the bullies language patterns and thus can generate a model for automatically detecting cyberbullying actions[2]. Body shaming, bad slurs are becoming the new normal, the social media platforms have to take certain actions against such behavior by implementing algorithms in the existing system to detect such behavior and discourage it by warning such bullies and taking strict actions.

Detection of cyber bullying and subsequent preventive measures are the main courses of action in the fight against cyber bullying. Techniques typically used for sentiment analysis can be used to detect electronic bullying using characteristics of messages, senders, and the recipients[3]. It should, however, be noted that cyberbullying detection is intrinsically more difficult than just detecting abusive content. The detection method should identify the presence of cyber bullying terms and classify cyber bullying activities in social networks such as Flaming, Harassment, Racism, and terrorism. Different forms of bullying triggers for various reasons, which have a varying impact on the person being bullied. Bully identification is not the big concern, detection of bully types is also a necessary goal in this field to mitigate the effects of cyber bullying. In order to overcome these problems, we design a website which uses lexicon-based approach to detect any bully comments by tokenization and lemmatization. Different approaches to creating dictionaries have been proposed, including manual[1] and automatic[2] approaches. NLTK tool sentiwordnet is used for analyzing and classifying the comments into positive and negative words and produce results. These words are given negative and positive values greater than or less than zero for positive and negative words respectively[4].

The warnings issued will make sure that such bullies get a chance to look into the behavior and improve it. But if they continue doing the same they will be barred from using the social media like interface The need of the hour is to make the web a safer place by eliminating the negativity to some extent by implementing a few changes and introducing algorithms to detect the bullies.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

II. ARCHITECTURE

The architectural diagram of the proposed methodology has been depicted in Fig.1. It includes the following components:

A. Social Media

It is an online platform where the users can login onto the system using their user ids and password. They can post pictures, view posts and even comment on the posts. It has a platform enabled in it which helps in identifying bad words and taking necessary actions to prevent it from happening again.

B. User Post

The users can comment on the pictures which will be taken as input by the software for further processing of the statement.

C. Pre-Processor

The main functionality of this pre-processor is to prune the words .i.e., extracting only necessary characters that will help in analyzing whether it is a good statement or a bad statement.
It also minimizes the overall time taken for detection as some of the unnecessary characters are removed as it need not be analyzed.

D. Analyser

In this module, we are making use of sentiwordnet algorithm in order to classify the words into two classes – Good words or bad words. Also, we are going to keep track of the number of times the user makes such comments. If the count exceeds 3, the user will be sent a warning mail. If he repeats even after sending the warning mail, he will be blocked.

E. Alert Mechanism

It is used to send alert messages to the user each time he makes an abusing comment. Alert is sent as mail to the user when he/she makes an abusing comment for the third time.

F. Database

It is used to store the posts, comments and user login information.

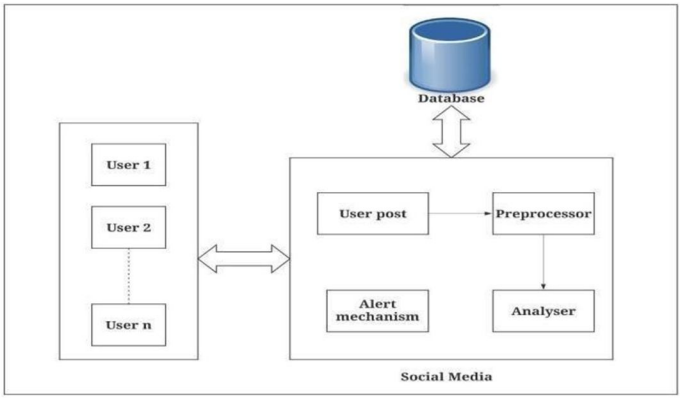


Fig.1. Architectural diagram

III. METHODOLOGY

In this section, we elaborate on the text pre-processing steps that has to be done in order to make the text feasible for sentiment analysis. There are four substeps: (1) tokenization, (2) stopping words removal, (3) pos tagging, (4) lemmatization.

1) In step (1), the larger body of text is spilt into smaller lines, words called as tokens. We define a list of unwanted characters like punctuation, special characters which will be compared with original text. If there is a match of characters, they will be removed from original sentence. The filtered sentence is then split so as to form a list of words. For e.g. Sentence1– Hello Friends, Welcome to the world of Natural Language Processing Word Tokenization of Sentence1 are as follows: ‘Hello’ ‘Friends’ ‘,’ ‘Welcome’ ‘to’ ‘the’ ‘world’ ‘of’ ‘Natural’ ‘Language’ ‘Processing’.

2) In step (2), useless words (such as “the, a, an, is, for”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of search query are removed as they do not play an important role in analysing sentiment associated with comments. We do not want such words to take up space or valuable processing time. For this purpose, we define a set of stopping words or they can be imported from nltk.corpus as stop words. These set of stop words are then compared with the tokens. If they match, they will be filtered from the list of tokens.

3) In step (3), the tokens are marked to a particular part of speech based on its definition and context. For this purpose, we make use of pos_tag() function from nltk which uses rule based tagging. It uses a dictionary or lexicon for getting possible tags for tagging each word. If the word has more than one possible tag, then rule- based taggers use hand written rules to identify correct tag. Disambiguation is also performed in rule-based tagging by analysing linguistic features of words along with its preceding as well as following words. For e.g. If the preceding word of a word is article, then the word is noun (such as A Mercedes, the party). Rule-based tagging follows a two-stage architecture –

a) *First Stage:* Usage of dictionary to assign each word a list of potential parts-of-speech.

b) *Second Stage:* Usage of large list of handwritten disambiguation rules to sort down the list to a single pos for each word.

4) In step (4), the tokens are lemmatised or stemmed to get the root/base words. Here we are aiming at removing inflectional endings and to return base form of the word which is called as lemma. It does morphological analysis which is a process of providing grammatical information about word on the basis of morpheme it contains. Morpheme is a smallest meaningful unit of a given word. For e.g. the word ‘unbreakable’ has three morphemes- ‘un’, ‘break’, ‘able’. Morphological analysis can be done using various methods like finite state automata, DAWG, Finite state transducer and so on. This is done so as to modify the root so that it fits properly in the context. Alongside morphological analysis, context sensitiveness of a word is considered. For e.g. if we have words like operate, operating, operation, operates, a stemming algorithm is going to return oper as a root word which is inappropriate, on the other hand, lemmatization returns operate as root word as it considers context. Another example which shows the importance of considering the context is shown below: Suppose we have words like operational research, operating system, if lemmatization process outputs operate research, operate system, it is not a perfect match for the query operation research or operating system respectively as it is a noun. Hence, if we have the knowledge of context (preceding words and following words), we can easily identify the lemma.

A. *WordNet database*

As large number of words are required for sentimental analysis, we acquire it from wordnet which is a lexical database of semantic relations between words. It links words into semantic relations including synonyms, hyponyms and meronyms. The synonyms are grouped into synsets with short definitions and usage examples. Therefore, wordnet is a combination and extension of a dictionary and thesaurus.

B. *Lexicon Based Sentiment Analysis*

Application of a lexicon is one of the two main approaches to sentiment analysis and it involves calculating the sentiment from the semantic orientation of word or phrases that occur in a text. With this approach a dictionary of positive and negative words is required, with a positive or negative sentiment value assigned to each of the words. Different approaches to creating dictionaries have been proposed, including manual[1] and automatic[2] approaches. Generally, in lexicon-based approaches a piece of text message is represented as a bag of words. Following this representation of the message, sentiment values from the dictionary are assigned to all positive and negative words or phrases within the message. A combining function, such as sum or average, is applied in order to make the final prediction regarding the overall sentiment for the message. In our work we have decided to apply a lexicon-based approach in order to avoid the need to generate a labelled training set. The main disadvantage of machine learning models is their reliance on labelled data. It is extremely difficult to ensure that sufficient and correctly labelled data can be obtained. Besides this, the fact that a lexicon-based approach can be more easily understood and modified by a human is considered a significant advantage for our work. Given that the data pulled from social media are created by users from all over the globe, there is a limitation if the algorithm can only handle English language. Consequently, sentiment analysis algorithm should be more easily transformable into different languages which is our future work.

A dictionary of positive and negative words is generated using sentiwordnet tool which is an automatic approach. Sentiwordnet makes use of semi-supervised learning step[3] and a random walk step[4] to generate a ternary classifier (three classes – positive, negative, neutral). The tool takes all the synsets from wordnet and assigns sentiment for each of them using ternary classifier.

As a result, each synset is associated with three numerical scores i.e., positive, negative and neutral which indicates how much positive, negative or neutral the terms contained in synsets are. The sum of these three scores equals to 1.0 and it ranges between [0, 1]. Also the scores for same word differs in different context as there are different senses of same term and in different context a word can mean a different thing. For e.g. [estimable (J, 3)] which is an adjective(J) means "may be computed or estimated" in context (3). Here sentiwordnet assigns the scores as follows: pos (0), neg (0), neutral (1.0). However, [estimable (J, 1)] in context (1) means "deserving of respect or high regard". In this case, the scores assigned are: pos (0.75), neg (0), neutral (0.25). Hence, it is equally important for us to consider the contextual sensitivity of words while doing the analysis.

In order to get the sentiment that is associated with each token, we are applying a synset() function on the lemmas (output of pre-processor) in order to get it's representation in wordnet. It outputs (name.n.01) where name represents the word for which we want the sentiment score, n represents part-of-speech associated with it, 01 represents the context and it varies according to different synonyms associated with each words. We are then applying senti_synset() function which takes output of synset() as parameter. It then outputs the score associated with each words. This process is repeated until all the words are analysed. The difference between positive and negative scores are calculated for each word and their sum is stored in a variable. If the value stored in variable is greater, then positive sentiment will be associated with sentence, otherwise negative sentiment is associated.

IV. EVALUATION AND ANALYSIS

To analyze the mechanism and results of the proposed methodology we consider different test cases and evaluate them. There are four categories that are considered as follows:

A. Website

The efficient simulation of the website depends on various factors and the user interface experience is essential. Hence certain test cases are used to determine the efficiency of the website.

Table 4.1 Test Cases for Website

No.	Test Cases	Expected Output	Observed Output	Result
1.	Clicking on Login Button	Directs the page to add post.	Successfully directs the user to next section.	Pass
2.	Clicking on the Register button	the user is directed to the registration page	Successful entry of user credentials.	Pass
3.	Validation of email entered	Valid emails should be accepted	Valid emails are accepted or else error message id displayed.	Pass
4.	Clicking on the Done button	It should redirect the registered user to the Login page	It directs the user to the Login page	Pass
5.	Clicking on the post button	The text should be posted on the public post section.	The text is posted on the public post section.	Pass
6.	Commenting	The user should be able to add comments	The is able to add comments	Pass

The above test cases are thoroughly tested to simulate the social media platform.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue X Oct 2021- Available at www.ijraset.com

B. Comments

For the user to have a smooth experience it is very essential for a commenting mechanism that makes sure that no user has to face any issue while commenting on a post.

Table 4.2 Test Cases for Commenting Mechanism

No.	Test Cases	Expected Output	Observed Output	Result
1.	Non offensive comments	The user should be able to comments casual and non-offensive comments	The user is able to comments casual and non-offensive comments	Pass
2.	Using offensive language to bully the owner of the post	A warning should be sent to the bully through the mail.	The bully gets a warning mail	Pass
3.	Bully continues using the abusive language and the count > 3	He/She should be barred from using the website	The bully is blocked from posting or commenting on public post.	Pass

C. Database

The database keeps record of the user information and the comments posted. Also, it maintains a counter to keep record of the warnings sent to the user also the status of the blocked use with 0 being not blocked and one being blocked. Since, the django server used the registration details are stored in the inbuilt application of django.

Table 4.3 Test Cases for Database

No.	Test Cases	Expected Output	Observed Output	Result
1.	Counts of negative comments made by user in database	The count should incremented for the user every time the bully words are used	The count is incremented for the user every time the bully words are used	Pass
2.	Storing User Informati on	All the user information has to be stored in the database	The user information gets stored in the database	Pass

D. NLTK

The NLTK helps with the classification of the comments with the wordnet inheriting all the dictionary of words used to classify, sentiwordnet is used to classify the words after lemmatization as positive, negative and neutral words. All of these functions are considered for the accuracy of the results generated and blocking of the user.

Table 4.4 NLTK Functions

No.	Test Cases	Expected Output	Observed Output	Result
1.	Breaking Down of Comments (Tokenization)	The comments should be broken down into words and comments	Successful break down of comments into words and sentences	Pass
2.	Removal of stopping words	Words such as like ,a , is connecting words are to be removed	Connected words are removed successful	Pass
3.	Getting polarity combination of word	Polarity of the word is to be determined by the sentiwordnet function	The polarity of the combination words is deteremined	Pass
4.	Classification of comments based on the value associated	Classification into negative and positive comments <0 is negative >0 is positive	Successful classificatio n of comments into negative and positive	Pass
5.	Lemmatization of tokens	Reducing the words to its stem	Successful Lemmatizati on	Pass
6.	Stemming	Reducing words to its root form	Successful stemming	Pass

V. FUTURE WORK

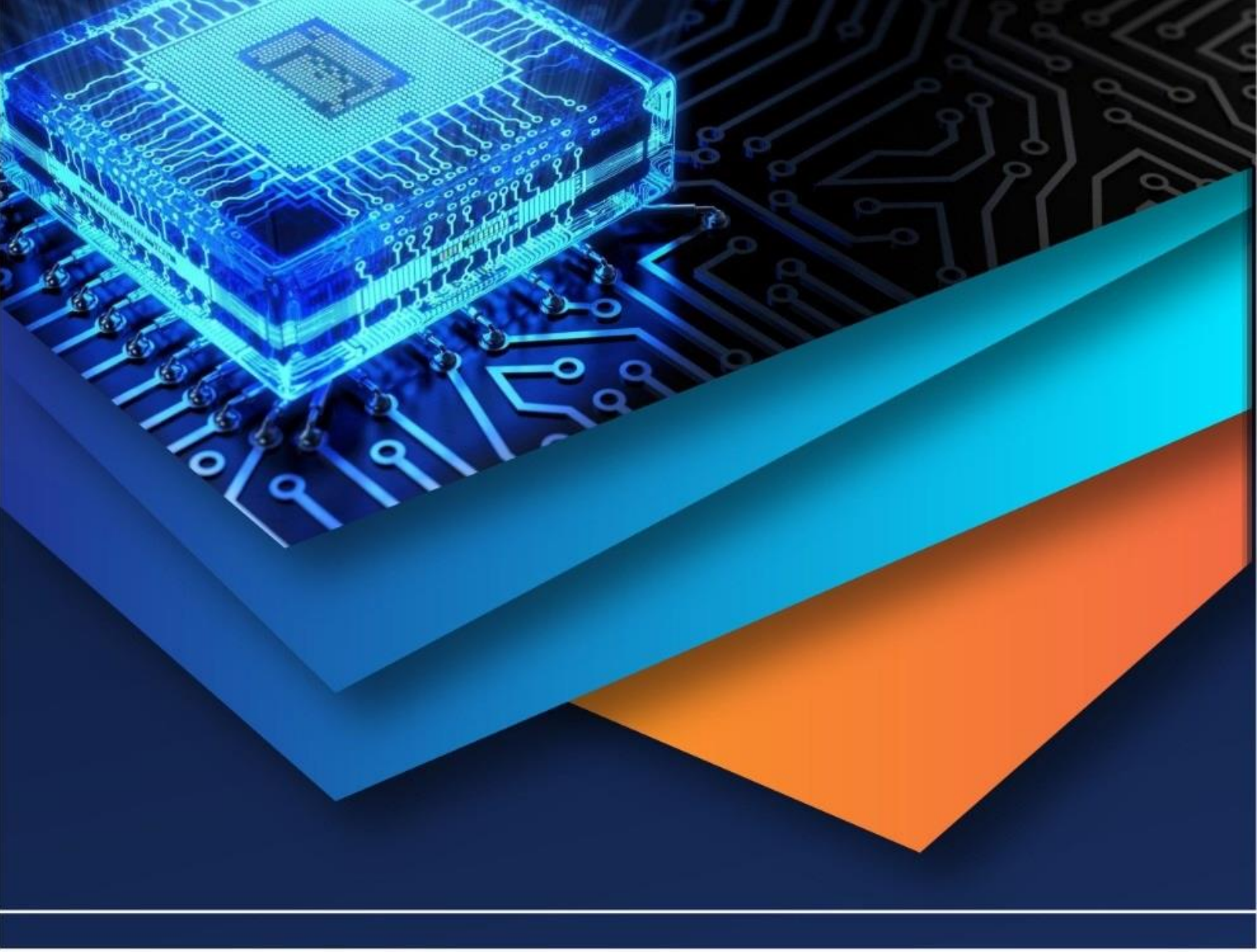
In future focus is more on multilingual sentiment analysis. Given that data pulled from social media are created by users from all over the globe, there is a consequent demand to perform sentiment analysis in more than just one language. The most challenging problem while trying to translate sentiment lexicon in a different language is inflection and conjugation of words applied in some of the languages. Unlike in English, some languages make use of grammatical gender and plural. Following this, verbs, nouns and adjectives are inflected for person or number and verbs are marked for tense. It is important to ensure that for ambiguous words, the appropriate meanings have been translated and included into new lexicon. If the user makes use of the English language script to comment in a native language, then this will be the limitation in the current system. In future, certain training can be provided to the system to detect multilingual slurs and words.

VI. CONCLUSION

The project achieved its stated goal of classifying bullying in comments, which is the key step toward automated systems for analyzing modern social environments that can adversely influence mental health. A lexicon-based approach of sentimental analysis is used to develop a framework for detecting harassment- based cyberbullying. A software is developed by us which identifies bullying by examining the language content of messages, and identify the bullies to provide warning to such users and no comments can be sent or added from their side if repeated. A biased model is presented by us for cyberbullying analysis using sentiwordnet tool with the aim of reducing the reflection and amplification of discriminatory biases in the data while learning the model. The more instances used to train the system, the more efficient the system will become thereby providing a reliable software that can be used to proactively avoid the bullies.

REFERENCES

[1] Tong RM (2001) An operational system for detecting and tracking opinions in on-line discussions. In: Working Notes of the SIGIR Workshop on Operational Text Classification, pp 1–6
[2] Turney P, Littman M (2003) Measuring praise and criticism: inference of semantic orientation from association. ACM Transact Inform Syst J 21(4):315– 346
[3] Esuli A, Sebastiani E (2006) SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of language resources and evaluation (LREC)
[4] Esuli A, Sebastiani E , Baccianella(2010) SentiWordNet 3.0: An enhanced lexical resource for opinion mining. In: Proceedings of language resources and evaluation (LREC)
[5] Elizabeth Denham, Mary Ellen Turpel-Lafond, "Cyberbullying: Empowering children and youth to be safe online and responsible digital citizens", Office of the Information and Privacy Commissioner; Representative for Children and Youth British Columbia, 2015
[6] Wayne MacKay, "Respectful and responsible relationships: There's no App for that", Nova Scotia Task Force on Bullying and Cyberbullying Nova Scotia, 2012.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY