



# **An Effective Approach for Cyberbullying Detection and avoidance**

Divyashree, Vinutha H, Deepashree N S

Student, Dept. of CSE, RVCE, Bangalore, Karnataka, India

Student, Dept. of CSE, RVCE, Bangalore, Karnataka, India

Student, Dept. of CSE, RVCE, Bangalore, Karnataka, India

**ABSTRACT:** The rapid growth of social networking is supplementing the progression of cyberbullying activities. Most of the individuals involved in these activities belong to the younger generations, especially teenagers, who in the worst scenario are at more risk of suicidal attempts. This proposes an effective approach to detect cyber bullying messages from social media through a SVM classifier algorithm. This present ranking algorithm to access highest visited link and also provide age verification before access the particular social media. The experiments show effectiveness of our approach.

**KEYWORDS:** Social Networks; Cyberbullying; Manchine leraing; Link Analysis

## **I. INTRODUCTION**

Cyberbullying (harassment on social networks) is widely recognized as a serious social problem, especially for adolescents. It is as much a threat to the viability of online social networks for youth today as spam once was to email in the early days of the Internet. Parents, being less enlightened about modern technology than young adults, are often in the dark about the horrific experiences of their children. In many cases the children themselves do not tell their parents either under threat from the bully or out of fear of social stigma. Since children often take the pressure alone on their young shoulders, a feeling of depression and self-isolation sets in. In extreme cases it leads one to commit suicide.

The biggest problem regarding cyberbullying is that the age group of the offenders ranges from as young as eight to the legal adult age of eighteen and beyond. In schools, children are bullied on the basis of their physical features, race, age or even their level of knowledge. Even if no actual long-lasting harm may be intended, the victims are often left permanently are not brought to justice and bullying in itself is left unchecked, there is a very high risk of the problem evolving into something more harmful and, maybe, uncontrollable. Hence, the proposed system focused on these issues to develop robust system. The objective of this proposed system are 1) Automatic detection and avoidance of cyberbully attack in internet. 2) Effective age authentication for website browsing and categorizing the links based on age. 3) Effective website filtering in search results based on ranking. 4) Enhanced searching procedure promisingly reduces the effort of user in searching indented websites.

## **II. RELATED WORK**

During recent years, many approaches have been developed for Cyber bullying. The [1] paper presents a research aiming to develop a systematic approach to Online Patrol by automatically spotting suspicious entries and reporting them to PTA members and therefore help them do their job. [1] Present some of the first results of analysis of the inappropriate data collected from unofficial school Web sites. The analysis is performed firstly with an SVM based machine learning method to detect the inappropriate entries. After analysis of the results we perform another analysis of the data, using an affect analysis system to find out how the machine learning model could be improved.

The [2] paper, we used a language-based method of detecting cyberbullying. By recording the percentage of curse and insult words within a post, we were able to correctly identify 78.5% of the posts that contain cyberbullying in a small sample of Formspring data. The [3] paper proposes an alternative approach to cyberbullying: we present a system composed of multiple agents that control users' norm adherence within virtual societies. Being physically present in the

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

virtual society, the agents continuously monitor the behavior of the visitors, communicate with each other to maintain shared beliefs of the visitors' characteristics, and apply punishments and rewards to influence their behavior. Computer software was developed to detect the presence of cyberbullying in online chat conversations.

Rules based on a dictionary of key words are used to classify a window of posts. The overall accuracy is 58.63% was proposed in [3]. The Lexical Syntactic Feature (LSF) architecture to detect offensive content and identify potential offensive users in social media. The overall accuracy is 77.8% in user offensive detection was proposed in [5]. Cross system analyses of the users' behavior - monitoring their reactions in different online environments - can facilitate this process and provide information that could lead to more accurate detection of cyberbullying was proposed in [6].

In [7] paper, method is utilizing a dataset of real world conversations in which each predator question is manually annotated in terms of severity using a numeric label. This paper approach the issue as a sequential data modeling approach, in which the predator's questions are formulated using a Singular Value Decomposition representation. In [8] paper we have devised methods to detect cyberbullying using supervised learning techniques. In [9] present two new hypotheses for feature extraction to detect offensive comments directed towards peers which are perceived more negatively and result in cyberbullying. This preventive approach can provide information about users of social networks and can be used to build monitoring tools to aid finding and stopping potential bullies.

In [10] used a support vector machine model to train a gender-specific text classifier. This paper demonstrated that taking gender-specific language features into account improves the discrimination capacity of a classifier to detect cyberbullying. The detection method can identify the presence of cyberbullying terms and classify cyberbullying activities in social network such as Flaming, Harassment, Racism and Terrorism, using Fuzzy logic and Genetic algorithm was proposed in [11]. This [12] studies negative user behavior in the Ask.fm social network, a popular new site that has led to many cases of cyberbullying, some leading to suicidal behavior.

We examine the occurrence of negative words in Ask.fm's question answer profiles along with the social network of "likes" of questions answers. In existing system we have several algorithms in detection of cyberbully using supervised and semi-supervised approaches. Since the identifying of cyberbullying is difficult, results indicate that it is possible to detect cyberbullying using web content mining techniques. Although a satisfactory level of accuracy has not been reached, the results are promising. The problem presented in existing papers is 1) In existing system even after detecting the cyberbully attack it wouldn't avoid. 2) In most of websites there is no proper validation on person's age. 3) Accuracy of the SVM algorithm is less.

### III. METHODOLOGY

The system consists of 3 major modules. Link filtering, age validation and comments validation. Link filtering is about displaying the search results by filtering according to ranking. When user wish to access the intended site, users age should be validated first. If age is not valid to access the site then site will blocked automatically. After visiting the site the comments validation module will be invoked. Here using cyberbully detection and avoidance algorithm unwanted comments are blocked.

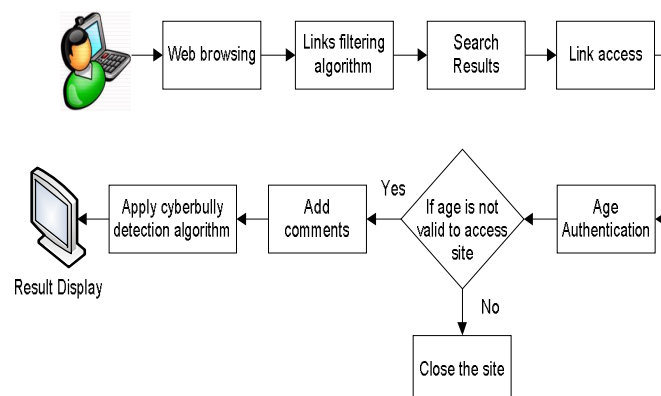


Figure 1: Block Diagram of Proposed System

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Algorithm consists of majorly two phases training phase and testing phase. In testing phase using data mining user comments are extracted from the web content then from each comment lexical and syntactic features are extracted these features are used to classify whether the comment belongs to cyberbully or not. For identification of cyberbully the system need previous knowledge hence system has to be trained first. In training phase input sentences are fed to feature extraction.

Then extracted features are trained using machine learning approach called SVM. Then trained data is stored in knowledge base. In testing phase using this knowledge the system identifies the cyberbully. Algorithm consists of majorly two phases training phase and testing phase. In testing phase using data mining user comments are extracted from the web content then from each comment lexical and syntactic features are extracted these features are used to classify whether the comment belongs to cyberbully or not. For identification of cyberbully the system need previous knowledge hence system has to be trained first. In training phase input sentences are fed to feature extraction. Then extracted features are trained using machine learning approach called SVM. Then trained data is stored in knowledge base. In testing phase using this knowledge the system identifies the cyberbully.

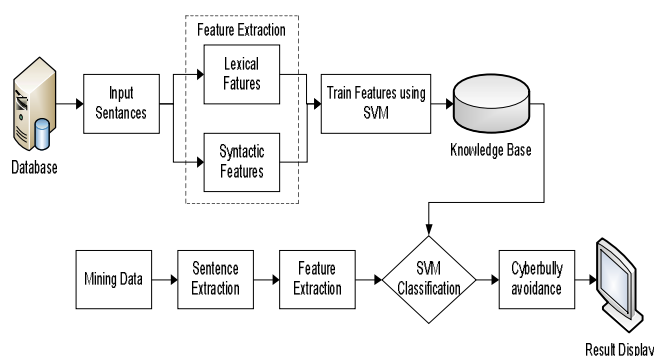


Figure 2: Flow Diagram for Cyberbully Detection and Avoidance

## A. SVM Training

Support vector machines (SVMs) are a set of supervised learning procedures which are best suited for higher dataset level. This paper search for training (Extracting negative words) the knowledge base cyber-bullying is based on the insight that repetitive negative words represent the core of the abusive text posted on profiles. Following the occurrence of negative words led us to many examples of cyber-bullying. Based on these two observations that (i) cyber-bullying is the behavior of posting questions with negative words and (ii) vulnerable targets of cyber-bullying (based on their answers) seem isolated, we sought to build and analyze.

## B. SVM Testing

Although users do not need to understand the underlying theory behind SVM, we introduce the basics necessary for explaining our procedure. A prediction task usually involves separating data into training and testing sets. Each instance in the training set contains the class labels and several the features or observed variables). The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given only the test data attributes.

## IV. IMPLEMENTATION

The implementation provides the step by step brief way of supporting your models in the proposed system. here we discuss about the two main algorithms in our implementation system. one among is the linear SVM model and the other one IS PAGE ranking algorithm. The linear SVM model provides how well the data points lie on the line However below we will produce the implementation of SVM which describes the aspects.

### A) Support vector machine

**Step 1:** identify the data points lying on the line obtained as features and identify the right type of SVM to be used



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

**Step 2:** Identify the right hyper-plane: Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.

**Step 3:** Identify the other hyper-plane and segregating the classes as well.

**Step 4:** Then apply proper kernel so that the data is segregated this is achieved by using kernel trick.

$$k(x, y) = \beta_0 + \beta_1 x \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (14)$$

**Step 5:** by applying the above necessary formula obtain the prediction.

## B) Page ranking algorithm

**Step 1:  $PR(Tn)$**  - Each page has a notion of its own self-importance. That's " $PR(T1)$ " for the first page in the web all the way up to " $PR(Tn)$ " for the last page

**Step 2:  $C(Tn)$**  - Each page spreads its vote out evenly amongst its entire outgoing links. The count, or number, of outgoing links for page 1 is " $C(T1)$ ", " $C(Tn)$ " for page n, and so on for all pages.

**Step 3:  $PR(Tn)/C(Tn)$**  - so if our page (page A) has a back-link from page "n" the share of the vote page A will get is " $PR(Tn)/C(Tn)$ "

**Step 4:** The weighted Page-Rank of pages  $Ti$  is then added up. The outcome of this is that an additional inbound link for page A will always increase page A's Page-Rank.

**Step 5:** Finally, the sum of the weighted Page-Ranks of all pages  $Ti$  is multiplied with a damping factor d which can be set between 0 and 1. Thereby, the extension of Page-Rank benefit for a page by another page linking to it is reduced.

**Step 5:** finally the page ranking formula would be  $PR(A) = (1 - d) + d (PR(T1)/C(T1) + \dots + PR(Tn)/C(Tn))$ .

## V RESULTS AND DISCUSSIONS

Web links are being filter are using page ranking algorithm. Age verification of a person done by authenticating the academic credentials provided by user.

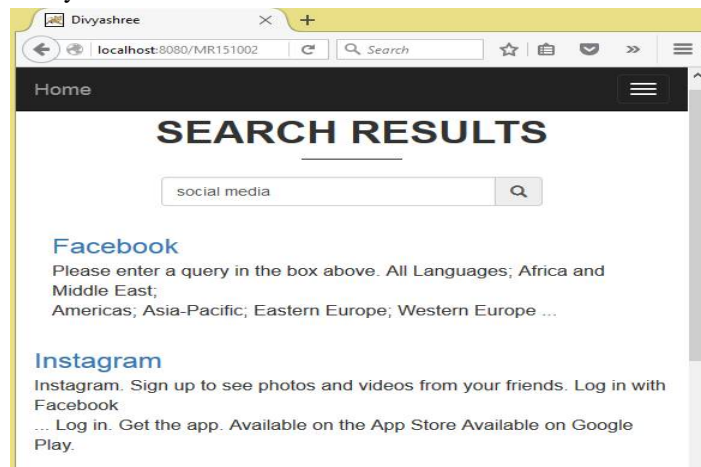


Fig: Link filtering

The first diagram indicates the age verification context which needs the age as obtained from the school credentials thus helping non-adults to prevent from entering the progress. An authenticated user can add a comments to the social media.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

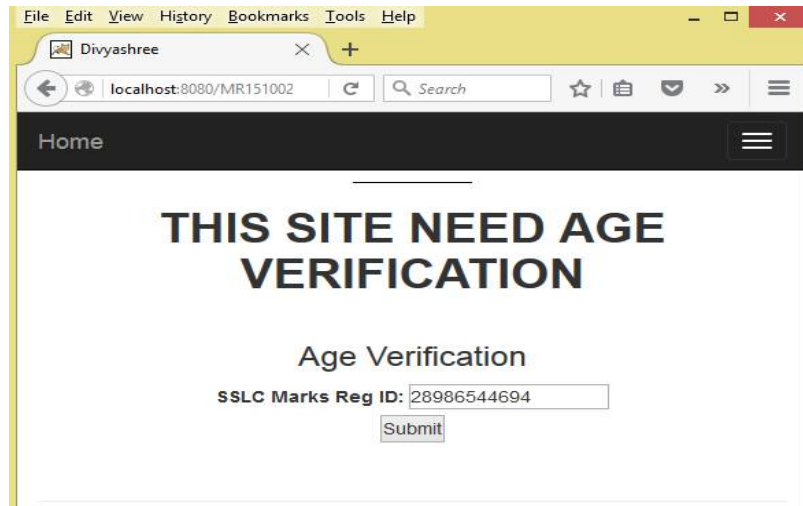


Fig: Age verification site

Once the age verification is obtained the adults are allowed to enter the necessary comments to test the system robustness further. Hence we add a text which contains filthy language and as we proposed we can see that the cyber-bullying is detected which is further not saved.



Figure 4: comment adding window interface

Comments are classified into positive or negative, if positive comments display on the site or if negative comments are deleted from the site. Overall with help of SVM classifier and predator this project achieves 87% accuracy.

## VI CONCLUSION AND FUTURE WORK

In this paper we represented a novel method on the current scenario of cyber-bullying and various methods available for the detection and prevention of cyber harassment. Our concept depends upon the text analysis, the data which is uploaded or text written by any user is first analyzed and after that, we estimate the roles of user, is it a bully? or a victim? and then provide help as required by the user using data mining techniques. Also we will be using a User Identity for registration on the system one will have to provide an identity proof for registering on the system else they will not be able to make an account. With this feature we will be able to check the problem of fake accounts and also cyber-bullying will be controlled to a limit as user accounts will be directly linked to their original identity. This mechanism will be very helpful for our society and the victims. The future work of this paper is to analyze the video comments and detect the positive video or negative video. If in case negative video is avoided from the social media.



ISSN(Online) : 2320-9801  
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

## REFERENCES

- [1] Michal Ptaszynski1 Pawel Dybala1 TatsuakiMatsuba, FumitoMasui ,RafalRzepka Kenji Araki, Machine Learning and Affect Analysis against Cyber-Bullying.
- [2] Kelly Reynolds, April Kontostathis, Lynne Edwards,"Using Machine Learning to Detect Cyberbullying".
- [3] TiborBosse and Sven Stam, "A Normative Agent System to Prevent Cyberbullying".
- [4] Jennifer Bayzick, Jennifer Bayzick, Lynne Edwards"Detecting the Presence of Cyberbullying Using Computer Software".
- [5] Ying Chen, Yilu Zhou, Sencun Zhu1, HengXu, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety"
- [6] MaralDadvar, Franciska de Jong" Cyberbullying Detection:A Step Toward a Safer Internet Yard"
- [7] NektariaPotha, ManolisMaragoudakis "Cyberbullying Detection using Time Series Modeling"
- [8] Vikas S Chavan, Shylaja S S," Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network"
- [9] MaralDadvarDolfTrieschniggFranciska de Jong,"Expert knowledge for automatic detection of bullies in social networks"
- [10] MaralDadvar, Franciska de Jong, Roeland Ordelman, DolfTrieschnigg" Improved Cyberbullying Detection Using Gender Information"
- [11] B.SriNandhini, J.I.Sheeba"Online Social Network Bullying Detection Using Intelligence Techniques"
- [12] HomaHosseinmardi, Richard Han, Amir Ghasemianlangroodi," Towards Understanding Cyberbullying Behavior in a Semi-Anonymous Social Network"
- [13] Sovers Singh Bisht1, Prof. (Dr.) Sanjeev Bansal. Optimization of Web Content Mining with an Improved Clustering Algorithm, International Journal of Emerging Technology and Advanced Engineering
- [14] Maher, D. (2008). Cyberbullying: an ethnographic case study of one Australian upper primary school class. Youth Studies Australia, 27(4), 50-57.
- [15] Patchin, J., &Hinduja, S. "Bullies move beyond the schoolyard; a preliminary look at cyberbullying." Youth violence and juvenile justice. 4:2 (2006). 148- 169.
- [16] Willard, Nancy E. Cyberbullying and Cyberthreats: Responding to the Challenge of Online Social Aggression, Threats, and Distress. Champaign, IL: Research, 2007.
- [17] The changing face of bullying: An empirical comparison between traditional and internet bullying and victimization, Daniell M. Law a , Jennifer D. Shapka
- [18] D. Yin, B. D. Davison, Z. Xue, L. Hong, A. Kontostathis, and L. Edwards, "Detection of Harassment on Web 2.0," In Proceedings of the Content Analysis In The Web 2.0 (CAW2.0) Workshop at WWW2009, 2009
- [19] V. Nahar, S. Al-Maskari, X. Li, and C. Pang. "Semi-supervised Learning for Cyberbullying Detection in Social Networks." In Databases Theory and Applications, pp. 160-171. Springer International Publishing, 2014.
- [20] Mcghee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride, and E. Jakubowski, "Learning to Identify Internet Sexual Predation," International Journal on Electronic Commerce 2011, vol. 15, pp. 103-122, 2011.