



# Robust Search Engine to Improve the Social Security Issue

DIVYASHREE

M. Tech - Student

Computer Science and Engineering,  
R V College of Engineering,  
Bangalore, Karnataka, India

**Abstract:** Cyber-bullying refers to the anonymous calling of any harassment that occurs through the web, mobiles, and other remote devices. Cyber-bullying takes the help of communication technologies to intentionally distort others through hostile behavior such as sending text messages and posting un-sensible or ugly comments on the Internet. The main definition of this phenomenon is derived from the concept of bullying. In this paper, current review of efforts in cyberbullying detection using web content mining techniques is presented [15]. The proposed system effectively overcomes the drawbacks of existing. Also our main contribution is providing a robust search engine that improves the search pattern as well improves the social security issues. Also robust feature extraction improves the accuracy in detecting cyberbully.

**Keywords-** SVM; Cyberbullying; Link filtering

## I. INTRODUCTION

Cyberbullying 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. Current work to tackle this problem has involved social and psychological studies on its prevalence as well as its negative effects on adolescents. Cyber-bullying can happen 24 hours a day, 7 days a week, and reach a kid even when he or she is alone. It can happen any time of the day or night. 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. Cyber-bullying makes a major impact on society; consequently it has become intensive field of research. Although many researchers analyze causes and consequences of cyber-bullying, only few suggest possible solutions for the prevention that include software systems beyond the usual ones based on key words. This paper reviews concepts of possible proposals how Internet may become a safer environment by using web-content mining techniques for detecting and tracking cyber-bullying [16][17]. If the problem of cyber-bullying comes to a solution or at least becomes minimized, social interaction would be much safer for many users on the web, especially for the most vulnerable ones, like teenagers. In brief this paper is organized to filter out the unauthorized content which usually appears in web thus obtained the most frequently used vulnerable words for the training. In addition to this, age authentication, to prevent the people of

certain below age is pinched off from accessing content and rank based search is characterized to make the search more feasible.[14]

## II. RELATED WORK

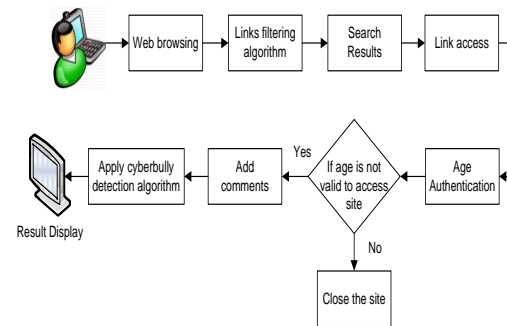
During recent years, many approaches have been developed for Cyber bullying. The [1] paper presents a research aiming to develop a systematic approach to Online Patrol by automatically spotting suspicious entries and reporting them to PTA members and therefore help them do their job. [1] Present some 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 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 prosed 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 are 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.



**Figure 1: Block Diagram of Proposed System**

#### A. Link filtering (Page ranking algorithm)

Link filtering is about displaying the search results by filtering according to ranking. The rank based search mainly works on the concept of the frequency of searchable times that particular website is been viewed by the USER. It keeps the count of the repetition of the websites and that repetition frequency is called the rank. Every time when you encounter a website, the search uses its complexity to store the address of the string rather than its content, so as to reduce the complexity and thus enable fast searching. Rank based search will help you to identify how the net is been used efficiently [22].

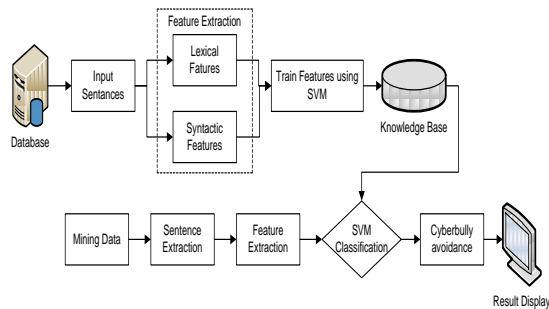
This present a rank based search that keeps track of record, keyword, and query frequency usage and uses this information to modify and improve its performance over time. After the user enters a query, the search engine looks for previously answered queries that are similar to the current query. They are shown to the user and if the user chooses one of the similar queries, he/she is immediately shown the location of the answer in the database. Otherwise, the user is shown a ranked list of records that possibly contain the answer. For each record, its first 80 characters of text are displayed. If the user cannot find an appropriate record or too many records are shown, the user can refine the search. The user will be shown a restricted list of keywords that are related to the query. The user has the option of adding/removing keywords in order to refine the search. Once the search is successful, the search engine will keep track of the record and keywords used to answer the query, and the query itself. This data is used to modify its future performance.

#### B. Age validation

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. With the help of the SSLC marks card roll number the easy to identify the age of user. This module helps the only adults can comments the social websites.

### C. Comments validation

After visiting the site the comments validation module will be invoked. Here using cyberbully detection and avoidance algorithm unwanted comments are blocked. The detection and avoidance of abuse comments can be done with the help of classification algorithm in data mining. Comment validation uses SVM for predicting the events. SVM model consists of two subsidies one is training and other is classify or prediction model.



**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 [18]. 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 [23].

### IV. SVM PREDICTION

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[23][24].

### V. IMPLEMENTATION

The implementation provides the step by step brief way of supporting your models in the proposed system. Here in this paper 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. This paper clearly mention the implementation steps below.

Step 1. Create Dynamic Web Project in eclipse.

Step 2. Install MySQL server and SQLyog.

Step 3. Install Tomcat server.

Step 4. Prepare the training set of the data using Matlab with help of SVM classifier.

Step 5. Rank based algorithm take the input as search query and provide the output as filtering links, based on the number of usage of that particular link.

Step 6. Select the link before access the link this paper provide the age authentication.

Step 7. The age authentication algorithm take input as registration number of academic certificate.

Step 8 . The registration number is used get the valid age for user.

Step 9.The age proof algorithm compare if the age is greater than or equal than allow to access the social media link otherwise leave the page.

Step 10. User can post comments on the social media page, these comments takes input to the robust search engine.

Step 11. The comments are classified into positive and negative comments with the help of SVM classifier algorithm.

Step 12. The comments are positive add to the site, if negative comments are deleted from the site.

### VI. 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.An authenticated user can add a comments to the social media. Comments are classified into positive or negative, if positive comments display on the site or if negative comments deleted from the site.Overall with help of SVM classifier and predator this project achieve 87% accuracy.

**Table: Result Analysis**

Comments	Excepted Test cases		Obtain output		Remarks
	Positive comments	Negative comments	Positive comments	Negative comments	
Wow, looking good	Yes		Yes		Display on the site
Not looking good		Yes		Yes	Deleted from the site
Not nice	S	Yes		Yes	Deleted from the site
Not photogenic face		Yes		Yes	Deleted from the site
Not bad	Yes			Yes	It should corrected

## VII. 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 analysis the video comments and detects the positive video or negative video. If in case negative video avoided from the social media.

## VIII. ACKNOWLEDGEMENTS

The author thanks Dr. S. Sridhar, Professor and Director, RV Cognitive & Central Computing, R.V. College of Engineering, Bangalore, India for communicating this article to this Journal for publication. The author also thanks Dr. G. Shobha, Professor and Head, Department of CSE, R. V. College of Engineering, Bangalore, India for the support given.

## IX. REFERENCES

[1]. Michal Ptaszynski1 Pawel Dybala1 Tatsuaki Matsuba, Fumito Masui , Rafal Rzepka Kenji Araki, "Machine Learning and Affect Analysis against Cyber-Bullying," AISB 2010 convention, pp. 1-5 29<sup>th</sup> march-1<sup>st</sup> April.

[2]. Kelly Reynolds, April Kontosthatis and Lynne Edwards "Using machine learning to detect cyberbullying", National Science

Foundation, Grant No. 0916152, pp. 1-4, Dec 2011.

[3]. Tibor Bosse and Sven Stam, "A Normative Agent System to Prevent Cyberbullying", in IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2011 © IEEE. Doi: 10.1109.

[4]. Jennifer Bayzick, Jennifer Bayzick, Lynne Edwards "Detecting the Presence of Cyberbullying Using Computer Software". 2011, 44(9), 93-96.

[5]. Ying Chen, Yilu Zhou, Sencun Zhu1, Heng Xu, "Detecting Offensive Language in Social Media to Protect Adolescent Online Safety", ASE/IEEE 2012, 978-0-7695-4848-7/12.

[6]. Maral Dadvar, Franciska de Jong "Cyberbullying Detection: A Step Toward a Safer Internet Yard" WWW 2012-PhD Symposium, IW3C2, 2012, pp. 1-4.

[7]. Nektaria Potha, Manolis Maragoudakis "Cyberbullying Detection using Time Series Modeling" IEEE international conference on data mining workshop, 978-1-4274-9/14.

[8]. Vikas S Chavan, Shylaja S S, "Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network", IEEE-2015, 978-1-4799-8792-4/15.

[9]. Maral Dadvar Dolf Trieschnigg Franciska de Jong, "Expert knowledge for automatic detection of bullies in social networks" EU COST Action IS0801 on Cyberbullying, 2014.

[10]. M. Dadvar, F. d. Jong, R. Ordelman, and D. Trieschnigg, "Improved cyberbullying detection using gender information," In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR 2012), pp. 23- 25, February 2012.

- [11]. B.Sri Nandhini, J.I.Sheeba”Online Social Network Bullying Detection Using Intelligence Techniques”,ICACTA-2015,pp.485-492
- [12]. Homa Hosseinmardi, Richard Han, Amir Ghasemianlangroodi,” Towards Understanding Cyberbullying Behavior in a Semi-Anonymous Social Network”,IEEE/ASONAM-2014,978-1-4799-5877-1/14.
- [13]. H. Lieberman, K. Dinakar & B. Jones (2011). Let’s Gang Up on Cyberbullying. *Computer*, 2011, 44(9), 93-96.
- [14]. A. Kovacevic, D. Nikolic, “Automatic detection of cyberbullying to make Internet a safer environment”. *Handbook of Research on Digital Crime*”, in *Handbook of Research on Digital Crime, Cyberspace Security, and Information Assurance* (pp. 1-675).editores Cruz-Cunha, M. M., & Portela, I. M., Hershey, PA: IGI Global, 2015
- [15]. Paridhi Singhal and Ashish Bansal “Improved Textual Cyberbullying Detection Using Data Mining,International Journal of Information and Computation Technology.ISSN 0974-2239 Volume 3, Number 6 (2013), pp. 569-576
- [16]. KARTHIK DINAKAR, BIRAGO JONES, CATHERINE HAVASI, HENRY LIEBERMAN, and ROSALIND PICARD, “Common Sense Reasoning for Detection, Prevention,and Mitigation of Cyberbullying”*ACM Transactions on Interactive Intelligent Systems*, Vol. 2, No. 3, Article 18, Publication date: September 2012.
- [17]. P. K. Smith, J. Mahdavi, M. Carvalho, S. Fisher, S. Russell, and N. Tippett, "Cyberbullying: Its nature and impact in secondary schoolpupils," *Journal of Child Psychology & Psychiatry*, vol. 49, pp. 376-385, 2008.
- [18]. 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.
- [19]. 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.
- [20]. Paridhi Singhal and Ashish Bansal, “Improved Textual Cyberbullying Detection Using Data Mining”, ISSN 0974-pp 2239 Volume 3, Number 6, *International Journal of Information and Computation Technology*, pp. 569-576, 2013.
- [21]. Vinita Nahar, Xue Li and Chaoyi Pang. “An effective approach for cyberbullying detection”, Volume 3, Issue 5, *Communications in Information Science and Management Engineering*, pp 238-247, May 2013.
- [22]. Ramesh Prajapati, “A Survey Paper on Hyperlink-Induced Topic Search (HITS) Algorithms for Web Mining”, Volume 1, Issue 2, *International Journal of Engineering Research and Technology*, pp.13-20, 2012.
- [23]. Robert Burbidge, Bernard Buxton “An Introduction to Support Vector Machines for Data Mining”, IEEE, PP. 1357-1360, 2008.
- [24]. Alexander Statnikov\*, Douglas Hardin, Isabelle Guyon†, Constantin F. Aliferis “A Gentle Introduction to Support Vector Machines in Biomedicine”, IEEE, PP. 833-840, 2009.