A DYNAMIC COGNITIVE SYSTEM FOR AUTOMATIC DETECTION AND PREVENTION OF CYBER-BULLYING ATTACKS

NaliniPriya. G[1] and Asswini. M[2]
[1]Department of Information Technology, Saveetha Engineering College, Anna University, Chennai, India
[2]Department of Software Engineering, Saveetha Engineering College, Chennai, India
E-Mail: Chennai.nalini.anbu@gmail.com

ABSTRACT

The phenomenal growth of the social networking sites has swept over the communication world. The rising popularity of the social networking sites have also contributed to the rise in offensive behaviours, giving birth to one of the most crucial problem called cyber-bullying. Most of the social networking users would have encountered a worst e-day experience .The victims of cyber-bullying, widely being the adolescents, suffer deep scars which has led to suicidal attempts in many cases. Detection of cyber-bullying has also been a challenging issue for the researchers. A few automated methods have been developed which mainly rely on textual features. This work aims to improvise the detection of cyber-bullying by developing a real-time application combining user based textual features along with the social networking features such as number of circles, number of contacts in the friend's list, bonding with the friends.

Keywords: cyber-bullying, textual features, cyber-bullying victims, cyber crime.

1. INTRODUCTION

An online social network (OSN) [1] shall be defined as the use of dedicated websites and applications that allow the users to interact with other users, or to find people with similar interests to one's own. With the advent of web 2.0, the social networks gained much popularity ever since the launch of the first social network SixDegrees.com in 1997[1].The social networking sites enable the people worldwide in stay in touch with each other irrespective of ages. The children in special are introduced to a bad world of worst experiences and harassments. The users of the social networking sites might be unaware of various vulnerable attacks hosted by the attackers in these sites.

The involvement of adolescents in the online social networking sites has increased as the result of technology gap between the parents and technically more updated children [3]. The children engage themselves more into the internet and thus are prompted to share their photos and personal details which in turn turn out to be a serious social problem. In the recent years the professional as well as personal tools for communication have witnessed dramatic changes [13].The social networks instigate the young people into a world of fatal threats such as cyber-bullying. Cyber-bullying is expanding as a crucial problem over the internet.

Cyber-bullying may be defined as an intentional and aggressive act accomplished over a period of time by an individual or a group of individuals through an electronic medium over feeble victim who cannot shield themselves [4]. Cyber-bullying can be mainly classified as follows. Cyber-bullying is basically carried out using two tools namely the internet sources and the mobile devices [5].

Various options in the internet sources and mobile devices such as messaging services, cameras, social networking sites, Dash boards etc serve as a medium to carry out these kinds of attacks on the victims. The person who is being bullied is called the victim. The cyber-bullying attacks may sometimes may more intense that the victims are forced towards suicidal attempts. The cyber-bullying attacks can be put forth over a wide range of users [6]. Bullying can be of several types such as physical, social, sexual, verbal, psychological etc [14]. However, the targeted victims are adolescents [8] in general according to the reports given by recent studies. Figure-1 gives out a brief classification of the cyber-bullying attacks. The cyber-bullying attacks are basically carried out through two sources namely the internet sources and the mobile networking sources. With the cloud nine popularity of the internet more number of users is attracted to it and thereby the number of social networking users also keep increasing. As far as mobile is concerned, it has become an essential part of everyone's day to day life. Also the users find it easy to install apps in their mobile and increase their contact network. The market also offers the users with wide range of applications free of cost.

Another form of cyber-bullying is called as cyber-grooming where adults employ the use of electronic medium to sexually seduce the adolescents [7]. Detecting such kinds of activities in the social networks becomes a difficult task because of lack of advancements in image level detection. According to research works [10] [11], Cyber-bullying can be classified into cyber-stalking, exclusion, impersonation, outing and trickery, harassment, denigration and flaming. Cyber-stalking [9] is a kind of cyber-bullying where victims is harassed or threaten over a period of time with the help of social media. The Figure-2 portraits the different types of cyber-bullying mentioned above. Flaming [11] is when a person uses aggressive and angry words to harass the victim. Harassment is when inappropriate or offensive language is used over a prolonged period of time. Denigration [12] is the spread of gossips and negative comment about a person with the intention of purely harassing the person.

Outing and trickery is carried out by disclosing a person's private information or tricking them to disclose the private information themselves in order to dishonour them in the social media. Impersonation is when a person steels the identity of another person and post offensive contents with the intention of defaming a person.

Exclusion [12] is when a person is removed from a social circle wantedly in order to hurt the sentiments of the person. Happy slapping and sexting are also a kind of cyber-bullying that wide spread over the web. Happy slapping consists of videos that depict conflicts and scuffles between the adolescents and their content intentionally over the web. Sexting is the circulation of photos of themselves or of others over the internet.

With the increasing number of users as well as the attackers, controlling these kinds of vulnerable attacks becomes a very difficult task for the online social networking providers. Cyber-bullying detection is on its initial stage as it is a much complicated task. The existing works mostly concentrate on the prevention of offensive words and also affect the privacy of the users. The proposed work aims to contribute to the detection of cyber-bullying attacks and also proposes a effective follow up strategy. Section III gives out a brief description about the proposed work. The section II describes the various related works available. Section IV describes the implementation and results. The section V gives out the conclusion.

2. RELATED WORKS

The cyber-bullying attacks occur frequently on the social networking site and lead to severe physical, emotional and mental abuses [24].the predator can approach the victim in the following way.

a) Connect to the victim through friend request or arouse the victim to send friend request.

b) Lure the victim into a delusive relationship.

c) Initiate abusive or sexual relationship.

Though the users work online, the consequences are experienced when the even when the user leaves the internet world. The cyber-bullying victims undergo humiliations over 24x7 in the web. These attacks are similar to old wine in new bottle types. They have always existed in the society. At present their impact as widened with the pervasiveness of the social networks. Many cases have been reported in a decade were adolescents have been projected to these fatal attacks [25].

The children feel reluctant to share these information with their parents or well wishers as the fear of losing the mobile or the internet connectivity engulfs them. It is essential to create awareness among young children as well as the adults to prevent or safeguard oneself from the attack. Parents and teacher play a main role in educating the children about these attacks.

Many governmental and nongovernmental organisations have opened up to stand against the cyber-bullying attacks. Child exploitation and online protection centre (CEOP) [26] is a professional organisation that helps the children to report about abusive conversations invoked on them by adults. Many tools such as net nanny, mobicip, near parent, my mobile watchdog etc which mainly offer parental over the children where the parents are allowed to watch the overall content or the data shared. This in turn questions the privacy of the user and humiliates the user's self-reliance. It motivates the children to de activate such tools since the children are more technically advanced than the parents.

Previous works have explored the text mining in cyber-bullying vastly. The number, values of foul words were used as features by Reynolds *et al* [19] to determine the cyber-bullying conversations. User based features such as user's activity log, content based features and cyber-bullying features were used for the purpose of cyber-bullying detection by Dadvar et al [18]. The gender factors such as male specific feature sets and female specific feature sets were also used for the purpose of cyber-bullying detection by the experts. Filters are used to filter out offensive contents. However the attackers find way to bypass the servers that maintain the filters [9].

3. PROPOSED WORK

The proposed work highlights the areas of importance that are necessary to create an automated system for the detection and prevention of cyber-bullying attacks. The previous works have concentrated only on the textual features of the social networking sites. The Table-1 gives a comparison of the existing works and the proposed approach in cyber-bullying. The proposed work combines the efficiency of both the textual features as well as the social networking features for the detection and prevention of cyber-bullying attack. The Figure-3 depicts the proposed system architecture.

3.1 Offensive posts detection module

The stop word removal [20] technique is employed in the process of detection of offensive words. The stop word removal technique enables the filtering of meaningless and offensive words. The list of offensive words can be found in the rejection list. The abusive words such as 'bitch','asshole','f**k' are obtained from the users as well as retrieved from noswearing.com .Blacklisting such abusive words enables the blocking of posts containing such words.

Based on age limits the usage of offensive words is restricted. When the user is a major that is a person whose age is 18 and above and attempts to post offensive words, they are warned before contents are being posted. The users themselves shall also report these offensive words directly.

The data collection subsystem is responsible for all collecting all types of data namely text data, image data and the social media data. The rule manager consists of the set of offensive words that cannot be posted. The rule manager consists of a set of restricted words that can be blocked whenever the user posts such words. Whenever an user attempts to post illegal contents the decision manager warns the user in prior.

If the predator attempts to posts even after the warning, the users are facilitated to either block the sender or seek the help of the trusted contact. The trusted contacts help the victims to recover from these kinds of vulnerable attacks. For those users feel hesitant to share their bad e-day experiences, the system provides the way of automatic alarming to their trusted contacts whenever the threshold level of the offensive words in a post increases or when the usage of abusive words from a particular contact keeps extending.

3.2 Social structure analysis module

An efficient age classification [21] is performed as the first step in the social structure analysis module. The users are required to submit a government approved valid age proof as a mandatory resource to create a profile in the social network. This enables the filtering of contents based on the age group. When a person suddenly blocks another person or when a group of people withdraw themselves from a person, this could be concluded as an indication of cyber-bullying attack according to the psychological studies on cyber-bullying.

The social structure takes into account the communication between different users. The temporal data module is used to compute the temporal data changes to detect abnormal patterns. The temporal data management module keeps into account the temporal activities of the user. The ego networks are employed for the purpose of computing the communication relationship.

Figure-4.A 1 ego-network around a node v depicted as marked in red. The ego node is represented by a triangle and its neighbours are represented by squares. The ego networks are generally used to compute the communication between the users. The centre user is called as the “ego”. The other users whom the ego is connected with are called its neighbours. The arrows establish the relationship between the ego node and its neighbour nodes. It can be represented as a graph G (V, E) where v is the number of nodes and E is the number of edges. The number of nodes and edges are used to determine the size and the connectivity of the user with the other users. The links determine the communication channel between different users. Links can be employed to determine the communication flow between two users

3.3 Follow-up strategy module

The present systems lack an effective follow-up strategy. Experts feel that follow-up strategy is most important in the detection and prevention of cyber-bullying attacks. They allow the user only to report to the online social networking providers. According to the cyber-bullying critics this follow-up strategy [22] is inefficient and does not leave a good impact on the cyber bullied victim. The proposed system aims to safeguard the privacy of the user and thus enhancing the user’s self reliance. The proposed approach enables the detection of offensive posts and provides an alert signal to the trusted contacts [23] of the victim and thus aims to empower the victims. Figure-5 shows user uploading a offensive post. It consists of a title and a description. The users can the text a crisp title so that the receiver could easily prioritize their messages.

The system is capable of evoking alert signals to the trusted contacts when the threshold level of the offensive message count is exceeded. The threshold is set to 5 message counts. When the illegal posts or messages are posted on the user’s page for more than five times, the system automatically signals the trusted contacts of the user seeking their aid to help the victim being bullied. The trusted contact may be the user’s close friends, relatives, colleague or parents. The trusted contacts are selected by the users themselves. And for minors it is essential have at least one adult in the trusted contact list. Figure-6 shows the restriction of offensive post when the user attempts to post them. The user is warned before they upload the text contain offensive words. However the user is allowed to post it on their own risk only when the receiver is classified as a major.

This helps the victim to seek proper advice, help and support from the people whom they believe and need in worst situations. This shall prevent the victim from breaking down and being depressed. Thus ensures that the victim is not subjected to any suicidal attempts. The victim experiences a crucial period after the cyber-bullying attack. The victims can have a moral support throughout the recovery period.

4. IMPLEMENTATION AND RESULT

Application is realised using Microsoft .Net framework with ASP.net. Microsoft .NET comprises of a set of Microsoft software technologies for efficient building and integration of XML based Web services, Microsoft Windows applications, and Web solutions. The .NET Framework is a language-equitable platform were programs can be written easily and ensures secure inter-operability of the programs. The application developed ensures the safety of the victims from the online intruders. Also it provides more support and mental strength to the victims being bullied with the help of trusted contacts.

The users are required to upload one of their government approved age proof as identity for the creation of their account. This submission of age proof is set to be a mandatory field in the creation of account. The admin verifies age submitted proof and gives permission to the user to create account. This enables the classification of the users as per the age group and also facilitates the segregation of contents as per age. This in turn ensures the online security of the adolescents. The application also prevents the users from uploading offensive words. Thus the users can rely on the application and have a pleasant e-day experience.

Table-2 gives out the security impact of online activities with regards to the cyber-bullying detection features. The security impact is classified as low, medium and high according to the level of protection offered to the user by the application of social networking features. A comparison of efficiency of existing works and the proposed work can be depicted in the graphical format. The table provides the data on the cyber-bullying features and the security impact of user's online activity. The chart shows that the security impact is high when the textual and the social networking features are combined.

5. CONCLUSIONS

This paper takes a step ahead with the employment of social relationship in the field of cyber-bullying detection. Text analysis has been the predominant fields of research in cyber bulling. The proposed work proves that the accuracy of bullying detection shall be increased with the advent of combining two or more cyber-bullying detection features. Human behaviour analysis and image analysis are the key factors in realising precision in cyber-bullying detection in the mere future. The involvement of social research groups and young generation plays a vital part in the prediction and reduction of these kinds of vulnerable attacks.

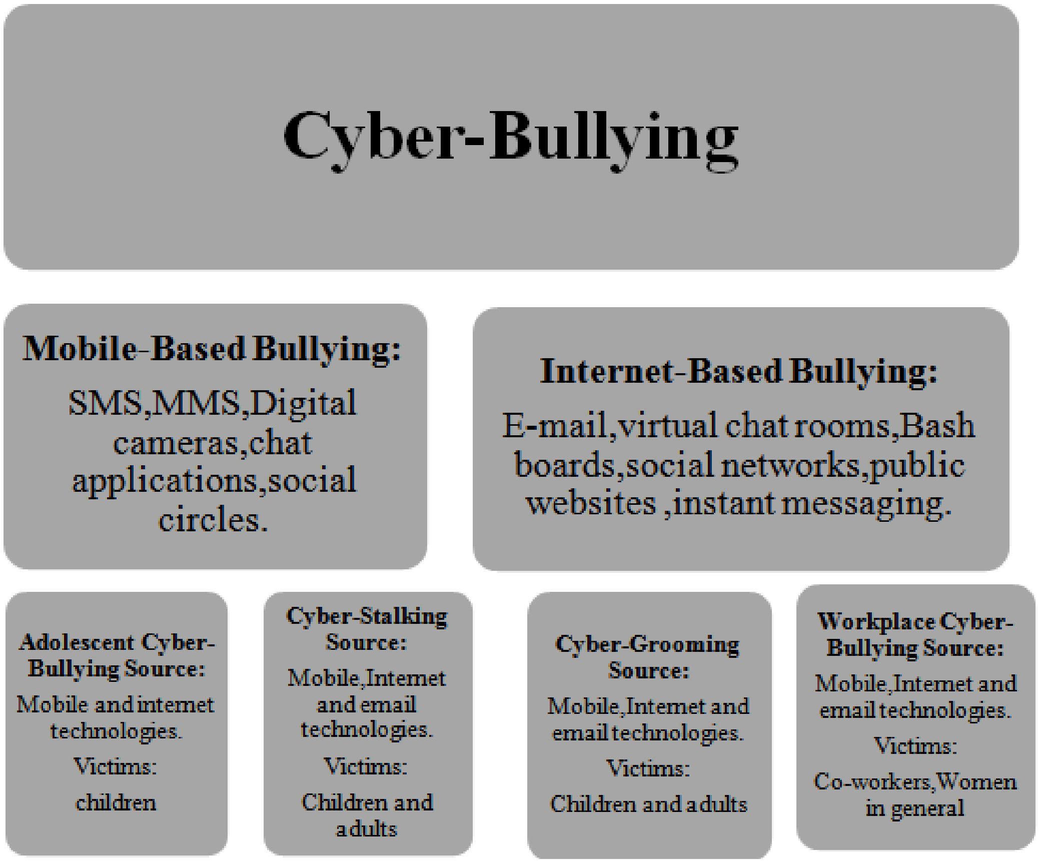

Cyber-Bullying
Mobile-Based Bullying:
SMS,MMS,Digital cameras,chat applications,social circles.
Internet-Based Bullying:
E-mail,virtual chat rooms,Bash boards,social networks,public websites ,instant messaging.
Adolescent Cyber-Bullying Source:
Mobile and internet technologies.
Victims:
children
Cyber-Stalking Source:
Mobile,Internet and email technologies.
Victims:
Children and adults
Cyber-Grooming Source:
Mobile,Internet and email technologies.
Victims:
Children and adults
Workplace Cyber-Bullying Source:
Mobile,Internet and email technologies.
Victims:
Co-workers,Women in general

Figure-1. Classification of cyber-bullying attacks.

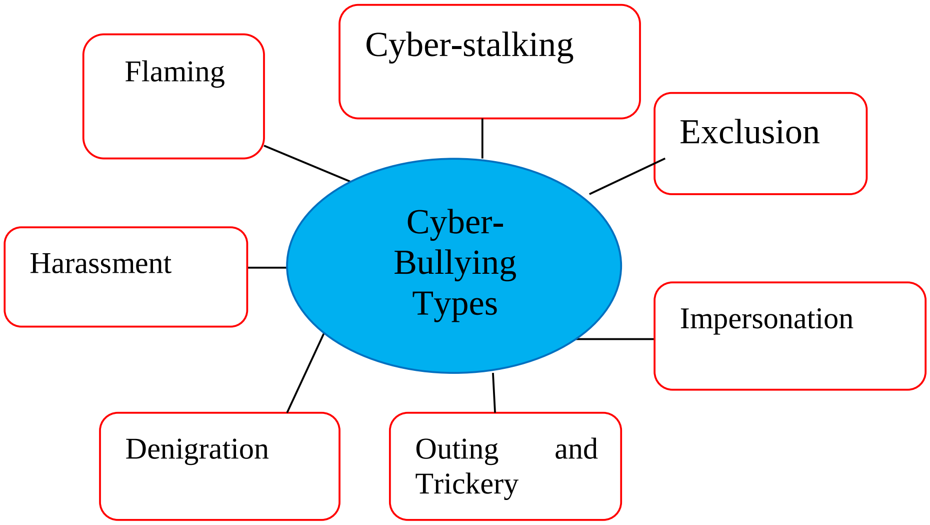


Figure-2. Types of cyber-bullying.

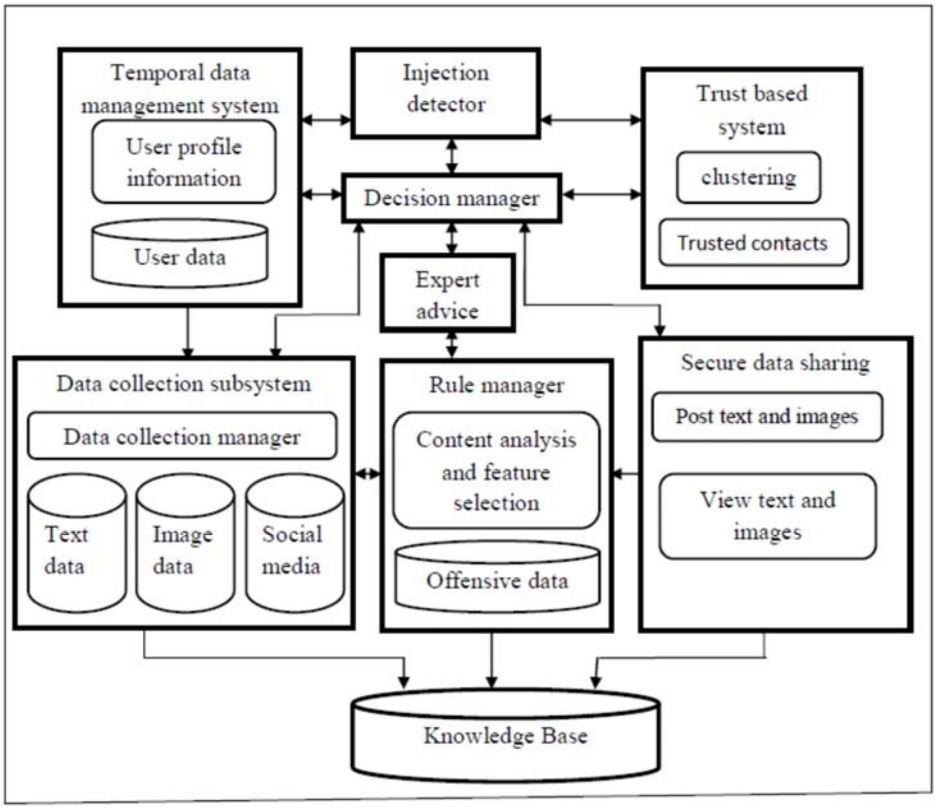


Figure-3. The proposed system architecture.

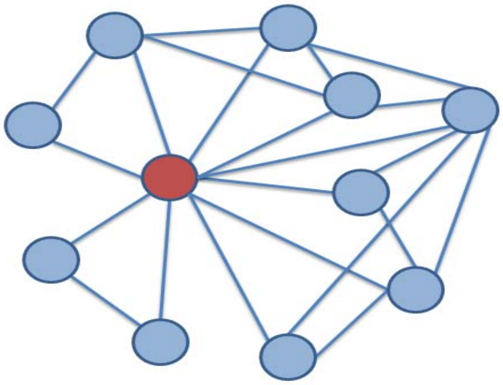
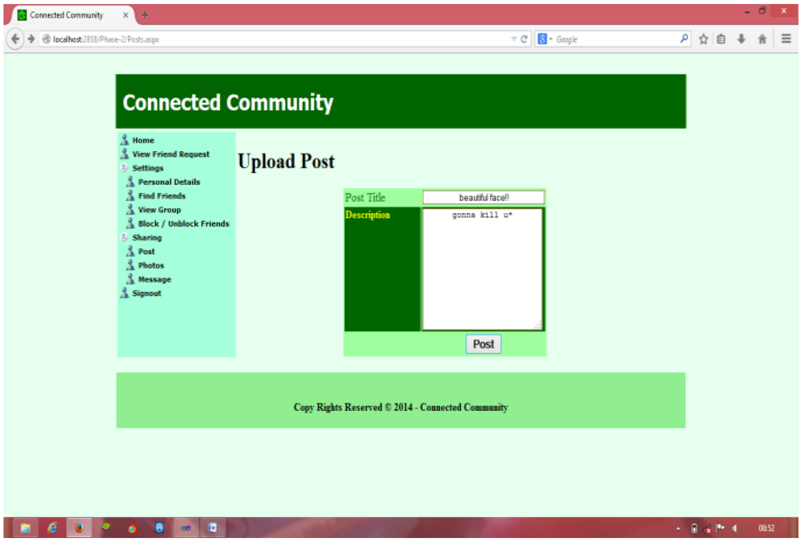


Figure-4. The 1 Ego network “v” marked in red and its neighbours marked in blue.



Connected Community

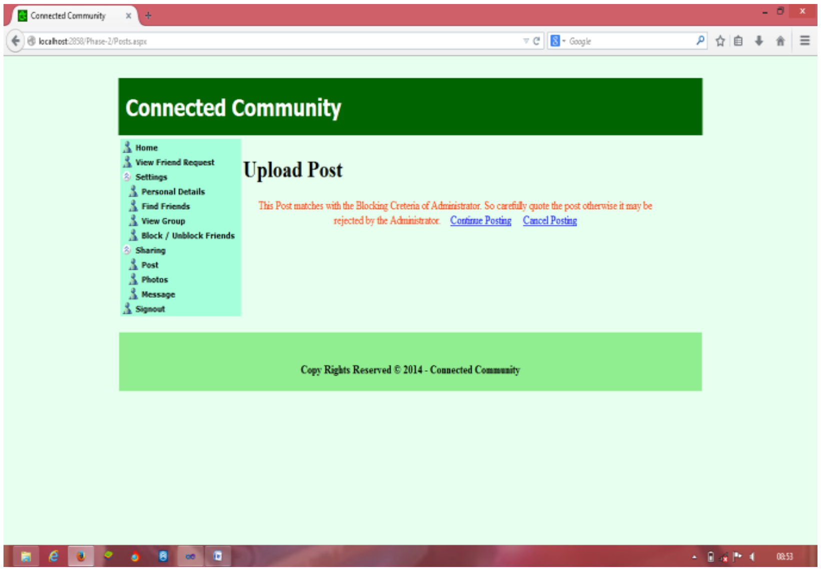
Upload Post

Post Title: beautiful face!!

Description: gonna kill u*

Copy Rights Reserved © 2014 - Connected Community

Figure-5. User uploads a post.



Connected Community

Upload Post

This Post matches with the Blocking Creteria of Administrator. So carefully quote the post otherwise it may be rejected by the Administrator. Continue Posting Cancel Posting

Copy Rights Reserved © 2014 - Connected Community

Figure-6. Restriction of offensive post.

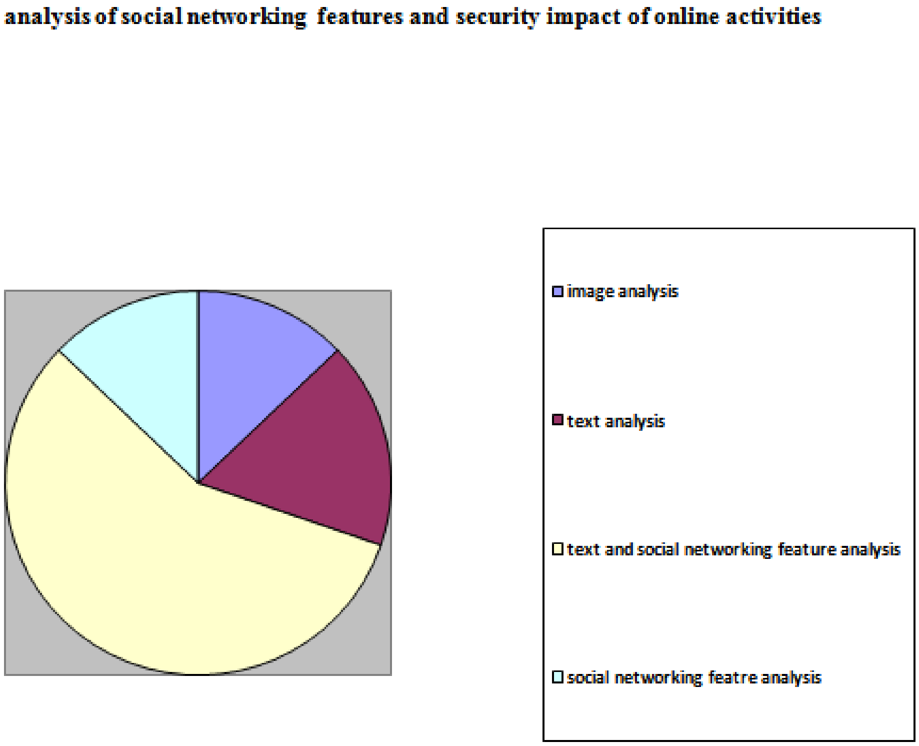


Figure-7. A comparison chart representing percentage of security impact over various social networking features.

Table-1. A comparison of present works and the proposed work.

The work	Textual features	Social network features
Dinakar *et al*. [11]	Yes	No
Nahar *et al*. [17]	Yes	No
Reynolds *et al*.	Yes	No
Dadvar *et al*. [18]	Yes	No
Net nanny [15]	Yes	No
D. Yin *et al*. [16]	Yes	No
Proposed work	Yes	Yes

Table-2. security impact of online social networking activities with respect to the cyber-bullying detection features.

Cyber-bullying detection features	Security impact of online activities
Text	Low
Image	Low
Social networking features	Low
User demography	Low
Text and social networking feature	High

REFERENCES

[1] D. Boyd and N.B. Ellison, "Social Networks Sites: Definition, History, and Scholarship", Computer-Mediated commun.vol.no:13, 2007.

[2] Y. Altshuler, M. Fire, E. Shmueli, Y. Elovici, A. Bruckstein, A. S. Pentland, and D. Lazer, " The social amplifier - reaction of human communities to emergencies", Journal of Statistical physics vol. no. 152(3): 399- 418, 2013.

[3] Great Britain: Parliament: House of Commons: Welsh Affairs Committee, House of Commons, Digital inclusion in Wales, Thirteen report of session 2008-2009, report, p. 135.

[4] Espelage, D. L., Swearer, S. M., “Research on school bullying and victimization: What have we learned and where do we go from here? In: School Psychology Review”, vol. 32, no. 3, pp. 365–383. 2003.

[5] E. Menesini, Smith, P K and Zukauskiene, R, "COST ACTION IS0801: Cyberbullying: Coping with negative and enhancing positive uses of new technologies, in relationships in educational settings”, Mykolas Romeris University Publishing Center, 2009.

[6] Saferinternet.at, “Sex and Violence in digital Media - Prevention, Help& Counselling”, Tech. Rep., 2012.

[7] S.Wachs, K. D. Wolf, and C. Pan, “Cyber-grooming: risk factors, coping strategies and associations with cyberbullying”, Psicothema, vol. 24, pp. 628-633, 2012.

[8] B. Lobe, S. Livingstone, K. lafsson, and H. Vodeb, “Cross-national comparison of risks and safety on the internet: Initial analysis from the EU kids online survey of European children,” EU Kids Online, Tech. Rep., 2011.

[9] W. Heirman and M. Walrave, “Predicting adolescent perpetration in cyberbullying: An application of the theory of planned behaviour,” Psicothema, vol. 24, pp. 614-620, 2012.

[10] Frank.E, “Data Mining: Practical machine learning tools and techniques, 2nd Edition”, Morgan Kaufmann, 2005.

[11] Dinakar, K; Reichart, R.; Lieberman, H, “Modeling the Detection of Textual Cyberbullying”, Massachusetts Institute of Technology, 2011.

[12] Willard, N. E, “Cyber bullying and Cyber threats: Responding to the Challenge of Online Social Aggression, Threats, and Distress” Champaign, IL: Research, 2007.

[13] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr, “Social media and mobile internet use among teens and young adults”, Pew Internet and American Life Project 20th September 2011.

[14] P. K. Smith, Madsen, K., and Moody, J., “What causes the age decline in reports of being bullied in school? Towards a developmental analysis of risks of being bullied”, Educational Research, vol. 41, pp.267-285, 1999.

[15] Content Watch, Inc., “Net nanny social,” http://www.netnanny.com/, Accessed January 31st, 2014.

[16] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, “Detection of harassment on web 2.0,” Proceedings of the Content Analysis in the WEB”, vol. 2, 2009.

[17] V. Nahar, X. Li and C. Pang, “An effective approach for cyberbullying detection”, Communications in Information Science and Management Engineering, 3(5): 238-247, 2013.

[18] M. Dadvar, D. Trieschnigg, R. Ordelman and F. de Jong, “Improving cyberbullying detection with user context”, In Advances in Information Retrieval, pages 693-696. Springer, 2013.

[19] K. Reynolds, A. Kontostathis and L. Edwards, “Using machine learning to detect cyberbullying” In Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on, volume 2, pp. 241-244. IEEE, 2011.

[20] X. Hu and H. Liu, “Text analytics in social media,” in Social Network Data Analytics, C. C. Aggarwal and C. Zhai, Eds. Springer US, 2012, pp. 385-414.

[21] Y. Fu, G. Guo, and T. Huang, “Age synthesis and estimation via faces: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 11, pp. 1955–1976, 2010.

[22] Kathleen Van Royen, Karolien Poels, Walter Daelemans, Heidi Vandebosch, “Automatic monitoring of cyberbullying on social networking sites: From technological feasibility to desirability”, Elsevier, 2014.

[23] T. Grandison and M. Sloman, “A survey of trust in internet applications,”IEEE Commun, vol. 3, no. 4, pp. 2-16, Oct. 2000.

[24] Hinduja, S., Patchin, J., “Offline consequences of online victimization”, Elsevier, 6 (3): 89-112, 2007.

[25] http://nobullying.com/six-unforgettable-cyber-bullying-cases.'accessed on 22/10/2014.

[26] Child Exploitation and Online Protection Centre, “Child exploitation and online protection centre,” http://ceop.police.uk/, 2013, Accessed September 10th, 2014.

[27] K. Reynolds, A. Kontostathis and L. Edwards, “Using Machine Learning to Detect Cyber bullying”, In: Proceedings of the 2011 10th International Conference on Machine Learning and Applications Workshops (ICMLA 2011), vol. 2, pp. 241-244, December 2011.