International Journal of Engineering & Technology, 7 (3.24) (2018) 497-500

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Detection of Cyber Harassment

P.Asha[1]*, A. Lakshmi Sai Prasanna[2], K.Vennela[3]

[1]*Asst. Prof, Sathyabama Institute of Science and Technology, Chennai.*
[2]*Student, Sathyabama Institute of Science and Technology, Chennai.*
[3]*Student, Sathyabama Institute of Science and Technology, Chennai.*
[*]*Corresponding author E-mail:ashapandian225@gmail.com*

Abstract

Social media usage is increasing day-by-day, besides its uses, many cyber harassment cases are being filed, leaving many adults as victims. It is better to stop these incidents before they happen. This paper main aim is to provide a secure and safe social media environment for its users. Due to some privacy issues and missing of data when deleted it is difficult to find the convict, so timeline posts are considered rather than messages. This paper uses a method semantic enhanced marginalized De-noising auto-encoder for automatic detection of bad posts which prevent them from being posted on the timeline. Whenever a user tries to post something bad in words or sentences, an alert box will be shown that usage of bad words not allowed. This reduces many vulgar posts in the social media.

Keywords: *Bad posts; Cyber harassment; Semantics based De-noising; Social Media.*

1. Introduction

Cyber harassment is a way of harming others mentally in a repeated way in the use of internet-related social media networks. Many people are being affected by cyber harassment, according to the past analysis in social media like Face book if 75% of youth are present among them 54% are being affected by this, In YouTube among 66% of adults 21% faced these incidents, In Twitter within 43% of young people more than half of them i.e., 28% are victims to it and in Instagram 24% of total have experienced cyber harassment. Analysis considering the gender is like 63% of female and 37% of male among the total social media users are facing this typical issue. Many female users are being affected by cyber harassment, only a few look forward to filing complaints against the convict, few report spam and leave, but there will be some victims who are severely affected and their situation is embarrassing to share, in these case some are taking suicidal decisions which are leading to deaths. All these are to be stopped before an incident happens; this project provides a method for automatic detection of bad words in posts and does not allow the user to post it on the timeline. Not only words but also in sentences it finds the bad content and shows an alert box that usage of bad words not allowed. Adapting this to social media reduces many cyber harassment cases.

2. Existing System

Homa Hosseinmardi et al. [1] proposed that "Instagram" is a trending application that is mostly used by the teenagers nowadays and also mostly suffering. They are collecting the data of images and their relative content like comments. Then, the people prepare the design for labeling the text, and also they provide the jobs too few people for their ideas and financials from a large group of net users is known as the crowded source. They divide the sessions based on cyber bullying and they provide the entire diagnosis of the content, which also constitutes the connection between cyber bullying and other similarities like temper, the way of behavior and data of an image. The main disadvantage is, they make the entire diagnosis after the incident, it is better to identify before the tragedy.

According to Dinakar et al. [2] cyber bullying score is being increased in number leaving adolescents as victims. They say that the behavior of being individual and lack of supervision in the electronic medium is the main causes of social menace. As comments and posts that containing sensitive topics which personally affecting the individual, they consider only them. Their way of approach is like decomposing the overall problem into some sensitive topics i.e., sub problems. As they considered text classification, they made an experiment by considering 4500 YouTube comments by applying binary and multi-class classifiers. They finally concluded that textual cyber bullying can be handled by building individual topic sensitive classifiers. They only just mentioned how the textual cyber bullying can be handled.

Bayzick et al. [3] informs that, according to them cyber bullying is becoming harmful these days through the electronic text medium. They developed software for detecting online chat conversations related to cyber bullying. They used a rule-based dictionary of keywords for classifying the posts. They used the data sets of MySpace social network. They identified 85.3% of cyber bullying and 51.91% of innocent words in a time interval, where the overall accuracy is of 58.63%. The main drawback is that their code is not so perfect for clarifying whether it is innocent conversion or not.

F.Godlin et al. [4] main aim is to find the named entities like names of persons, organizations, locations etc in the posts of Twitter, but the posts contain small content and errors which makes a challenging task for them. As they are using unlabelled data it is semi-supervised learning. They used a large number of Twitter posts to generate the best output for given input. They secured 4th position in the final ranking.

B.Sri Nandhini et al. [5] chose social networking sites for their mission, as it begins gradually from the past years. Which provides the program to people for communicating and sharing their

likes and dislikes, along with this cyber bullying is also gradually increased. They are conveying that cyber bullying is harassing or insulting a person by sending messages of bullying words or threatening mentally by their electronic communication, which leads to the physical and mental depression of the one. They proposed a method which is used for identifying the cyber bullying activities and their terms. They used fuzzy logic and genetic algorithm; these are used to collect the relative data for the classification and improving the circumstances to produce an effective outcome. The main disadvantage is, there is no guarantee for identifying the appropriate results. Pierre Baldi [6] said that it is difficult to understand nonlinear data, but by using these auto-encoders we can understand both the linear and nonlinear data as it provides some mathematical forms in an easy manner. They named the auto-encoder as Boolean which understands nonlinear data. They also say that their method is equal to cluster process. In cluster process, it works fast when the number of clusters less and slow when more. It replaced all the old methods.

Sandra Bosackia et al. [7] tried to say whether peer and internal problems like depression, anxiety related to self-esteem. They made a survey among 7290 adults where 3756 are girls about peer relations like direct, indirect victims, socially isolation, friendship, trust and self-esteem like anxiety and depression. They gave regression analysis based on both, whereas friendship analysis gave partial depression only. They gave theoretically and applied implications and their discussions. M. Dadvar et al. [8] proposed improved cyberbullying detection with respect to user context. As the negative results of cyber bullying are winding up all the more disturbing each day and specialized technologies that take action by means of robotized recognition are still limited. According to them up-to then, finding cyber bullying concentrated on only comments, slighting setting, for example, clients qualities and profile data. They concluded that considering user context improves the detection.

M. Kaplan et al. [9] discussed that nowadays social media is trending and popular as many people are using for their business purpose. The companies are creating the profile in social media applications like Face book, Twitter, YouTube, and Wikipedia by producing ads. By doing that, they are going to gain popularity and as well as profit. Some of the companies are like Myntra, Amazon, Banks, Health insurance, Educational applications, further Consultants. Social media means, the start-up is to describe the ideas of social media clearly and discuss how it is different from connecting ideas based on user context. Their like to produce a classification of social media which divides the group of applications which are based on present terms into more specific categories by their specifications: Blogs, Channels, Social Networking Sites, Virtual Game Worlds. Finally, they approved for 10 recommendations for business applications to use the Social Network.

T. Mikolov et al. [10] proposed a model is used for the representation of text in the form of vectors, as an extension, they performed this on large datasets. This can be used in filtering or getting the information from the text documents. When compared to the old models it proved as the best as it can make many changes in less time and cost. Also, they proved that it works better for finding the semantic similarities [11-14].

3. Comparative Study on Existing Methods

Table 1: Comparison of Existing system

Methodology	Advantages	Disadvantages
Bag of words	Represents text data for classifying documents. It is simple and flexible. It compares the words which are represented in vector form.	Fails in the context of words.
Sentiment analysis	This method determines the opinion or feeling of the text. It helps in finding the user response as negative or positive for giving reviews.	Few drawbacks due to grammatical mistakes, spelling mistakes, and different slangs.
Naive Bayes	This classifies whether the given text is interesting or not using probability.	Only applicable for few data sets.
Term frequency and Inverse document frequency	It gives values for the frequently occurring words to find the similarity between documents. Also gets the descriptive words from the given document in an easy way.	It cannot find the position in the text, bad at semantics and the same occurrence in various documents.
Latent Semantic Analysis	It performs well with the data sets of different topics. It is easier and faster to implement.	Not efficient with deep learning techniques. A linear model, not the best solution.
Genetic algorithm	It considers the initial state as parent and final state as a child, by performing techniques it obtains the best solution in less time.	There is no guarantee for the optimal solution.
Fuzzy logic	Don't require to train lots of data as it works with the rules of natural language. It is simple in the computation of words	One should not think with a Boolean brain as it is efficient.
Named entity recognition	It categorizes the given unstructured data into groups. It makes easier to locate information.	Computational cost is more.
Sparse autoencoder	Uses word embeddings which is a better representation.	It collects more information other than relevant.
Vector space model	Used in filtering the information, retrieval of information and ranking at the relevant level.	Poor at representing long documents.

4. Proposed System

In the proposed work, we are able to stop the bad posts or content on the timeline. An automatic detection method named semantic marginalized de-noising auto-encoder is used where this helps in detecting the bad words in given sentence automatically and stops from being posted. On creating an online social network we perform how the method works. User registration to the social network is given where the user can fill his details and while login user details are checked whether authorized or not. User homepage consists of many features like posting on the timeline, searching friends, friend requests, message requests, change profile picture and details. Admin login consists of malicious users list – which displays the table of users who are using bad words with an option of function block, blocked user list – which consists of list who are blocked by admin when frequently using bad words and add bad words – where admin can add both illegal or vulgar words list which is helpful for the comparison.

All the users are provided with the features of the social network, but when a user tries to post bad content on the timeline, for that specific user an alert box will be showing that " you are using bad words " and the user will be unable post the particular content. Here the sentence the user types along with the image is checked with the database words, if any of the words match all through the sentence an alert box will be displayed, all this happens in the backend using the automatic detection method. Using this we can detect the malicious users in the social network and for the safe environment, it is better to block the users who are trying to post bad posts often (Figure 1).

4.1. Advantages of Proposed System

1. Human observation is not required.
2. It is better to stop bad posts at the starting level of posting on

the timeline, the proposed method helps in doing this.
3. It is an efficient way for the cyber harassment detection.

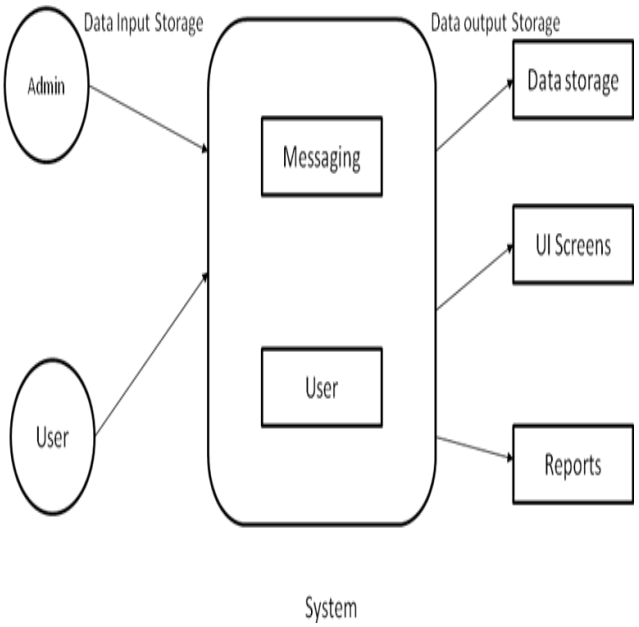

Fig. 1: System Architecture

5. Results and Discussion

The implementation has been done in Java with Server side technologies, Servlets and JSP and the Client side technologies such as Html and CSS, with MySQL databases.

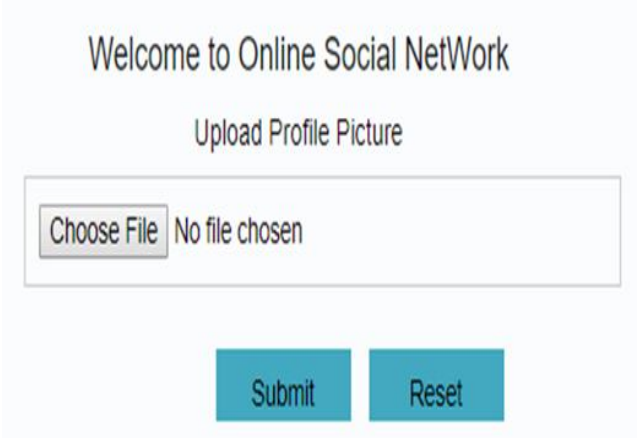

Fig. 2: User profile picture upload page

User profile picture upload page: Here user can upload a profile picture by browsing the image from the files in their systems (Figure 2).

Alert message: This is displayed when a user tries to post bad content on the timeline. The user will not able to post and his details are sent to the admin (Figure 3).

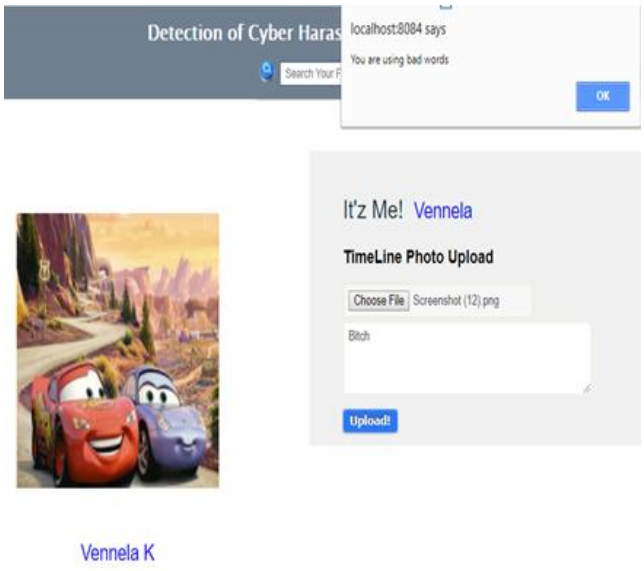

Fig. 3: Alert box

MALICIOUS USER DETAILS

User ID	Name	Email	Gender	Unwanted Message	Block
2	Vennela	vennie@gmil.com	female	Kill	block
2	Vennela	vennie@gmil.com	female	kill	block
4	Bhanu	bhanu@yahoo.com	female	slut	block
2	Vennela	vennie@gmil.com	female	It Is a boozer	block
2	Vennela	vennie@gmil.com	female	it is boozer	block
2	Vennela	vennie@gmil.com	female	He is slut	block
2	Vennela	vennie@gmil.com	female	Bitch	block
2	Vennela	vennie@gmil.com	female	He is boozer	block
2	Vennela	vennie@gmil.com	female	he is a boozer	block
8	A. Lakshmi	prasanna@gmail.com	female	he is boozer	block
8	A. Lakshmi	prasanna@gmail.com	female	she is a slut and boozer	block
8	A. Lakshmi	prasanna@gmail.com	female	Is she looking hot	block
8	A. Lakshmi	prasanna@gmail.com	female	he is boozer and hot	block

Fig. 4: Admin page with malicious user details

Admin page with malicious user details: Displays the table of information about malicious users mentioning the unwanted message they use and a block option is provided. On clicking the block option, the malicious user cannot login again (Figure 4).

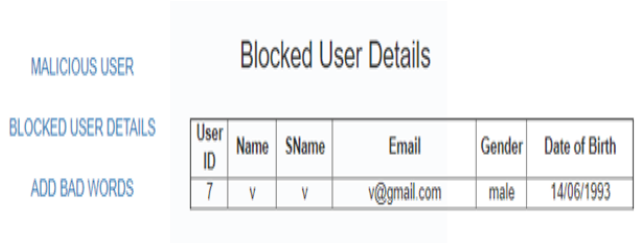

Fig. 5: Admin page with blocked user details

Admin page with blocked user details: Displays the table of all the blocked users by the admin, who are using bad words. These users cannot login to their accounts anymore (Figure 5).

6. Conclusion

This can be implemented in any social media and it is mainly best for the social networks having the timeline like Twitter, Instagram, and Facebook. Using this we can reduce many harassing cases about the timeline posts. Sometimes people will be using their regional language and different slangs, in these cases, it will not able to detect, further modifications can make this work more efficient.

References

[1] Hosseinmardi, Homa, Mattson, S.A., Rafiq, R.I., Han, R., Lv, Q. and Mishra, S(2015), Analyzing labeled cyberbullying incidents on the instagram social network." *International Conference on Social Informatics*. Springer, Cham, 2015.

[2] Dinakar, K., Reichart, R. and Lieberman, H., 2011. Modeling the detection of Textual Cyberbullying. *The Social Mobile Web*, *11*(02), pp.11-17.

[3] Bayzick, J., Kontostathis, A. and Edwards, L., 2011, June. Detecting the presence of cyberbullying using computer software. In *3rd Annual ACM Web Science Conference (WebSci '11)* (pp. 1-2).

[4] Godin, F., Vandersmissen, B., De Neve, W. and Van de Walle, R., 2015. Multimedia Lab $@ $ ACL WNUT NER Shared Task: Named Entity Recognition for Twitter Microposts using Distributed Word Representations. In *Proceedings of the Workshop on Noisy User-generated Text* (pp. 146-153).

[5] Nandhini, B.S. and Sheeba, J.I., 2015. Online social network bullying detection using intelligence techniques. *Procedia Computer Science*, *45*, pp.485-492.

[6] Baldi, P., 2012, June. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning* (pp. 37-49).

[7] Bosacki, S., Dane, A., Marini, Z. and YLC-CURA, 2007. Peer relationships and internalizing problems in adolescents: mediating role of self-esteem. *Emotional and Behavioural Difficulties*, *12*(4), pp.261-282.

[8] Dadvar, M., Trieschnigg, D., Ordelman, R. and de Jong, F., 2013, March. Improving cyberbullying detection with user context. In *European Conference on Information Retrieval* (pp. 693-696). Springer, Berlin, Heidelberg.

[9] Kaplan, A.M. and Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, *53*(1), pp.59-68.

[10] Mikolov, T., Chen, K., Corrado, G. and Dean, J., 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[11] Huang, Q., Singh, V.K. and Atrey, P.K., 2014, November. Cyber bullying detection using social and textual analysis. In *Proceedings of the 3rd International Workshop on Socially-Aware Multimedia* (pp. 3-6). ACM.

[12] P. Asha, Roshni Sridhar and Rinnu Rose P. Jose, Click Jacking Prevention in Websites using Iframe Detection and IP Scan Techniques, *ARPN Journal of Engineering and Applied Sciences* 11(15), 9166-9170, 2016.

[13] P.Asha, Prasanth R and Pranath P, Ranking the Product Details and its Application using Sentiment Classification, *Research Journal of Pharmaceutical, Biological and Chemical Sciences* 7(6), 1399-1405, 2016.

[14] P. Asha, Madhavi N. Latha, and K. Architha, Design And Implementation Of IOT Based Security Aware Architecture Using IDS, *Research Journal of Pharmaceutical, Biological and Chemical Sciences* 8(2), 2293-2300, 2017.