International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 21 No. 3 (1) 2025

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 21, No. 3(1), 2025

Research Paper

DETECTION OF CHILD PREDATORS / CYBER HARASSERS ON SOCIAL MEDIA

[1]B. SNEHALATHA,[2]CH. LAXMI PRASANNA,[3]B. KALYAN,[4]Mrs. DURGA BHAVAN
[123] *Students,* [4] *Assistant Professor*
Department Of Information Technology
Teegala Krishna Reddy Engineering College, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

The use of social media has grown exponentially over time with the growth of the Internet and has become the most influential networking platform in the 21st century. However, the enhancement of social connectivity often creates negative impacts on society that contribute to a couple of bad phenomena such as online abuse, harassment cyberbullying, cybercrime and online trolling. Cyberbullying frequently leads to serious mental and physical distress, particularly for women and children, and even sometimes force them to attempt suicide. Online harassment attracts attention due to its strong negative social impact. Many incidents have recently occurred worldwide due to online harassment, such as sharing private chats, rumours, and sexual remarks. Therefore, the identification of bullying text or message on social media has gained a growing amount of attention among researchers. The purpose of this research is to design and develop an effective technique to detect online abusive and bullying messages by merging natural language processing and machine learning. Two distinct freatures, namely Bag- of - Words (BoW) and term frequency-inverse text frequency (TFIDF), are used to analyse the accuracy level of four distinct machine learning algorithms.

Received: 10-7-2025 Accepted: 18-8-2025 Published: 25-8-2025

I. INTRODUCTION

Social media is a platform that allows people to post anything like photos, videos, documents extensively and interact with society . People connect with social media using their computers or smart phones. The most popular social media includes Facebook1, Twitter2, Instagram3, TikTok4 and so on. Nowadays, social media is involved in different sectors like education, business , and also for the noble cause . Social media is also enhancing the world's economy through creating many new job opportunities . Although social media has a lot of benefits, it also has some draw backs. Using this media, malevolent users conduct unethical and fraudulent acts to hurt others feelings and damage their reputation. Recently, cyber bullying has been one of the major social media issues. Cyber bullying or cyber-harassment refers to an electronic method of bullying or harassment. Cyber bullying and cyber-harassment are also known as online bullying. As the digital realm has grown and technology has progressed, cyber bullying has become relatively common, particularly amongst adolescents

social media, as defined as ''a group of Internetbased applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content.'' Via social media, people can enjoy enormous information, convenient communication experience and so on. However, social media may have some side effects such as cyberbullying, which may have negative impacts on the life of people, especially children and teenagers.

Cyberbullying can be defined as aggressive, intentional actions performed by an individual or a group of people via digital communication methods such as sending messages and posting comments against a victim. Different from traditional bullying that usually occurs at school during faceto-face communication, cyberbullying on social media can take place anywhere at any time. For bullies, they are free to hurt their peers' feelings because they do not need to face someone and can hide behind the Internet. For victims, they are easily exposed to harassment since all of us, especially youth, are constantly connected to Internet or social media. As reported, cyberbullying victimization rate ranges from 10% to 40%. In the United States, approximately 43% of teenagers were ever bullied on social media. The same as traditional bullying, cyberbullying has negative, insidious and sweeping impacts on children. The outcomes for victims under cyberbullying may even be tragic such as the occurrence of self-injurious behaviour or suicides.

tomatically detect and promptly report bullying messages so that proper measures can be taken to prevent possible tragedies. Previous works on computational studies of bullying have shown that natural language processing and machine learning are powerful tools to study bullying. Cyberbullying detection can be formulated as a supervised learning problem. A classifier is first trained on a cyberbullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message. Three kinds of information including text, user demography, and social network features are often used in cyberbullying detection. Since the text content is the most reliable, our work here focuses on text-based cyberbullying detection.

In the text-based cyberbullying detection, the first and also critical step is the numerical representation learning for text messages. In fact, representation learning of text is extensively studied in text mining, information retrieval and natural language processing (NLP). Bag-of-words (BoW) model is one commonly used model that each dimension corresponds to a term. Latent Semantic Analysis (LSA) and topic models are another popular text representation models, which are both based on BoW models. By mapping text units into fixed-length vectors, the learned representation can be further processed for numerous language processing tasks. Therefore, the useful representation should discover the meaning behind text units. In cyberbullying detection, the numerical representation for Internet messages should be robust and discriminative. Since messages on social media are often very short and contain a lot of informal language and misspellings, robust representations for these messages are required to reduce their ambiguity. Even worse, the lack of sufficient high-quality training data, i.e., data sparsity make the issue more challenging. Firstly, labeling data is labor intensive and time consuming. Secondly, cyberbullying is hard to describe and judge from a third view due to its intrinsic ambiguities. Thirdly, due to protection of Internet users and privacy issues, only a small portion of messages are left on the Internet, and most bullying posts are deleted. As a result, the trained classifier may not generalize well on testing messages that contain nonactivated but discriminative features. The goal of this present study is to develop methods that can learn robust and discriminative representations to tackle the above problems in cyberbullying detection.

Some approaches have been proposed to tackle these problems by incorporating expert knowledge into feature learning. Yin et.al proposed to combine BoW features, sentiment features and contextual features to train a support vector machine for online harassment detection. Dinakar et.al utilized label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis. In addition, common sense knowledge was also applied. Nahar et.al

presented a weighted TF-IDF scheme via scaling bullying-like features by a factor of two. Besides content-based information, Maral et.al proposed to apply users' information, such as gender and history messages, and context information as extra features. But a major limitation of these approaches is that the learned feature space still relies on the BoW assumption and may not be robust. In addition, the performance of these approaches rely on the quality of hand-crafted features, which require extensive domain knowledge.

In this paper, we investigate one deep learning method named stacked denoising autoencoder (SDA). SDA stacks several denoising autoencoders and concatenates the output of each layer as the learned representation. Each denoising autoencoder in SDA is trained to recover the input data from a corrupted version of it. The input is corrupted by randomly setting some of the input to zero, which is called dropout noise. This denoising process helps the autoencoders to learn robust representation. In addition, each autoencoder layer is intended to learn an increasingly abstract representation of the input. In this paper, we develop a new text representation model based on a variant of SDA: marginalized stacked denoising autoencoders (mSDA), which adopts linear instead of nonlinear projection to accelerate training and marginalizes infinite noise distribution in order to learn more robust representations. We utilize semantic information to expand mSDA and develop Semantic- enhanced Marginalized Stacked Denoising Autoencoders (smSDA). The semantic information consists of bullying words. An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labor can be reduced. During training of smSDA, we attempt to reconstruct bullying features from other normal words by discovering the latent structure, i.e. correlation, between bullying and normal words. The intuition behind this idea is that some bullying messages do not contain bullying words. The correlation information discovered by smSDA helps to reconstruct bullying features from normal words, and this in turn facilitates detection of bullying messages without containing bullying words.

II. LITERATURE SURVEY

Author:] N. Selwyn

Title: SOCIAL MEDIA IMPACT ON LANGUAGE LEARNING FOR SPECIFIC PURPOSES: A STUDY IN ENGLISH FOR BUSINESS ADMINISTRATION

Description: Nowadays, social media are dominating the life of people. Facebook has become noticeably widespread among the youth, and students in particular. Research has indicated that Facebook could be an effective platform for language learning. This study, therefore, comes to explore the effects of Facebook-assisted teaching on learning English for specific purposes by students at the University of Tabuk, Saudi Arabia. A sample of 64 students from the Faculty of Business Administration, taking a Business Letters course in English, were divided into a Facebook-tutored group and a traditional classroom tutored group and were given the same vocabulary content. The two groups were given pre- and post-tests to measure their vocabulary learning, and were subjected to an interview to gauge their attitudes towards the instructional methods which were put to use. However, no significant difference between the two groups was found in terms of achievement in spite of the positive response and the high satisfaction level the Facebook-tutored students showed towards the use of such a platform.

Author:] H. Karjaluoto, P. Ulkuniemi, H. Keinanen, and O. Kuivalainen

Title: Antecedents of social media B2B use in industrial marketing context: customers' view

Description: Purpose – The purpose of this study is to clarify business-to-business (B2B) customers' behavior regarding their social media

use for B2B purposes and the antecedents of this behavior in the industrial marketing setting. It explores the influence of corporate culture, colleagues' support and personal and psychological factors on customer behavior toward social media business use. Design/methodology/approach – The authors conducted an online questionnaire survey among key customer accounts of an information technology service company (N 82). Partial least squares (PLS) path modeling was utilized to analyze the relationship between the dependent variable (social media business use) and the independent variables. Findings – Results show that private social media usage has the most significant relationship with the social media business use. Colleagues at work are also Supporting B2B social media use and personal characteristics are also of importance. Surprisingly, perception of usability of social media for B2B use did not explain social media business use within our sample. Research limitations/implications – The chosen methodology, sampling frame and sample size may limit generalizability. Therefore, researchers are encouraged to test the proposed hypothesis in other settings, particularly as the diffusion of B2B social media increases. Practical implications – The paper provides insights into how marketing managers can make an impact with their social media marketing. For example, when planning social media activities, companies need to consider which social media services could serve their marketing and communication targets and would reach the customers. Originality/value – Studies related to social media in B2B, especially from a customer's perspective, are still limited, and the authors do not know how customer firms value industrial marketing activities in social media. This novel paper provides insights into managers' reasons for using social media and gives guidelines for B2B marketers on how to conduct social media marketing in the future.6

Author: W. Akram, R. Kumar,
Title: "A Study on Positive and Negative Effects of Social Media on Society,"
Description: Social media is a platform for public around the World to discuss their issues and opinions. Before knowing the actual aspects of social media people must have to know what does social media mean? Social media is a term used to describe the interaction between groups or individuals in which they produce, share, and sometimes exchange ideas, images, videos and many more over the internet and in virtual communities. Children are growing up surrounded by mobile devices and interactive social networking sites such as Twitter, myspace, and Facebook, Orkut which has made the social media a vital aspect of their life. Social network is transforming the behavior in which youthful people relate with their parents, peers, as well as how they make use of technology. The effects of social networking are twofold.[1] On the positive side, social networks can act as invaluable tools for professionals. They achieve this by assisting young professionals to market their skills and seek business opportunities. Social networking sites may also be used to network efficiently. On the negative side, the internet is laden with a number of risks associated with online communities. Cyber bullying, which means a type of harassment that is perpetrated using electronic technology, is one of the risks. In this paper we cover every aspect of social media with its positive And negative effects. Focus is on the particular field like health, business, education, society and youth. During this paper we explain how these media will influence the society in a broad way
Title: Improved cyberbullying detection using gender information
Author Names: M. Dadvar, F. De Jong, R. Ordelman, and R. Trieschnigg
Description: ne may have hundreds of "friends" without even seeing their faces. Meanwhile, alongside this transition there is increasing

evidence that online social applications are used by children and adolescents for bullying. State-of-the-art studies in cyberbullying detection have mainly focused on the content of the conversations while largely ignoring the characteristics of the actors involved in cyberbullying. Social studies on cyberbullying reveal that the written language used by a harasser varies with the author"s features including gender. In this study we used a support vector machine model to train a gender-specific text classifier. We demonstrated that taking gender-specific language features into account improves the discrimination capacity of a classifier to detect cyberbullying.

Title: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,"

Author Names: P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol

Description: The resulting algorithm is a straightforward variation on the stacking of ordinary autoencoders. It is however shown on a benchmark of classification problems to yield significantly lower classification error, thus bridging the performance gap with deep belief networks (DBN), and in several cases surpassing it. Higher level representations learnt in this purely unsupervised fashion also help boost the performance of subsequent SVM classifiers. Qualitative experiments show that, contrary to ordinary autoencoders, denoising autoencoders are able to learn Gabor-like edge detectors from natural image patches and larger stroke detectors from digit images. This work clearly establishes the value of using a denoising criterion as a tractable unsupervised objective to guide the learning of useful higher level representations.

EXISTING SYSTEM

A classifier is first trained on a cyber bullying corpus labeled by humans, and the learned classifier is then used to recognize a bullying message. Three kinds of information including text, user demography, and social network features are often used in cyberbullying detection . Since the text content is the most reliable, our work here focuses on text-based cyberbullying detection.

In the text-based cyberbullying detection, the first and also critical step is the numerical representation learning for text messages. In fact, representation learning of text is extensively studied in text mining, information retrieval and natural language processing (NLP). Bag-of-words (BoW) model is one commonly used model that each dimension corresponds to a term. Latent Semantic Analysis (LSA) and topic models are another popular text representation models, which are both based on BoW models. By mapping text units into fixed-length vectors, the learned representation can be further processed for numerous language processing tasks.

PROPOSED SYSTEM

Some approaches have been proposed to tackle these problems by incorporating expert knowledge into feature learning. Proposed to combine BoW features, sentiment features and contextual features to train a support vector machine for online harassment detection.

It can utilized label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis. In addition, common sense knowledge was also applied. Nahar et.al presented a weighted TF-IDF scheme via scaling bullying-like features by a two factor. Besides content-based information, Maral et.al proposed to apply users' information, such as gender and history messages, and context information as extra features. But a major limitation of these approaches is that the learned feature space still relies on the BoW assumption and may not be robust. In addition, the performance of these approaches rely on the quality of hand-crafted features, which require extensive domain knowledge.

Advantages

1) Most cyber bullying detection methods rely on the BoW model. Due to the sparsity problems of both data and features, the classifier may not be trained very well. Stacked densoing autoencoder (SDA), as an unsupervised representation learning method, is able to learn a robust feature space. In SDA, the feature correlation is explored by the reconstruction of corrupted data. The learned robust feature representation can then boost the training of classifier and finally improve the classification accuracy. In addition, the corruption of data in SDA actually generates artificial data to expand data size, which alleviate the small size problem of training data.

2) For cyberbullying problem, we design semantic dropout noise to emphasize bullying features in the new feature space, and the yielded new representation is thus more discriminative for cyberbullying detection.

3) The sparsity constraint is injected into the solution of mapping matrix W for each layer, considering each word is only correlated to a small portion of the whole vocabulary. We formulate the solution for the mapping weights W as an Iterated Ridge Regression problem, in which the semantic dropout noise distribution can be easily marginalized to ensure the efficient training of our proposed smSDA.

4) Based on word embeddings, bullying features can be extracted automatically. In addition, the possible limitation of expert knowledge can be alleviated by the use of word embedding.

III. MODULES DESCSRIPTION:

Admin:

In this module admin can login in to application and he can view the accounts which are involving cyber crime, and admin can control the accounts of users in online social network by activating or deactivating based on the user behavior.

User:

This module is used for new user registrations and after registrations the users can login with their authentication. Where after the existing users can send messages to privately and publicly, options are built. Users can also share post with others. The user can able to search the other user profiles and public posts. In this module users can send or accept send friend requests, user can also comment or like the posts posted by others in the social network.

Construction of Bullying Feature Set:

The bullying features play an important role and should be chosen properly. In the following, the steps for constructing bullying feature sets are given, in which the first layer and the other layers are addressed separately. For the first layer, expert knowledge and word embeddings are used. For the other layers, discriminative feature selection is conducted. In this module firstly, we build a list of words with negative affective, including swear words and dirty words. Then, we compare the word list with the BoW features of our own corpus, and regard the intersections as bullying features. Finally, the constructed bullying features are used to train the first layer in our proposed smSDA. It includes two parts: one is the original insulting seeds based on domain knowledge and the other is the extended bullying words via word embeddings.

Cyber bullying Detection:

In this module we propose the Semantic-enhanced Marginalized Stacked Denoising Auto-encoder (smSDA). In this module, we describe how to leverage it for cyber bullying detection. smSDA provides robust and discriminative representations The learned numerical representations can then be fed into our system. In the new space, due to the captured feature correlation and semantic information, even trained in a small size of training corpus, is able to achieve a good performance on testing documents.

Based on word embeddings, bullying features can be extracted automatically. In addition, the

possible limitation of expert knowledge can be alleviated by the use of word embedding

Semantic-Enhanced Marginalized Denoising Auto-Encoder:

An automatic extraction of bullying words based on word embeddings is proposed so that the involved human labor can be reduced. During training of smSDA, we attempt to reconstruct bullying features from other normal words by discovering the latent structure, i.e. correlation, between bullying and normal words. The intuition behind this idea is that some bullying messages do not contain bullying words. The correlation information discovered by smSDA helps to reconstruct bullying features from normal words, and this in turn facilitates detection of bullying messages without containing bullying words. If bullying messages do not contain such obvious bullying features, the correlation may help to reconstruct the bullying features from normal ones so that the bullying message can be detected.

IV. SYSTEM ARCHITECTURE:

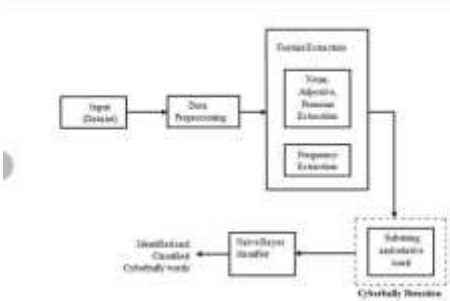
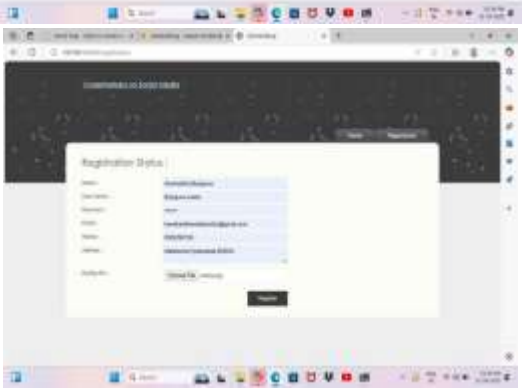


Fig: System Architecture

V. OUTPUT SCREENS



VI. CONCLUSION

CONCLUSION In particular, cyber bullying has become more common and has begun to raise significant social issues with the rising prevalence of social media sites and increased social media use by teenagers. There needs to design automatic cyberbullying detection method to avoid bad consequences of cyber harassment. Considering the significance of cyberbullying detection, in this study, we investigated the automated identification of posts on social media related to cyberbullying by considering two features BoW and TF-IDF. Four machine learning algorithms are used to identify bullying text and SVM for both BoW and TF-IDF. In future we are planning to design a framework for automatic detection and

classification of cyberbullying from Bengali texts using deep learning algorithms.

REFERENCES

1. C. Fuchs, Social media: A critical introduction. Sage, 2017.
2. N. Selwyn, "Social media in higher education," The Europa world of learning, vol. 1, no. 3, pp. 1– 10, 2012.
3. H. Karjaluoto, P. Ulkuniemi, H. Keinanen, and O. Kuivalainen, "An- ¨ tecedents of social media b2b use in industrial marketing context: customers' view," Journal of Business & Industrial Marketing, 2015.
4. W. Akram and R. Kumar, "A study on positive and negative effects of social media on society," International Journal of Computer Sciences and Engineering, vol. 5, no. 10, pp. 351–354, 2017.
5. D. Tapscott et al., The digital economy. McGraw-Hill Education,, 2015. [6] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet, and I. De Bourdeaudhuij, "Cyberbullying on social network sites. an experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully," Computers in Human Behavior, vol. 31, pp. 259–271, 2014.
6. D. L. Hoff and S. N. Mitchell, "Cyberbullying: Causes, effects, and remedies," Journal of Educational Administration, 2009.
7. S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide," Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.
8. 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, pp. 1–7, 2009.
9. K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying," in In Proceedings of the Social Mobile Web. Citeseer, 2011.
10. K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying." in The Social Mobile Web, 2011.
11. V. Nahar, X. Li, and C. Pang, "An effective approach for cyberbullying detection," Communications in Information Science and Management Engineering, 2012.
12. M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, "Improved cyberbullying detection using gender information," in Proceedings of the 12th -Dutch-Belgian Information Retrieval Workshop (DIR2012). Ghent, Belgium: ACM, 2012.
13. M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, "Improving cyberbullying detection with user context," in Advances in Information Retrieval. Springer, 2013, pp. 693–696.
14. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," The Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.
15. P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7, p. 43, 2012.
16. M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," arXiv preprint arXiv:1206.4683, 2012.
17. T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," Discourse processes, vol. 25, no. 2-3, pp. 259–284, 1998.
18. T. L. Griffiths and M. Steyvers, "Finding scientific topics," Proceedings of the National academy of Sciences of the United States of America, vol. 101, no. Suppl 1,

pp. 5228–5235, 2004.
19. D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” the Journal of machine Learning research, vol. 3, pp. 993–1022, 2003.