DETECTION OF CYBER BULLYING IN DIGITAL FORUMS

#V. RAVI KUMAR[1], Email:ravivannoj@tkrcet.com
#SAMYU MAN CHIKANTI[2], Email:samyusamyu3082000@gmail.com
#N.PURNA PRASAD[3], Email:prasadpurna8@gmail.com
# K.NITHIN REDDY[4], Email:nithinreddy537@gmail.com
#BTech Student, TKR College of Engineering and Technology, Hyderabad,INDIA.

Abstract:-Prior to the innovation of Information Communication Technologies (ICT), social interactions evolved within small cultural boundaries, such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications. These social technologies have created a revolution in user-generated information, online human networks, and rich human behavior-related data. However, the misuse of social technologies, such as social media (SM) platforms, has introduced a new form of aggression and violence that occurs exclusively online. A new means of demonstrating aggressive behavior in SM websites is highlighted in this work. The motivations for the construction of prediction models to fight aggressive behavior in SM are also outlined. We comprehensively review cyberbullying prediction models and identify the main issues related to the construction of cyberbullying prediction models in SM. This paper provides insights on the overall process for cyberbullying detection and most importantly overviews the methodology. Though data collection and feature engineering process has been elaborated, yet most of the emphasis is on feature selection algorithms and then using various machine learning algorithms for prediction of cyberbullying behaviors. Finally, issues and challenges have been highlighted as well, which present new research directions for researchers to explore.

Keywords: social media (SM) platforms, cyberbullying,

INTRODUCTION

Social media, as defined as “a group of Internet based 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 face-to-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.

This will automatically 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 makes 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.

2.Problem definition

Approximately 50% of the teenagers in America experience cyberbullying. This bullying has a physical and mental impact on the victim. The victims choose self-destructive acts like suicide because the trauma of cyberbullying which is hard to be endured. Thus, the identification and prevention of cyberbullying is important to protect teenagers.

Low Accuracy, More time for Dataset Training.

3.METHODOLOGY

We describe the cyberbullying detection framework which consists of two major parts: The first part is called NLP (Natural Language Processing) and the second part is named as ML (Machine learning).

In the first phase, datasets containing bullying texts, messages or post are collected and prepared for the machine learning algorithms using natural language processing.

The processed datasets are then used to train the machine learning algorithms for detecting any harassing or bullying message on social media including Facebook and Twitter.

Advantage:

High Accuracy, Faster Dataset Training than before.

Implementation

The development of the project is based on the Dataset considered and effective tuning of parameters of Machine Learning Algorithms. The system consists of basically 4 phases: 1. Data Gathering 2. Data processing 3. Training Phase 4. Testing Phase

Data Gathering: The dataset represented here is a collection of tweets which was collected using Twitter API. The number of data entries exceeded 1000 tweets which belong to different time periods. The following images depict the datasets indicating Text Labels.

Datasets:

We have collected Facebook comments from different posts (Dataset-1) and the twitter comments dataset from kaggle.com [27] for (Dataset-2). The texts or comments were classified into two types as follows:

Non-bullying Text: This type of comments or posts are non-bullying or positive comments. For example, the comment like ”This photo is very beautiful” is positive and non-bullying comments.

Bullying Text: This type belongs to bully type comments or harassments. For example, ”go away bitch” is a bullying text or comment and we consider as negative comment.

Data Processing: Adaptation of the raw data according to our need is important before implementing the regression model. Since, raw data is most of the time inconsistent or incomplete or lacking in certain behavior or lacking in attributes or may contain noises. So, we need to remove all these abnormalities and convert the dataset into something which can be used by the machine learning algorithms. So, we processed data obtained from online sources to obtain useful data metrics, related to profanity in the output, on a daily basis which can be used to train our models. The comment data which we downloaded was in xlsx format. So, we had to convert the xlsx file format to csv format which is usual format used to train machine learning models. Further sometimes data contain various inconsistencies such as noisy data which model cannot interpret and value dominances of a variable over another which can cause model's inconsistency to predict accurately.

Training Phase:

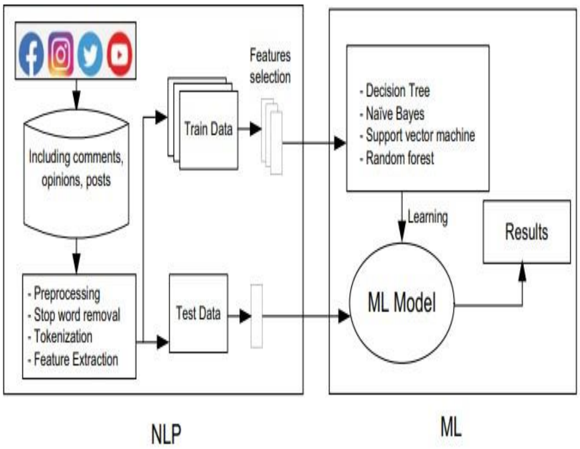
For training the model, first we import a specific algorithm class/module and create an instance of it. Then using that instance, we fit the model to the training data. Then we validate it by testing its accuracy score and fine tune its parameters till we get required results.

Testing Phase:

For testing the model, we compare its predicted values after the training phase with test data. Then input some different value for prediction and check whether it predicts it right. If it didn't predict right then, fine tune the algorithmic parameters and fit the model again.

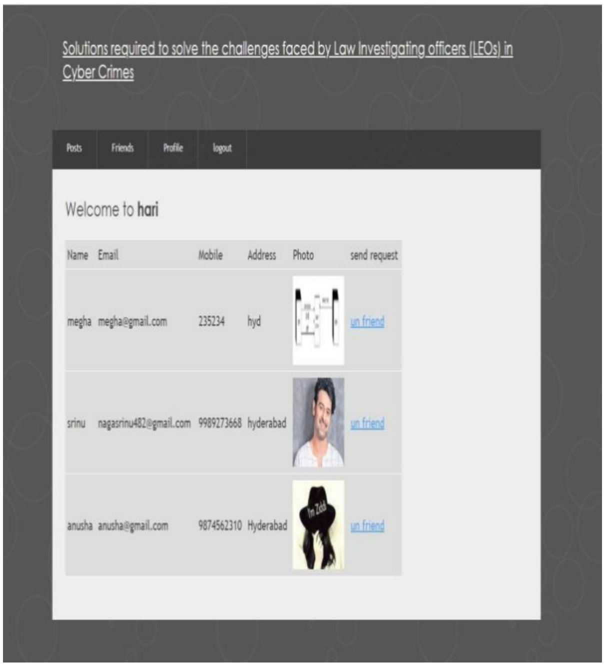
4.SYSTEM ARCHITECTURE

Figure 1: System architecture



It is used to abstract the overall outline of the software system and the relationships, constraints, and boundaries between components. It is an important tool as it provides an overall view of the physical deployment of the software system and its evolution roadmap.

5.Results

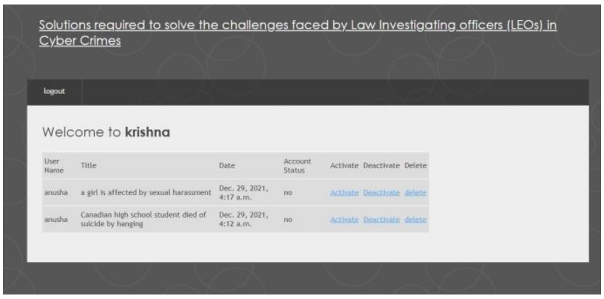


Solutions required to solve the challenges faced by Law Investigating officers (LEOs) in Cyber Crimes

Welcome to **hari**

Name	Email	Mobile	Address	Photo	send request
megha	megha@gmail.com	235234	hyd		un friend
srinu	nagasrinu482@gmail.com	9989273668	hyderabad		un friend
anusha	anusha@gmail.com	9874562310	Hyderabad		un friend

Figure3:send Request



Solutions required to solve the challenges faced by Law Investigating officers (LEOs) in Cyber Crimes

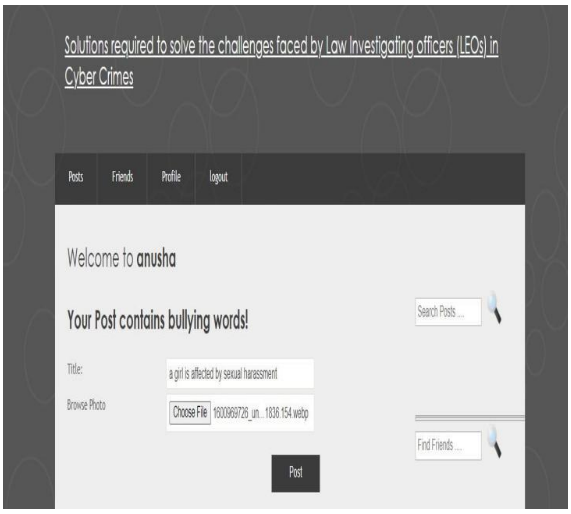
Welcome to **krishna**

User Name	Title	Date	Account Status	Activate	Deactivate	Delete
anusha	a girl is affected by sexual harassment	Dec. 29, 2021, 4:17 a.m.	no	Activate	Deactivate	delete
anusha	Canadian high school student died of suicide by hanging	Dec. 29, 2021, 4:12 a.m.	no	Activate	Deactivate	delete

Figure 4:Activate or deactivate account

6.CONCLUSION

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.



Solutions required to solve the challenges faced by Law Investigating officers (LEOs) in Cyber Crimes

Welcome to **anusha**

Your Post contains bullying words!

Title: a girl is affected by sexual harassment

Browse Photo: Choose File 1600969726_un...1836.154.webp

7.REFERENCES

[1] A. M. Kaplan and M. Haenlein, “Users of the world, unite! the challenges and opportunities ofsocialmedia,”Businesshorizons,vol.53 ,no.1,pp.59–68,2010.

[2] R. M. Kowalski, G. W. Giumetti, A. N. Schroeder, and M. R. Lattanner,

"Bullying in the digitalage:Acriticalreviewandmetaanalysisofcyberbullyingresearchamongyouth. "2014.

[3] M.Ybarra,"Trendsintechnology-basedsexualandnon-sexualaggressionovertimeandlinkagestonontechnology aggression," National Summit on Interpersona l Violence and Abuse Across theLifespan:ForgingaSharedAgenda,2010.

[4] B. K. Biggs, J. M. Nelson, and M. L. Sampilo, "Peer relations in the anxiety–depression link:Testofamediationmodel,"Anxiety,Stress,&Coping,vol.23,no.4,pp.431–447,2010.

[5] S.R. Jimerson, S. M.Swearer, and D. L. Espelage, Handbook of bullying inschools: Aninternationalperspective.Routledge/Taylor&FrancisGroup,2010.

[6] G. Gini and T. Pozzoli, "Association between bullying and psychosomatic problems: A me ta-analysis,"Pediatrics,vol.123,no.3,pp.1059–1065,2009.

[7] A. Kontostathis, L. Edwards, and A. Leatherman, "Text mining and cybercrime," Text Mining:ApplicationsandTheory.JohnWiley&Sons,Ltd,Chichester,UK,2010.

[8] J.-M. Xu, K.-S. Jun, X. Zhu, and A. Bellmore, "Learning from bullying traces in social media,"inProceedings ofthe2012 conference of theNorthAmericanchapter oftheassociation forcomputationallinguistics:Humanlanguagetechnologies.AssociationforComputationalLinguistics,2012,pp.656–666.

[9] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textualanalysis," in Proceedings of the 3rd International Workshop on Socially-Aware Multimedia. ACM,2014,pp.3–6.

[10] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Edwards, "Detection ofharassmentonweb2.0," ProceedingsoftheContentAnalysisintheWEB,vol.2,pp.1–7,2009.

[11] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying."inTheSocialMobileWeb, 2011.

[12] V.Nahar,X.Li,andC.Pang,"Aneffective approach forcyberbullying detection,"CommunicationsinInformati

onScienceandManagementEngineering, 2012.

[13] M. Dadvar, F. de Jong, R. Ordelman, and R. Trieschnigg, “Improved cyberbullying detectionusinggenderinformation,” inProceedingsofthe12th-Dutch-BelgianInformation RetrievalWorkshop(DIR2012).Ghent,Belgium:ACM,2012.

[14] M. Dadvar, D. Trieschnigg, R. Ordelman, and F. de Jong, “Improving cyberbullying detectionwithusercontext,”inAdvancesin InformationRetrieval.Springer,2013,pp. 693–696.

[15] P. Vincent, H.Larochelle, I.Lajoie, Y.Bengio, and P.-A. Manzagol, “Stacked denoisingautoencoders: Learning useful representations in a deep network with a local denoising criterion,”TheJournalofMachineLearningResearch,vol.11,pp.3371–3408,2010.

[16] P.Baldi, “Autoencoders, unsupervised learning, and deep architectures,” Unsupervised andTransferLearningChallengesinMachineLearning,Volume7,p.43,2012.

[17] M.Chen,Z.Xu,K.Weinberger,andF.Sha,“Marginalizeddenoisingautoencodersfordomainadaptation,”arXivpreprintarXiv:1206.4683,2012.

[18] T. K. Landauer, P. W. Foltz, and D. Laham, “An introduction to latent semantic analysis,”Discourseprocesses,vol.25,no. 2-3,pp.259–284,1998.