Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7583875

DISTINGUISH AND RESTRICT THE CYBERBULLYING CONVERSATION ON SOCIAL NETWORKS USING SUPPORT VECTOR MACHINE ALGORITHM

M. Jayanthi Rao[1], A. Venkata Mahesh[2], P. Prasanthi[1], B. Ramakrishna[3*], M. Ramanaiah[4] and M. Balakrishna[5]

1. *Department of Computer Science and Engineering, Aditya Institute of Technology and Management, Tekkali-532201, India.*
2. *Department of Computer Science and Engineering, GIET University, Gunupur, Odisha 765022, India.*
3. *Department of Mechanical Engineering, Aditya Institute of Technology and Management, Tekkali-532201, India.*
4. *Department of Chemistry, Aditya Institute of Technology and Management, Tekkali-532201, India.*
5. *Department of Chemistry, Lendi Institute of Engineering and Technology, Vizianagaram-535005, India.*

* *CORRESPONDING AUTHOR:* ***BRKBTECH@GMAIL.COM***

ABSTRACT

In current days there was a lot of abused communication found in social media. A recent survey report confirmed that more than 80 percent of online social networks are having abused orvulgar communication on their user accounts. These types of messages are mainly posted on user walls in order to harass teens, preteens other children by posting these types of offensive messages. Till now no application is providing a solution for this cyber content not to spread on social media, so this motivated me to design this current application for stopping vulgar communication in online social networks. In this proposed application, we mainly try to propose a new representation learning method to tackle this problem for identifying and stopping the abused messages not to communicate in online chat. Here we try to use well-known machine learning algorithms suchas Support Vector Machine for classifying the abused messages and normal messages and, we use Porter Stemming Algorithm to pre-process the text messages. This Porter Stemming is a well-known NLT Package, which will divide the whole message into parts and then assign tokens for each individual word. Here, we classify the cyber bullied dialogue into five categories based on literature such as hate, vulgar, offensive, sex and violence.

KEYWORDS: Cyberbullying**,** Communication, Natural Language Toolkit, Support Vector Machine, Porter Stemming, Vulgar, Offensive

1. INTRODUCTION

Online social media refers to a collection of web-based tools that support the production and sharing of user-generated content and are founded on the conceptual and technical underpinnings of Web 2.0. Most of this information can be exchanged through social media, where users have access to a wealth of knowledge, convenient communication tools, and other things. Social media may, however, have certain unintended consequences, such as

cyberbullying, which can have a severe impact on people's lives, particularly those of children and teenagers. Because they do not have to confront anyone and can hide behind the internet, bullies are free to hurt their peers' feelings. Because we are all always connected to the Internet or social media, especially young people, victims are easily exposed to harassment. According to [2], victimization rates for cyberbullying range from 10% to 40%. Nearly 43% of teenagers in the US have experienced cyberbullying at some point [3]. Similar to traditional bullying, cyberbullying has detrimental, pervasive effects on kids [4], [5], [6], as seen in figure 1. The results of cyberbullying for victims may even be devastating, such the prevalence of self-harming behaviour or suicides. Automated detection and fast reporting of bullying communications is one strategy for combating the issue, allowing for the right action to be done to avert potential catastrophes. Natural language processing and machine learning are effective techniques for studying bullying, according to earlier computational studies [7, 8]. The problem of detecting cyberbullying can be expressed as supervised learning.

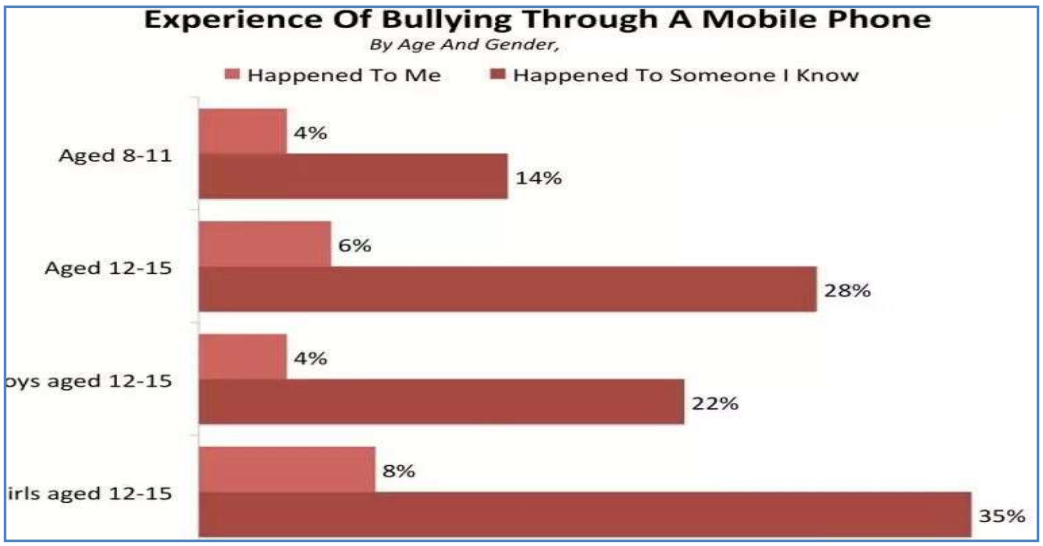


FIGURE 1. REPRESENT THE PERCENTAGE OF USERS WHO ARE GETTING BULLYING MESSAGES

The major goal of the current study is to create a unique technique that can train discriminative and robust representations to address the aforementioned issues with cyberbullying detection. In order to address these issues, expert knowledge is incorporated into feature learning [10]. Generally speaking, when we use certain words in sentence building, some words give positive meaning in one way while the same word give negative meaning in another way. To train a support vector machine for online cyberbullied detection, Dynacare g. Yin et. al proposed combining BoW features, sentiment features, and contextual features. They used label specific features to extend the general features, where the label specific features are learned by Linear Discriminative Analysis [11]. Additionally, we attempt to discover the weighted TF-IDF scheme by scaling bullying-like features by a factor of two using a mechanism from a toolkit for natural language processing [12]. In addition to content-based data, such as gender and history messages, we also try to find out the other information such as extracting the features [13], [14] and then form a new bullying set. In this current work we only

design the model for English literature words and in future we want to derive some more new methods which can be applied on multi languages.

2. LITERATURE SURVEY

The literature review is the most crucial phase in the software development process. Before designing a new application or model, it is vital to calculate the timeline, budget, and company's strength. Once all of these elements have been reviewed and approved, application development can commence. The literature review focuses mostly on all of the past work performed by multiple users, as well as the benefits and drawbacks of those models. This literature review serves primarily to establish the list of resources that will be used to create this suggested application.

2.1. MOTIVATION

A.M. Kaplan and Michael Haenlein[15] Proposed that the concept of social media is top of the agenda for many business executives today. Decision makers, as well as consultants, try to identify ways in which firms can make profitable use of applications such as social media flatforms. This article intends to provide some clarification. Begin by describing the concept of social media, and discuss how it differs from related concepts such as web 2.0 and user generated content. Finally, present 10 pieces of advice for companies which decide to utilize social media.

K. Reynolds, A. Kontostathis and L. Edwards[16] reported a language-based system for identifying online bullying in a small sample of Form spring data, we were able to correctly identify 78.5% of the posts that contain cyberbullying by keeping track of the proportion of swear and insult terms within a post. Our findings show that while our features can detect cyberbullying in Form spring postings to a respectable degree, there is still much space for advancement in this timely and significant use of machine learning to web data.

J. HANI AND M. NASHAAT[17] ET. AL., SUGGESTED USES MACHINE LEARNING TO IDENTIFY CYBERBULLYING. WE USED TFIDF AND SENTIMENT ANALYSIS ALGORITHMS TO EXTRACT FEATURES AND TWO CLASSIFIERS, SVM AND NEURAL NETWORKS, TO EVALUATE OUR MODEL. THE CATEGORIZATIONS WERE ASSESSED USING SEVERAL N-GRAM LANGUAGE MODELS.

3. PROPOSED SVM ALGORITHM FOR CYBER BULLYING DETECTION IN ONLINE SOCIAL NETWORKS

For feature learning, the suggested system made advantage of expert knowledge. The suggested system employs an ML-approach to categorise the semantic meanings of posted messages and attempts to combine BoW features, sentiment features, and contextual data to train an online harassment support vector machine.

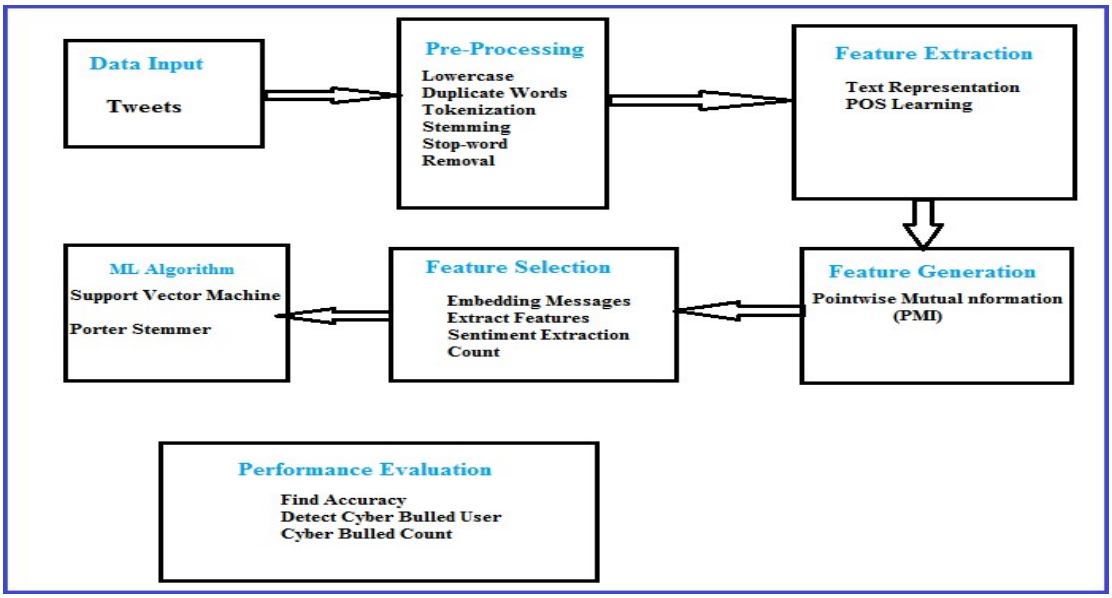


FIGURE 2. REPRESENT THE FLOW OF PROPOSED MODEL FOR CYBERBULLYING DETECTION

There are numerous machine learning techniques for text classification with machine learning platforms in general. Support vector machines, which we employ in our current application, are one of the best. Support vector machines are algorithms that find the best decision border between vectors that belong to a particular group (or category) and vectors that do not. To transform text messages into vectors, we can assume vectors on any type of data. Vectors are thought of as lists of integers that represent a collection of coordinates in space. By dividing the complete list into two halves, the SVM algorithm performs best in classifying text messages into two or more subgroups. We attempt to apply the SVM method to text categorization issues and achieve excellent results.

STATISTICS OF THE DATASET

Total number of Conversations	1800
Number of cyberbullying	900
Number of non-Cyberbullying	900
Number of distinct words	6300
Number of token	54675
Maximum Conversation size	865 Characters
Minimum Conversation size	66 Characters

3.1. STEP BY STEP PROCEDURE

The Step-by-Step Procedure of SVM Classification Algorithm in order to detect Cyber bulled Messages from Online Social Networks are as follows:

STEP 1: CHOOSE THE MODEL

Initially we need to choose which type of model we want to design with SVM Algorithm. Theremainly three types of models are such as:

A) Classifier Model

B) Extractor Model

C) Workflow Model

In our current work, we try to use SVM as classification model to classify the tweets into cyberbulled or not.

STEP 2: CHOOSE CLASSIFICATION TYPE

At this stage we need to choose the type of classification task what we would like to perform. Based on the topic we need to choose the type of classification type. Here in our current work, we try to classify the text messages and find out cyber bulled and normal messages. The content classification is divided into three types such as:

A) Topic Classification

B) Sentiment Analysis

C) Intent Classification

STEP 3: IMPORT THE DATA

In this stage, we try to import necessary data which is required for the current application. In thiscurrent application we try to import the tweets data either collected from well-known social media or directly from social user profiles. Here the data can be imported either from excel file, CSS file or data library.

STEP 4: LOAD THE INPUTS FOR FEATURE EXTRACTION

In this stage we try to load the input by taking some sample keywords such as cyber bulled words and these words are matched with feature extraction. Here we try to identify these features extracted from sentences and then try to load these features to match with the corresponding Bag of Words(BOW) which is present in the database.

STEP 5: TRAIN THE MODEL

In this stage we try to train the model by using SVM Classifier and then try to divide the messages into parts and then verify the message whether come under cyber bulled category or normal category. Here we try to train the model with more than hundreds of cyber bulled words and then make the model ready to classify any type of messages.

STEP 6: TEST THE MODEL

In this stage we try to test the model by giving some sample inputs which contain both cyber bulled words and also normal meaning and then test the efficiency of our current application by using SVM Algorithm.

3. IMPLEMENTATION PHASE

The implementation stage is where the theoretical design is translated into a programmed format. At this point, the programme will be divided into modules and coded for deployment. The front end of the application uses JSP, HTML, and Java Beans, and the back-end data base is My SQL. The application is broken into the five parts listed below. They are listed below.

1. Network Construction Module
2. User Registration Module
3. Bag of Words Construction Module
4. Identify the Bullying Feature Set Module
5. Cyber Bulled Detection Module

Now let us discuss about each and every module in detail as follows:

1) NETWORK CONSTRUCTION MODULE

In this module, we must first build a network with a single administrator and several users. Where the administrator has the option to add a group of words depending on certain category to each BoW database. Every single word must be individually entered into the database by the administrator. A word should not be added to another category after being added to one before. Therefore, the admin should perform this step as a must before adding new words to the database. Additionally, admin has the option to authorize each user upon registration. Only the user that the admin activated can access his profile by logging into the website. Users who are not permitted cannot, under any circumstances, access their own accounts.

2) USER REGISTRATION MODULE

In this module the users need to register with all their basic details and once the user get registered then they need to request for activation by the admin or server. Once the server activate the users then only the user can able to login into their account and try to perform the operations such as: Login into their account, Send friend request and receive response, Add Post, Receive Comments and Send replies and so on.

3) BAG OF WORDS CONSTRUCTION MODULE

Here in this section we try to construct a bag of words(BOW) model in which the admin try to add all the set of abused or vulgar words into the database. Here the admin try to categorize the cyber bulled communication into 5 categories such as Vulgar, Hate, Violence, Offensive and Sex.Based on the type of word the admin try to add those words into that appropriate category and finally try to maintain the words in the BoW for classification using Support Vector Machine Classification Algorithm.

4) IDENTIFY THE BULLYING FEATURE SET MODULE

The characteristics of bullying have a significant influence and should be carefully considered. The processes for building the bullying feature set Zb are provided here, with the first layer and the other layers being discussed separately. Word embedding and expert knowledge are employed for the first layer. Discriminative feature selection is carried out for the other layers.

5) CYBER BULLED DETECTION MODULE

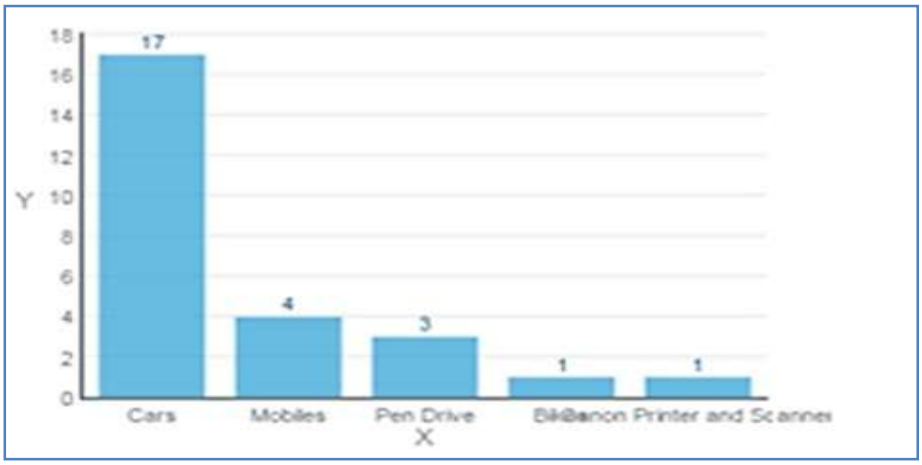
Here we try to apply Support Vector Machineand PTStemmer method to divide the messages into parts and then classify which message contain cyber bulled content and which messages contain normal content. Based on the words which are present in the message the system try to identify the cyber bulled user and normal users separately.Finally the admin try to maintain all the log information about cyber bulled user and normal users and then present that information for end users.

5. EXPERIMENTAL RESULTS

In this section we try to design our current model using JAVA as programming language and taking MY-SQL as storage database. Here we try to construct the application in the form of online social networks simulated with same functionalities and then test the current application on that current model.In order to test the efficiency using SVM Algorithm we try to maintain Bag of Words(BOW) model and then add all set of cyber bullying words inside the database.

BAG OF WORDS

CYBER BULLED USERS GRAPH



From the above figure we can clearly find out the set of cyber bulled users who try to post

cyberbulled messages on several post.

6.CONCLUSION

In this current work we finally concluded our model can give accurate results in order to identifythe list of cyber bulled users and normal users in an accurate manner. These types of messages are mainly posted on user walls in order to harass teens, preteens other children by posting these types of offensive messages. In this current application we try to design a model by using SVM and PT Stemming techniques in order to classify the tweets based on some BOW model and then find out which messages come under cyber bulled and which messages come under normal mode. Here we classify the cyber bulled conversation into five categories that are available in the literature like hate, vulgar, offensive, sex, and violence. By conducting various experiments on our proposed model by taking online social networks as simulated manner, our simulation results clearly state that our proposed model is very efficient in identifying the cyber bulled messages.

REFERENCES

1. 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.
2. K. Dinakar, R. Reichart, and H. Lieberman, Modeling the detection of textual cyberbullying in the social mobile web, 2011.
3. V. Nahar, X. Li, and C. Pang, "An effective approach for cyberbullying detection," Communications in Information Science and Management Engineering, 2012.
4. 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.
5. 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.
6. 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.
7. P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," Unsupervised and Transfer Learning Challenges in Machine Learning, Volume 7, p. 43, 2012.
8. M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," arXiv preprint arXiv: 1206.4683, 2012.
9. M. Jayanthi Rao, R. Kiran Kumar, Follicle Detection in Digital Ultrasound Images Using BEMD and Adaptive Clustering Algorithms, Lecture Notes in Mechanical Engineering, 651-659, 2020.
10. M. Jayanthi Rao, R. Kiran Kumar., J. Harikiran, Method for follicle detection and ovarian classification in digital ultrasound images using geometrical features, Journal of Advanced Research in Dynamical and Control Systems, 11(2), 1249–1258, 2019.

11. M. Balakrishna, M. Ramanaiah, B. Ramakrishna, M. Jayanthi Rao and R. Neeraja, Inductively Coupled Plasma-Mass Spectroscopy: Machine Learning Screening Technique for Trace Elemental Concentrations in Hemidesmus Indicus. 65(1), 4431-4445, 2022.
12. M. Jayanthi Rao, P. Prasanthi, P. Suresh Patnaik, M. Divya, J. Sureshkumar and M. Ramanaiah, Forecasting systems for heart disease using advanced machine learning algorithms. Int. J. Food and Nut. Sci., 11(7), 1257-1268, 2022.
13. M. Jayanthi Rao, M. Divya, M. Ratnan Mohitha, P. Prasanthi, S. Paparao, M. Ramanaiah, Analyzing the effectiveness of convolutional neural networks and recurrent neural networks for recognizing facial expression. Int. J. Food and Nut. Sci., 11(7), 1269-1282, 2022.
14. 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.
15. 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.
16. T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," Machine learning, vol. 42, no. 1-2, pp. 177–196, 2001.
17. Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," Pattern Analysis and MachineIntelligence, IEEE Transactions on, vol. 35, no. 8, pp. 1798–1828, 2013.
18. B. L. McLaughlin, A. A. Braga, C. V. Petrie, M. H. Moore et al., Deadly Lessons:: Understanding Lethal School Violence. National Academies Press, 2002.
19. A.M. Kaplan and M. Haenlein, Users of the world, unite! The challenges and opportunities of social media, Business Horizons, 53, 59-68, 2010.
20. K. Reynolds, A. Kontostathis and L. Edwards, Using Machine Learning to Detect Cyberbullying, 2011 10th International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, USA, 2011, pp. 241-244.
21. J. Hani1, M. Nashaat, M. Ahmed, Z. Emad, E. Amer and A. Mohammed, Social media cyberbullying detection using machine learning, international journal of advanced computer science and applications, 10(5), 703-707, 2019