Volume 9, No. 1, January-February 2018



International Journal of Advanced Research in Computer Science

**RESEARCH PAPER** 

Available Online at www.ijarcs.info

# CYBERBULLYING REVELATION IN TWITTER DATA USING NAÏVE BAYES **CLASSIFIER ALGORITHM**

Vandana Nandakumar MTech Student Department of Information Technology, CUSAT Kerala, India

Binsu C. Kovoor Assistant Professor Department of Information Technology, CUSAT Kerala, India

Sreeja M. U. **Research Scholar** Department of Information Technology, CUSAT Kerala, India

Abstract: Cyberbullying can be visualized as a potential issue affecting children and all categories of people. One demanding concern is effective representation for learning of content messages. The proposed system deals with cyberbullying revelation in email application using Naive Bayes Classifier Algorithm. The Classification Algorithm is a baseline method for content classification; the method of analyzing documents as relating to one classification or the other with word prevalence as features. The technique deals with the identification and filtering of spam words. The denoised messages are classified with the help of Naive Bayes Classifier Algorithm. The messages are processed under feature set extraction method. The feature probabilities are found out using Naive Bayes Classifier Algorithm .The efficiency factor is compared among the two algorithms, Naive Bayes Classifier Algorithm and Support Vector Machine and a graph is plotted. Comparison on the basis of precision factor is also done with the fact that the probabilities for each feature set are calculated independently from the twitter dataset and can evaluate the performance by predicting the output variable.

Keywords: e-mail; twitter; spam; Naïve Bayes classifier; support vector machine

#### INTRODUCTION I.

Cyberbullying can be explained as the technique of bullying a person or a character with the advent of internet technologies [1, 2]. Cyberbullying Revelation is a technique which is implemented in email application with the help of Naive Bayes Classifier Algorithm. The algorithm uses the classification method in order to categorize the messages which are having spam words. The proposed system categorizes the emails which are having cyberbullying content versus the emails which are not having cyberbullying content. The denoised value for each word is calculated by grouping the messages and is done by using the classification method. The feature set extraction technique is done for each twitter message which selects data attributes that best characterize a predicted variable [3]. The feature probabilities are calculated using the naïve Bayes classifier algorithm. The proposed system alerts the sender if they are using any vulgar languages and the messages are redirected as such. The technology provides hope to concerned parents and is a sign of relief to all categories of people who are affected.

Cyberbullying problem is also occurring in school premises. Teachers try to make their students aware about cyberbullying practices and its negative effects [4]. The proposed system uses Word Embedding Technique as its framework which obtains the bullying characters automatically. Finally these specific alterations make the new feature space more selective and thus facilitate Cyberbullying Revelation.

Machine Learning Techniques can make automatic revelation of bullying messages in social media networks possible and will create a clear social environment [5]. Data mining can be explained as an absolute subfield of computer science. It is the technique of analyzing patterns in large data sets including techniques at the intersection of machine

© 2015-19, IJARCS All Rights Reserved

learning methodologies and database system techniques. The main aim of data mining technology is to extract information from a large data set and convert it into useful methods so that it can be used for further extensions [6]. Cyberbullying can be compared with traditional bullying but the latter encompasses a range of public areas like college, school with the victim often experiencing it. The predator is the first person who is capable of molesting the victim in both cases. But, cyberbullying is done with the help of online methods where physical presence of victim is not a relevant factor.

In [7], a real time system has been implemented which minimizes the amount of bullying rather than detecting and preventing them. An analysis of common users in social networks is done in [8]. The posting activity of common users and relation with negativity is examined here. Negativity in anonymous messages is also analyzed. The accuracy of predicting the level of cyberbullying attack using classification methods is studied in [9]. A Facebook watchdog application is developed in [10] that make use of image analysis, social media analytics, and text mining techniques to detect cyberbullying activity.

The rest of the paper is organized as follows. Section 2 describes the methodology adopted for the proposed system. Experimental results based on different parameters are described in section 3. Conclusions along with future enhancements possible are detailed in section 4.

# **II. METHODOLOGY**

The various modules for the proposed system are GUI designing, Training dataset, classification and analyzing the twitter messages for the presence of spam content. The classification technique is implemented using Naïve Bayes Classifier Algorithm. The Revelation consists of the following steps. The primary step is to accept data sets from numerous online network sites. The dataset deals with the twitter messages and the wordcount for each message are calculated by grouping the messages. In the layers, the values of words are found out using the wordcount. The datasets include the comments that are posted by users, images, video clips on networking sites, social networks network, etc. Using Twitter API, tweets can be easily analyzed and verified. The next stage is Preprocessing of Data where the dataset is processed so that data contains only required information. Subsequently, denoised words are analyzed. The removal of whitespaces and stop words can be considered as a way of data preprocessing after which tokenization and lemmatization occurs. Various other methods are also there to clear the datasets. The final stage deals with the classification of data. The classification is done using classifiers and a classification algorithm is considered as a part of it. Messages are classified into a set of classes. The probabilities are found out using the feature set extraction method. A message with a value less than the threshold value which is 1.5 is considered as a cyberbullied message. Data is classified into positive and negative instances by comparing text content having cyberbullying content with data which has no admissible cyberbullying content. After processing the messages the denoised values are found out. The message classification is done using Naïve Bayes Classifier Algorithm. The denoised values are trained using the Naïve Bayes Classifier. Before a new data is classified, the classification algorithm is in need of training sets for training a classifier and thus in turn facilitates potent and discriminative representation of learning of text messages. A classifier is in need of labeled examples. So that it could analyze the label of an input and this learned classifier is then used to validate a bullying message. Numerous algorithms and techniques are classification of data like used for the Bag-of-Words(BoW),Support Vector Machine(SVM),Naive Bayes Classifier Algorithm etc. Each technique deals with both the positive and negative versions of cyberbullying aspects.

Data Preprocessing improves the data set so that the dataset includes only required information. The various stages in data preprocessing are:

#### A. Tokenization

Tokenization is the process of distributing large set of unstructured messages into a small subset of tokens. These are classified with the help of various aspects such as white spaces, punctuation marks and is categorized as phrases, sentences etc.

# B. Stop words Removal

The most common words that are used in a text are words such as 'a' 'and' 'are' and so on. The main drawback of these words are that such words only contribute very little meaning to text and aids only a very small value in classifying text. Stop words removal from messages results in more convenient recognition of text in further steps.

# C. Replacement of Special Characters

The method deals with the replacement of special characters like '@' with its exact word 'at'. In tweets, this step has larger importance because of the extensive occurrences of special symbols.

# D. Stemming and Lemmatization

This method finds the root of a single word and is considered as a heuristic technique it simply abridges prefixes as well as suffixes. It uses word-based approach in order and is so called dictionary based approach. Lemmatization method can be considered as the further extension of stemming technology. For the grammatical categorization of characters to get the base method of a single word called lemma, this technique is widely used. One of the algorithms which are broadly used for this purpose is Porters algorithm and can be more adequately used and is more specific.

#### E. Coreference Resolution

Coreference Resolution is the method of analyzing all expression that focuses to the same entity in text content. One of the relevant co-referential equipment in a written document which can be considered is repetition and it could make stringcomparison characteristics more relevant to all co-reference resolution methods. This is one of the most promising steps in advanced Natural Language Processing methods that comprise minor language analysis such as a text content summarization, answering to a particular question and retrieval of the correct information.

The flowchart of the proposed system is shown in fig.1.



Figure 1. Flowchart of proposed system

The main limitation in the research related to cyberbullying is that the bullying begins when there is rivalry between the bullying person and victims in real life. Since real life incidents are difficult to be deducted from social networks, the main reason for bullying is difficult to be analyzed. But there are some researches depicting the common reasons to bully a person like love failure, envy, etc. The first step here is to identify the popular newsmakers for a week and term frequency methods can be used as an output. The next stage is to identify the news that made the newsmakers popular and fetch the corresponding news and cluster them. Later, extract the comments and posts related to that news using the similarity index. The last method is to identify the cyberbullying terms and negative words so it can be more extensively recognized. This can be considered as the challenging parts in cyberbullying revelation .A graph can be introduced after the classification of data by considering the data which definitely has cyberbullying content with the data which has no significant cyberbullying content.

#### **III. RESULTS AND DISCUSSION**

The results and discussion deals with the analysis of comparison graph on the basis of precision and run time complexity. When the algorithm runs it takes each twitter messages and breaks it down into individual words. Each word is compared to the words in the bully dictionary. If it matches any of the words then it is added to the precision value. The precision values are compared with the threshold value and if it is less than the threshold value it is considered as cyberbullied message. Finally, the algorithm adds all the twitter messages having precision values less than the threshold. Fig. 2 compares precision values obtained using Naïve Bayes classifier and support vector machine.



Figure 2. Comparison based on Precision

Using the precision factor, the probabilities for each feature set are calculated independently from the twitter dataset and performance is evaluated by predicting the output variable. The Feature extraction selects the data attributes that best characterize a predicted variable. It can be done more conveniently using the precision factor analysis of Naïve Bayes Classifier Algorithm. The identification and separation of segments can also be done more effectively using this technique.

The time complexity graph in fig.3 shows that Naïve Bayes Classifier is having lower run time complexity. Run time complexity is calculated as the absolute difference between the time before the algorithm starts and time at which the algorithm finishes running. It is calculated in milliseconds.





The rapid increase of social networks has shown a consistent growth in cyberbullying activities. Cyberbullying has become a major social problem. Cyberbullying has become an important area of research due to its impact on society. Various researches try to recognize the reason of cyberbullying and its aftereffect. But only a few try to enhance software to prohibit cyberbullying. Robust and selective representation of learning of text messages is crucial for consistent detection system [11]. Machine Learning representation and authentication makes automatic revelation of bullied messages in online media possible and ensures building a relevant and clear social media environment. The Email based cyber stalking is also a huge problem. Email based cyber stalking detection involves two phases; the first is to analyze and detect cyber stalking emails and the second phase is to verify the proof for finding out the cyber stalkers as a prohibition and detection mechanism [12].

Cyberbullying is a major problem that is happening on the Internet. Internet is a convenient environment for bullying the most vulnerable community. The main procedure for cyberbullying revelation is web based mining technologies. An acceptable level of precision can be acquired with the proposed system and the results are promising. For social network medias when the level of precision is met it can be copied and verified in software and can be done for miscellaneous implementation phases. The proposed system can be modified for cyberbullying revelation in Non-English applications. The productive (effective) visualization can be promisingly met with the help of simpler notification about the occurrences of bullies. After identifying cyberbullying problem, relevant measures should be taken to prohibit further molestation of victim, preventing the spread of vulgar and immense messages. Analyzing, verifying and providing additional information helps the victims to take measures for getting rid of the problem.

#### V. REFERENCES

- [1] Weider D. Yu, Maithili Gole, Nishanth Prabhuswamy, Sowmya Prakash, Vidya Gowdru Shankaramurthy, "An Approach To Design and Analyze the Framework for Preventing Cyberbullying", IEEE International Conference on Services Computing, 2016, pp.864-867, doi: 10.1109/SCC.2016.125.
- [2] Samaneh Nadali, Masrah Azrifah Azmi Murad, "A Review of Cyberbullying Detection :An Overview", International Cmiference on Intelligent Systems Design and Applications (ISDA), 2013, pp.325-330, doi:10.1109/ISDA.2013.6920758.
- [3] Amrita Mangoankar, Allenoush Hayrapetian, "Collaborative Detection of Cyberbullying Behaviour In Twitter Data", 2015, IEEE Conference, pp.612-616, doi:10.1109/EIT.2015.7293405.
- [4] Silvana Tambosi, Gustavo Borges, Vanessa Mondin, Maria José Domingues, "Cyberbullying: Concerns of Teachers and School Involvement", International Conference on Complex, Intelligent, and Software Intensive Systems Cyberbullying, 2015, pp.416-420, doi: 10.1109/CISIS.2015.92
- [5] Kelly Reynolds, April Kontostathis, Lynne Edwards, "Using Machine Learning to Detect Cyberbullying", International Conference on Machine Learning and Applications, 2011, pp.242-244, doi:10.1109/ICMLA.2011.152.
- [6] Ana Kovacevic, "Cyberbullying Detection Using Web content Mining", Telecommunications forum TELFOR, 2014, pp. 25-27, doi:10.1109/TELFOR.2014.7034560.
- [7] Tibor Bosse, Sven Stam ,"A Normative Agent System To Prevent Cyberbullying", IEEE/WIC/ACM International

Conferences on Web Intelligence and Intelligent Agent Technology, 2011, pp. 425-430, doi: 10.1109/WI-IAT.2011.24.

- [8] Homa Hosseinmardi Shaosong Li, Zhili Yang and Qin Lv, Rahat Ibn Rafiq Richard Han and Shivakant Mishra," A Comparison of Common Users across Instagram and Ask.fm to Better Understand Cyberbullying", IEEE International Conference on Big Data and Cloud Computing, 2014, pp. 355-362, doi: 10.1109/BDCloud.2014.87.
- [9] Nektaria Potha, Manolis Maragoudakis," Cyberbullying Detection using Time Series Modeling", IEEE International Conference on Data Mining Workshop, 2014, pp.373-382, doi:10.1109/ICDMW.2014.170.
- [10] Marlies Rybnicek, Rainer Poisel and Simon Tjoa, "Facebook Watchdog:A Research Agenda For Detecting Online Grooming And Bullying Activities", IEEE International Conference on Systems, Man, and Cybernetics, 2013, pp.2854-2859, doi: 10.1109/SMC.2013.487.
- [11] Rekha Sugandhi, Anurag Pande, "Method for Detection of Cyberbullying:A Survey", International Conference on Intelligent Systems Design and Applications (ISDA), 2015, pp.173-177, doi: 10.1109/ISDA.2015.7489220.
- [12] Zinnar Ghasem, Ingo Frommhoz, "A Hybrid Aprroach to Combat Email Based Cyberstalking", FGCT International Conference on Future Generation Communication Technologies, 2015, pp.106-111, doi:10.1109/FGCT.2015.7300257