International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 07 | July - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Cyber Bullying Detection for Twitter Using MI Classification Algorithms

Mamatha B H [1], Dr. Geetha M [2]

[1]Student, Department of MCA, BIET, Davangere,
[2]Associate professor, Department of MCA, BIET, Davangere.

ABSTRACT: Cyber bullying is a major problem encountered on internet that affects teenagers and also adults. It has led to mis- happenings like suicide and depression. Regulation of content on Social media platforms has become a growing need. The following study uses data from two different forms of cyber bullying, hate speech tweets from Twittter and comments based on personal attacks from Wikipedia forums to build a model based on detection of cyber bullying in text data using Natural Language Processing and Machine learning. Three methods for Feature extraction and four classifiers are studied to outlinethe best approach. For Tweet data the model provides accuracies above 90% and for Wikipedia data it gives accuracies above 80%.

Keywords:Wikipedia,Twitter,NaturalLanguagePr ocessingandCyberbullying.

1. INTRODUCTION

Introduction With therapid growth of social media, users especially adolescents are spending significant amount of time on various social networking sites to connect with others, to share information, and to pursue common interests. It has been found that 70% of teens use social media sites on a daily basis and nearly one in four teen shit their favorite social media sites 10 or more times a day. 19% of teens report that someone has written or posted mean or embarrassing things about them on social networking sites. As adolescents are more likely to be negatively affected by biased and harmful contents than adults, detecting online offensive contents to protect adolescent's online safety becomes an urgent task. To address concerns on children, access to offensive contents over Internet, administrators of social media often manually review online contents to detect and delete offensive materials; however, manually reviewing is labour intensive and time consuming.

Some automatic contents filtering software packages, such as Appen, eblaster, iambigbrother, Internet SecuritySuite etc, have been developed to detect and filter online offensive contents. Most of them simply blocked webpages and paragraphs that contained dirty words. These word based approaches not only affect the readability and usability of web sites, but also fail to identify subtle offensive messages.

Sites for social networking are excellent tools for communication within individuals. Use of social networking has become widespread over the years, though, in general people find immoral and unethical ways of negative stuff. We see this happening between teens or sometimes between young adults. One of the negative stuffs they do is bullying each other over the internet. In online environment we cannot easily said that whether someone is saying something just for fun or there may be other intention of him. Often, with just a joke, "or don't take it so seriously," they'll laugh it off. Cyber bullying is the use of technology to harass, threaten, embarrass, or target another person. Often this internet fight results intoreal life threats for some individual. Some people haveturned tosuicide. It is necessary to stop such activitie at the

beginning. Anyactionscould betaken toavoid this for exampleif an individual'stweet/post is found offensive then maybe his/her account can be terminated or suspended for a particular period.

I. RELATED WORK

1) In recent years, users are widely intended to express and share their opinions over the Internet. However, due to the characters of social media, it appears negative use of social media. Cyberbullying is one of the abuse behaviors in the Internet as well as a very serious social problem.
2) Under this background and motivation, it can help to prevent the happen of cyberbullying if we can develop relevant techniques to discover cyberbullying in social media. Thus, in this paper we propose an approach based on social networks analysis and data mining for cyberbullying detection. In the approach, there are three main techniques for cyberbullying discovery will be studied, including keyword matching technique, opinion mining and social network analysis. In addition to the approach, we will also discuss the experimental design for the evaluation of the performance.
3) The use of new technologies along with the popularity of social networks has given the power of anonymity to the users. The ability to create an alter-ego with no relation to the actual user, creates a situation in which no one can certify the match between a profile and a real person.
4) This problem generates situations, repeated daily, in which users with fake accounts, or at least not related to their real identity, publish news, reviews or multimedia material trying to discredit or attack other people who may or may not be aware of the attack. These acts can have great impact on the affected victims' environment generating situations in which virtual attacks escalate into fatal consequences in real life. In this paper, we present a methodology to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to a real profile within the same network by analyzing the content of comments generated by both profiles.
5) Accompanying this approach we also present a successful real life use case in which this methodology was applied to detect and stop a cyberbullying situation in a real elementary school.
6) As the size of Twitter© data is increasing, so are undesirable behaviors of its users. One of such undesirable behavior is cyberbullying, which may even lead to catastrophic consequences. Hence, it is critical to efficiently detect cyberbullying behavior by analyzing tweets, if possible in real-time. Prevalent approaches to identify cyberbullying are mainly stand-alone and thus, are time-consuming.
7) This research improves detection task using the principles of collaborative computing. Different collaborative paradigms are suggested and discussed in this paper. Preliminary results indicate an improvement in time and accuracy of the detection mechanism over the stand-alone paradigm. With the increasing use of social media, cyberbullying behavior has received more and more attention. Cyberbullying may cause many serious and negative impacts on a person's life and even lead to teen suicide.
8) To reduce and stop cyberbullying, one effective solution is to automatically detect bullying content based on appropriate machine learning and natural language processing techniques. However, many existing approaches in the literature are just normal text classification models without considering bullying characteristics. In this paper, we propose a representation learning framework specific to cyberbullying detection. Based on word

embedding's, we expand a list of pre-defined insulting words and assign different weights to obtain bullying features, which are then concatenated with Bag-of-Words and latent semantic features to form the final representation before feeding them into a linear SVM classifier. Experimental study on a twitter dataset is conducted, and our method is compared with several baseline text representation learning models and cyberbullying detection methods. The superior performance achieved by our method has been observed in this study.

9) Innovation is developing quickly today. This headway in innovation has changed how individuals cooperate in an expansive way giving communication another dimension.

10) But despite the fact that innovation encourages us innumerous parts of life, it accompanies different effects that influence people in a few or the other way. Cyberbullying is one of such effects. Cyberbullying is a wrongdoing in which a culprit focuses on an individual with online provocation and loathes which has antagonistic emotional, social and physical effects on the victim. So as to address such issue we proposed a novel cyberbullying detection method dependent on deep neural network. Convolution Neural Network is utilized for the better outcomes when contrasted with the current systems.

II. METHODOLOGY

3.1 System Architecture

We use SVM for twitter data and Random Forest classifier for Wikipedia data for identifying cyber bullying. SVM offers very high accuracy compared to other classifiers such as logistic regression, and decision trees. It is known for its kernel trick to handle nonlinear input spaces.It is used in a variety of applications such as face detection, intrusion detection, classification of emails, news articles and web pages, classification of genes, and handwriting recognition.

SVM is an exciting algorithm and the concepts are relatively simple. The classifier separates data points using a hyper plane with the largest amount of margin. That's why an SVM classifier is also known as a discriminative classifier. SVM finds an optimal hyperplane which helps in classifying new data points. Random forests are a supervised learning algorithm. It can be used both for classification and regression. It is also the most flexible and easy to use algorithm. A forest is comprised of trees. It is said that the more trees it has, the more robust a forest is. Random forests creates decision trees on randomly selected data samples, gets prediction from each tree and selects the best solution by means of voting. It also provides a pretty good indicator of the feature importance.

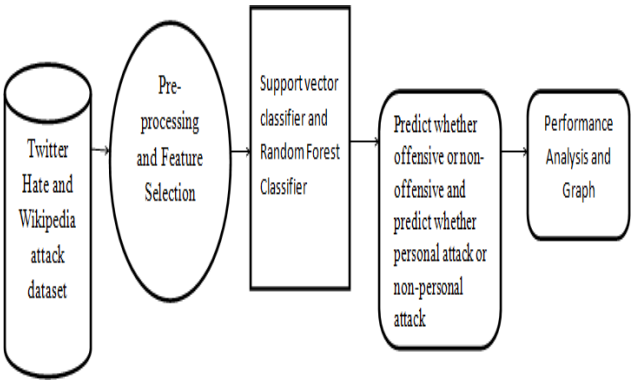

Fig. 3.1.1 Architecture of our proposed System

3.2 Dataflow Diagram

Figure 2 depicts DFD is also called as bubble chart. It is a simple graphical formal is math

At can be use do to represent a systemin terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

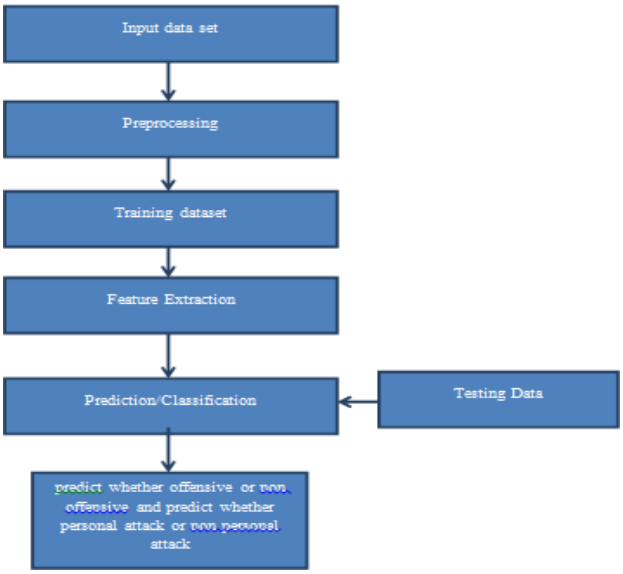

Fig.3.2.1 Dataflow Diagram

3.3 DATASET USED

The dataset consists of 115864 individual

Therefore 4 columns in the dataset, which are described below

1. Review Id: unique id
2. comment: comment about Wikipedia titles
3. yearn: yearn of comment
4. attack: Personal attack or non-persona ln attack

3.4 DATA PRE-PROCESSING

Dataset is pre-processed by removing null values and error values. The training phase can be summarized as follows:

1) Extract features from the pre-processed data set
2) Traina SVM for twitter datat set and Random Forest classifier for Wikipedia data set using this feature set.

Theoutputofthetrainingphaseisatrainedclassifier capableofpredictingclassificationpersonalattack sandoffensive comments.

The performance of the trained classifier can be valu at reducing measures like accuracy, sensitivity and specifi city.

3 .5 ALGORITHAM USED

The Random Forests Algorithm

It works in four steps:

- Select random samples from a given dataset.
- Construct a decision tree for each sample and get a prediction result from each decision tree.
- Perform a vote for each predicted result.

Select the prediction result with the most votes as the final prediction

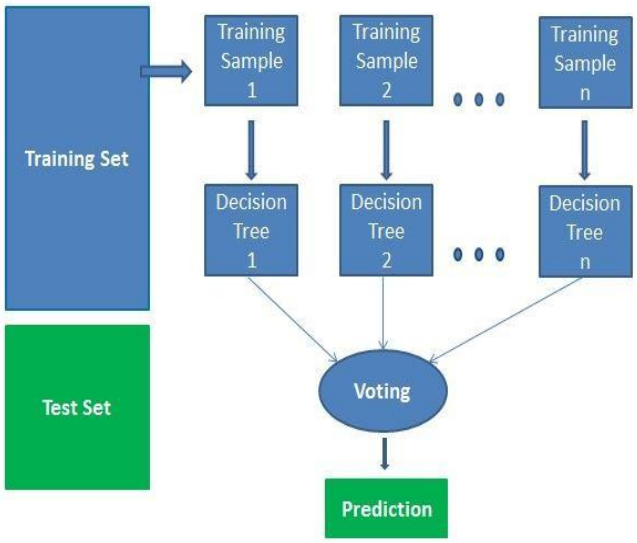

Fig.3.5.1Random forest classifier working Diagram

Let's understand the algorithm in layman's terms. Suppose you want to go on a trip and you would like to travel to a place which you will enjoy. So what do you do to find a place that you will like? You can search online, read reviews on travel blogs and portals, or you can also ask your friends.

Let's suppose you have decided to ask your friends, and talked with them about their past travel experience to various places. You will get some recommendations from every friend. Now you have to make a list of those recommended places. Then, you ask them to vote (or select one best place for the trip) from the list of

recommended places you made. The place with the highest number of votes will be your final choice for the trip.

In the above decision process, there are two parts. First, asking your friends about their individual travel experience and getting one recommendation out of multiple places they have visited. This part is like using the decision tree algorithm. Here, each friend makes a selection of the places he or she has visited so far.

The second part, after collecting all the recommendations, is the voting procedure for selecting the best place in the list of recommendations. This whole process of getting recommendations from friends and voting on them to find the best place is known as the random forests algorithm.

It technically is an ensemble method (based on the divide-and-conquer approach) of decision trees generated on a randomly split dataset. This collection of decision tree classifiers is also known as the forest. The individual decision trees are generated using an attribute selection indicator such as information gain, gain ratio, and Gini index for each attribute. Each tree depends on an independent random sample. In a classification problem, each tree votes and the most popular class is chosen as the final result. In the case of regression, the average of all the tree outputs is considered as the final result. It is simpler and more powerful compared to the other non-linear classification algorithms.

IV. RESULTS

The figure depictsthe output screen of the proposed system. It is the homepage in which we get login button. Weused jango framework for fron-end. Once we start the django server we get this page in our local host port.

Fig.4.1.1 Home Page

Home page got two separate buttons for logging twitter data set and Wikipedia data set cyberbullying identification.

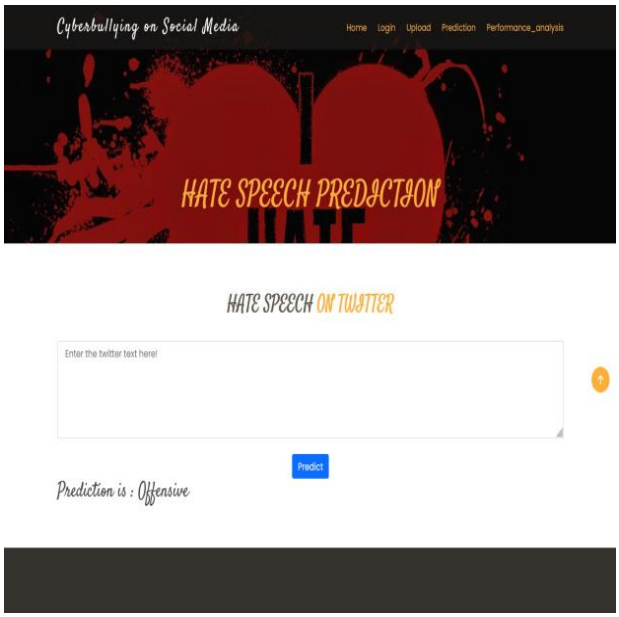

Fig.4.1.2 Predicted resul to the input comment.

Figure4.1.2shows the predicte dresult for the twitte comment. Its shows the commentis offensive.

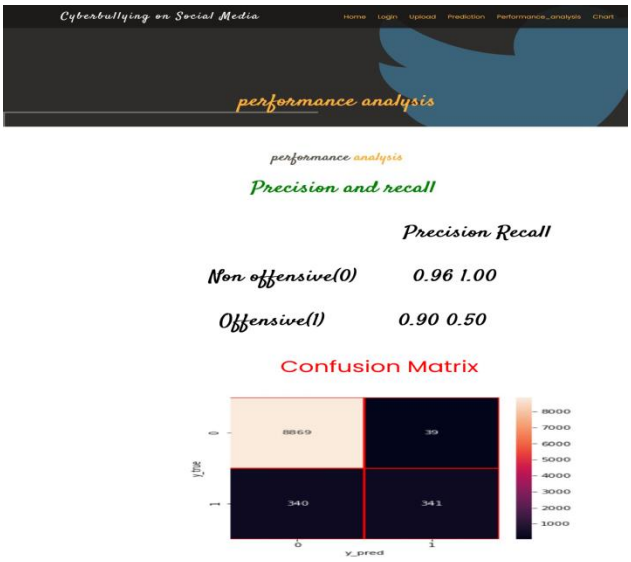

Fig.4.1.3 Performance analysis of SVM for twitter data cyber bullying attack identification

Figure 4.1.3portrays the perfromance of our created cycberbullying discoveryfor twitter information. Ourmodel hasexactness of 96% with SVM calculation. Likewise it shows Derived Confusion Matrix of the System for SVM

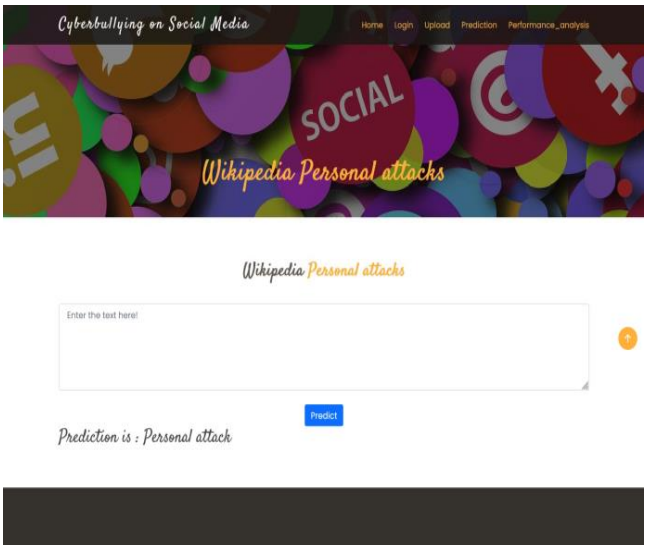

Fig.4.1.4 Predicted outcome of the input Wikipedia comment

Fig.4.1.5 Performance analysis of Random Forest for Wikipedia data cyberbullying attack identification

Figure 4.1.5 depicts the perfromance of ourdevelopedcycberbullying detectionforWikipedia data. Our model has accuracy of 76% with Random forest algorithm. Also it shows Derived Confusion Matrix of the System for RF algorithm.

V. CONCLUSION

Cyber bullying across internet is dangerous and leads to mis-happenings like suicides, depression etc and therefore there is a need to control its spread. Therefore, cyber bullying detection vital on socil media platforms. With viability of more data and better classified user information for various other forms of cyber-attacks Cyber bullying detection can be used on social media websites to ban users trying to take part in such activity In this paper we proposed an architecture for detection of cyber bullying to combat the situation. We discussed the architecture for two types of data: Hate speech Dataon Twitter and Personal attacks on Wikipedia. For Hate speech Natural Language Processing techniques proved effective with

accuracies of over 90 precent using basic Machine learning algorithms because tweets containing Hate speech consisted of profanity which made it easily detectable. Due to this it gives better results with Bow and Tf-Idf models rather than Word2Vec models However, Personal attacks were difficult to detect through the same model because the comments generally did not use any common sentiment that could be learned however the three feature selection methods performed similarly.

VI.REFERENCES

[1] H. Ting, W. S. Liou, D. Liberona, S. L. Wang, andG. M. T. Bermudez, "Towards the detection of cyberbullying based on social networkmining techniques,"in Proceedings of 4th International Conference on Behavioral, Economic, and SocioCultural Computing, BESC 2017, 2017, vol. 2018-January, doi:10.1109/BESC.2017.8256403.

[2] P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machinelearning for the detection of troll profiles in twitter socialnetwork: Application to a real case of cyberbullying," 2014, doi: 10.1007/978-3-319-01854-6_43.

[3] A. Mangaonkar A. Hayrapetian, and R.Raje, "Collaborative detection of cyber bullying behaviorinTwitterdata, 2015, doi:10. 1109/EIT. 2015.7293405.

[4] R. Zhao, A. Zhou and K. Mao, "Automatic de tecti of cyber bullying on socialnetworks based on bullying features,"2016, doi:10.1145/2833312.2849567.

[5] V. Banerjee, J. Telavane P. Gaikwad, and P. Vartak, "Detection of Cyber bullying Using Deep Neural Network," 2019, doi:10. 1109/ICACCS.2019.8728378.

[6] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machinel earning to detect cyber bullying," 2011, doi:10. 1109/ICMLA. 2011. 152.

[7] J. Yadav D Kumar, and D Chauhan, "Cyberbullying Detection using Pre-Trained BERT Model," 2020, doi: 10. 1109/ICESC48915. 2020. 9155700.

[8] M.Dadvar and K. Eckert, "Cyberbullying Detectionin Social Networks Using Deep Learning BasedModels; A Reproducibility Study," arXiv .2018.

[9] S.Agrawal and A.Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms,"arXiv. 2018.

[10] 10.Y.N. Silva, C.Rich, and D.Hall, "BullyBlocker:Towards the identification of cyber bullying in social networking sites,"2016,doi:10.1109/ASONAM.2016.7752420.

*****