RESEARCH ARTICLE OPEN ACCESS

Detection of Cyber Bullying on social media using Machine Learning

Mrs R Jhansi Rani MCA [1], M Narendra [2]

[1] Asst. Professor, Department Computer Application
[2] Student, Department of Computer Application
[1], [2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRAC
Cyber bullying is a major problem encountered on internet that affects teenagers and also adults. It has lead to mishappenings 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 cyberbullying, 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.
Keywords: - Cyber bullying, machine Learning, social media.

I. INTRODUCTION

Now more than ever technology has become an integral part of our life. With the evolution of the internet. Social media is trending these days. But as all the other things mis users will pop out sometimes late sometime early but there will be for sure. Now Cyber bullying is common these days. 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 into real life threats for some individual. Some people have turned to suicide. It is necessary to stop such activities at the beginning. Any actions could be taken to avoid this for example if an individual's tweet/post is found offensive then maybe his/her account can be terminated or suspended for a particular period.

So, what is cyber bullying??

Cyber bullying is harassment, threatening, embarrassing or targeting someone for the purpose of having fun or even by well-planned. Researches on Cyber bullying Incidents show that 11.4% of 720 young peoples surveyed in the NCT DELHI were victims of cyber bullying in a 2018 survey by Child Right and You, an NGO in India, and almost half of them did not even mention it to their teachers, parents or guardians. 22.8% aged 13-18 who used the internet for around 3 hours a day were vulnerable to Cyber bullying while 28% of people who use internet more than 4 hours a day were victims. There are so many other reports suggested us that the impact of Cyber bullying is affecting badly the peoples and children between age of 13 to 20 face so many difficulties in terms of health, mental fitness and their decision making capability in any work. Researchers suggest that every country should have to take this matter seriously and try to find solution. In 2016 an incident called Blue Whale Challenge led to lots of child suicides in Russia and other countries . It was a game that spread over different social networks and it was a relationship between an administrator and a participant. For fifty days certain tasks are given to participants. Initially
they are easy like waking up at 4:30 AM or watching a horror movie . But later they escalated to self harm which let to suicides. The administrators were found later to be children between ages 12-14.

II. RELATEDWORKS

H Ting and S L Wang proposed Towards the detection of cyberbullying based on social network mining techniques . In recent years, users are widely intend 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 behavior in the Internet as well as a very serious social problem. Under this background and motivation, it can help to prevent the happen of cyberbullying if can develop relevant techniques to discover cyberbullying in social media. Thus, in this project 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. R Raje and A Mangonakar proposed Collaborative detection of cyberbullying behavior in Twitter data. 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 realtime. Prevalent approaches to identify cyberbullying are mainly standalone and thus, are time-consuming. This research improves detection task using the principles of collaborative computing. Different collaborative paradigms are suggested and discussed in this project. Preliminary results indicate an improvement in time and accuracy of the detection mechanism over the stand-alone paradigm. Pooja Gaikwad and Vijay Benarjee Proposed Detection of Cyberbullying Using Deep Neural Network. Innovation is developing quickly today. This headways in innovation has changed how individuals cooperate in an expansive way giving communication another dimension. But despite the fact that innovation encourages us in numerous parts of life, it accompanies different effects that influence people in a few or the other way. So as to address such issue 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. In Existing system used an approach using keyword matching, opinion mining and social network analysis and got a precision of 0.79 and recall of 0.71 from datasets from four websites and proposed a hypothesis that a troll(one who cyberbullies) on a social networking sites under a fake profile always has a real profile to check how other see the fake profile. They proposed a Machine learning approach to determine such profiles. The identification process studied some profiles which has some kind of close relation to them. The method used was to select profiles for study, acquire information of tweets, select features to be used from profiles and using ML to find the author of tweets. 1900 tweets were used belonging to 19 different profiles. It had an accuracy of 68% for identifying author. Later it was used in a Case Study in a school in Spain where out of some suspected students for Cyberbullying the real owner of a profile had to be found and the method worked in the case. The following method still has some shortcomings. For example a case where trolling account doesnt have a real account to fool such systems or experts who can change writing styles and behaviours so that no patterns are found . For changing writing styles more efficient algorithms will be needed. In existing system proposed a collaborative detection method where there are multiple detection nodes connected to each other where each nodes uses either different or same algorithm and data and results were combined to produce results AND suggested a B-LSTM technique based on concentration. used KNN with new embeddings to get an precision of 93%. A vocabulary is not designed from all the documents. The vocabulary may consist of all words (tokens) in all documents or some top frequency tokens. Tf-Idf method is not similar to the bag of words model since it uses the same way to
create a vocabulary to get its features.

III. PROPOSED SYSTEM ARCHITECTURE

Cyberbullying detection is solved in this project as a binary classification problem where we are detecting two majors form of Cyberbullying: hate speech on Twitter and Personal attacks on Wikipedia and classifying them as containing Cyberbullying or not. Tokenization: In tokenization we split raw text into meaningful words or tokens. For example, the text “we will do it” can be tokenized into ‘we’, ‘will’, ‘do’, ‘it’. Tokenization can be done into words called word tokenization or sentences called sentence tokenization. Tokenization has many more variants but in the project we use Regex Tokenizer. In regex tokenizer tokens are decided based on rule which in the case is a regular expression. Tokens matching the following regular expression are chosen Eg For the regular expression ‘\w+’ all the alphanumeric tokens are extracted. Stemming: Stemming is the process of converting a word into a root word or stem. Eg for three words ‘eating’ ‘eats’ ‘eaten’ the stem is ‘eat’. Since all three branch words of root ‘eat’ represent the same thing it should be recognized as similar. NLTK offers 4 types of stemmers: Porter Stemmer, Lancaster Stemmer, Snowball Stemmer and Regexp Stemmer. The following project uses PorterStemmer. Stop word Removal: Stop words are words that do not add any meaning to a sentence eg. Some stop words for english language are: what, is, at, a etc. These words are irrelevant and can be removed. NLTK contains a list of english stop words which can be used to filter out all the tweets. Stop words are often removed from the text data when we train deep learning and Machine learning models since the information they provide is irrelevant to the model and helps in improving performance.

1.Common Bag of Words model takes as input of multiple words and predicts the word based on the context. Input can be one word or multiple words.
2.CBOW model takes a mean of context of input words but two semantics can be clicked for a single word. i.e. two vector of Apple can be predicted. First is for the firm Apple and next is Apple as a fruit.
In this proposed system, there are two modules. They are :

➢ Cyber Security Analyst

➢ End User

1.Cyber Security Analyst

In this Module, it Provides following functionalities:

- Login
- Train and Test Data sets with SVM and Naïve Bayes
- View Trained and Tested Accuracy
- View Cyber Bullying Predict Type Details
- Find Cyber bullying prediction ratio on dataset
- View All Remote Users
- Logout

2. End User

In this Module, it Provides following functionalities:

- Register
- Login
- Predict cyber bullying
- View Profile
- Logout

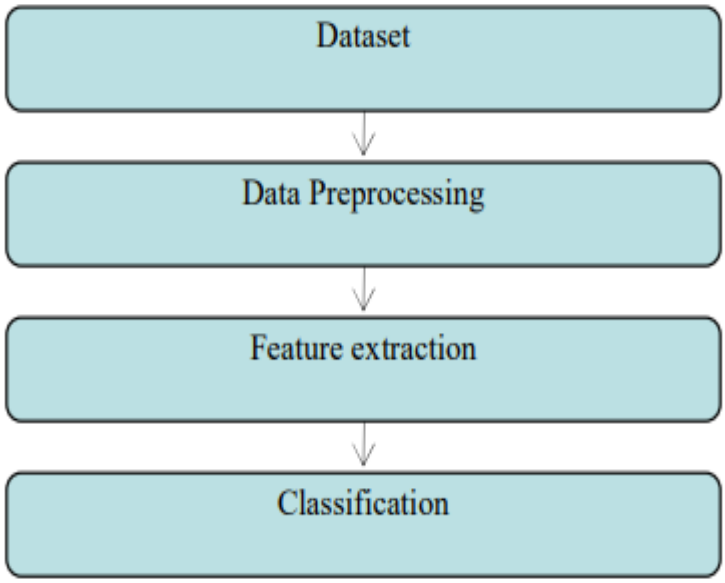

Fig.1 Proposed system architecture

IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system are shown from Fig.2 to Fig.9

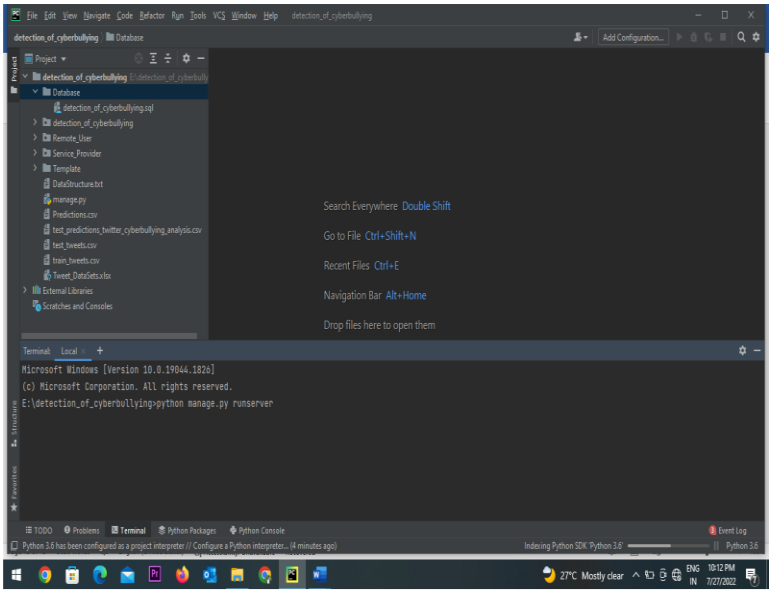

Fig.2 Running program on PyCharm

Fig.3 Admin login

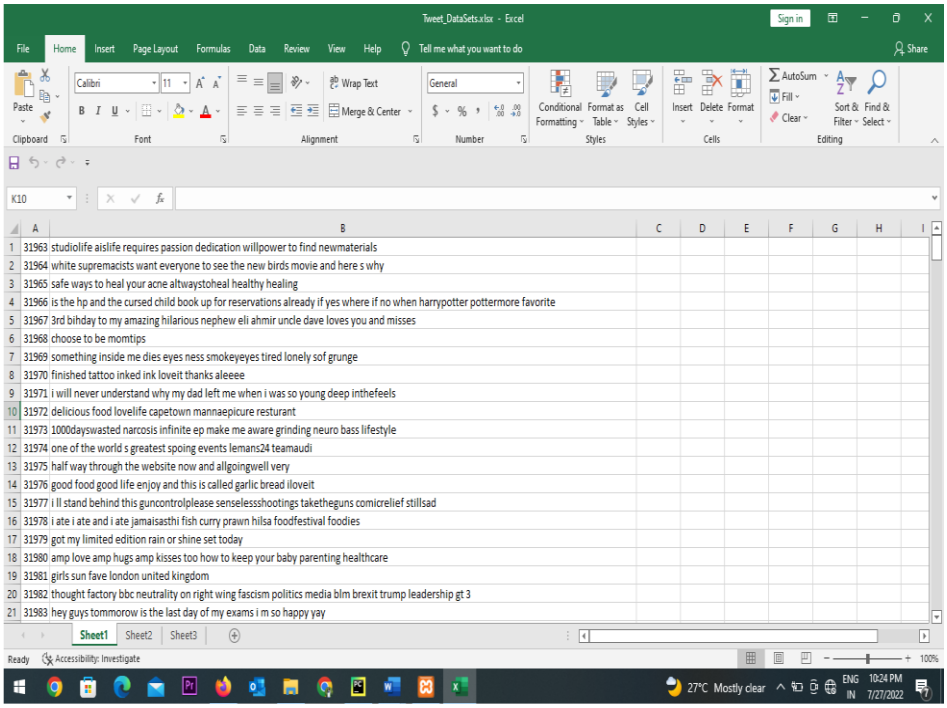

Fig.4 Tweet data set

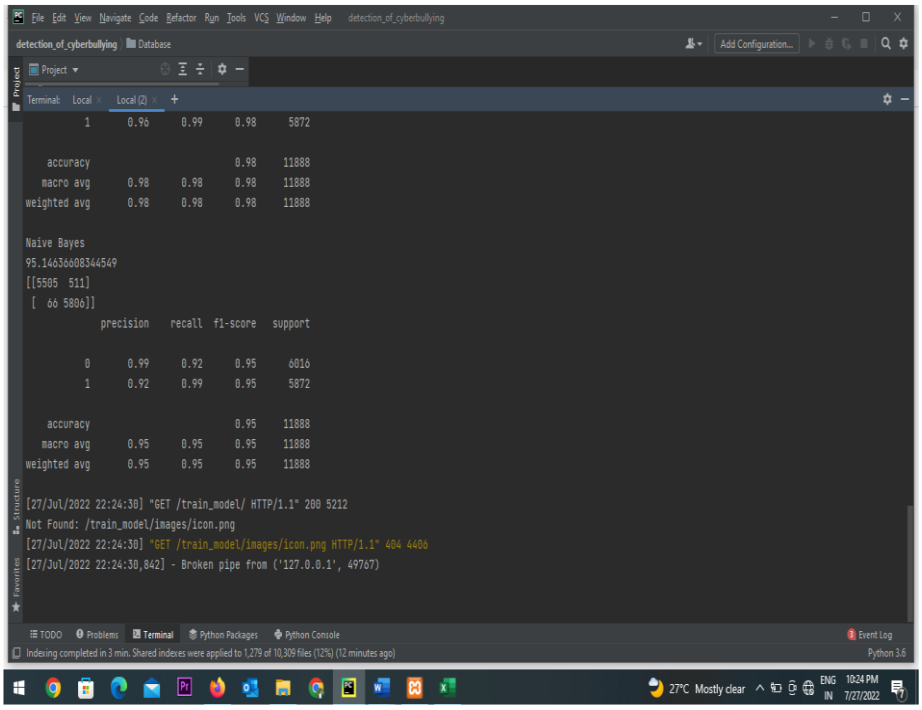

Fig. 5 Training data set

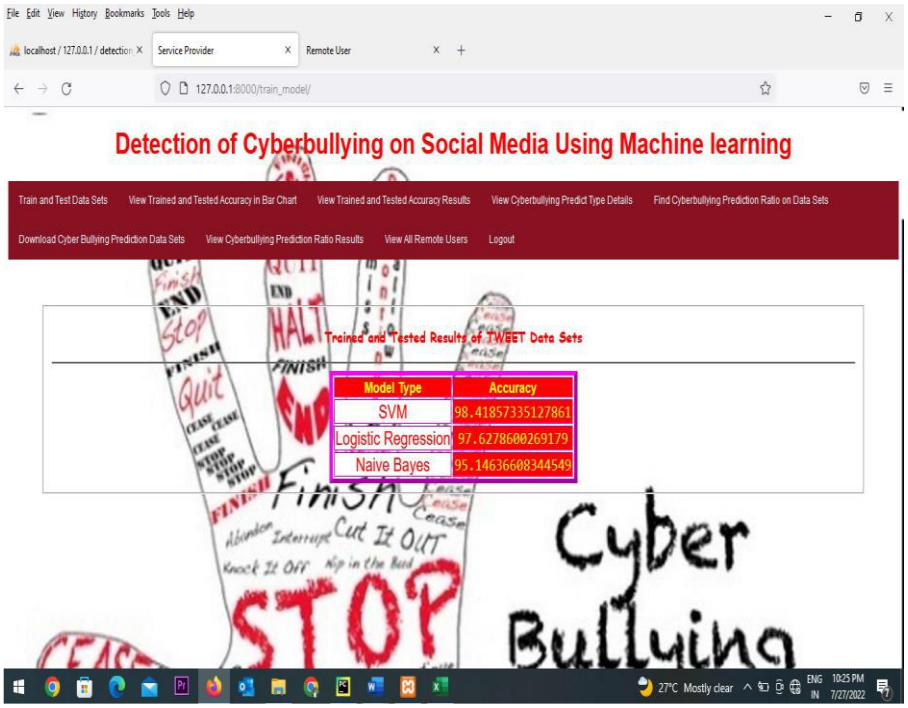

Fig.6 Training and tested accuracy

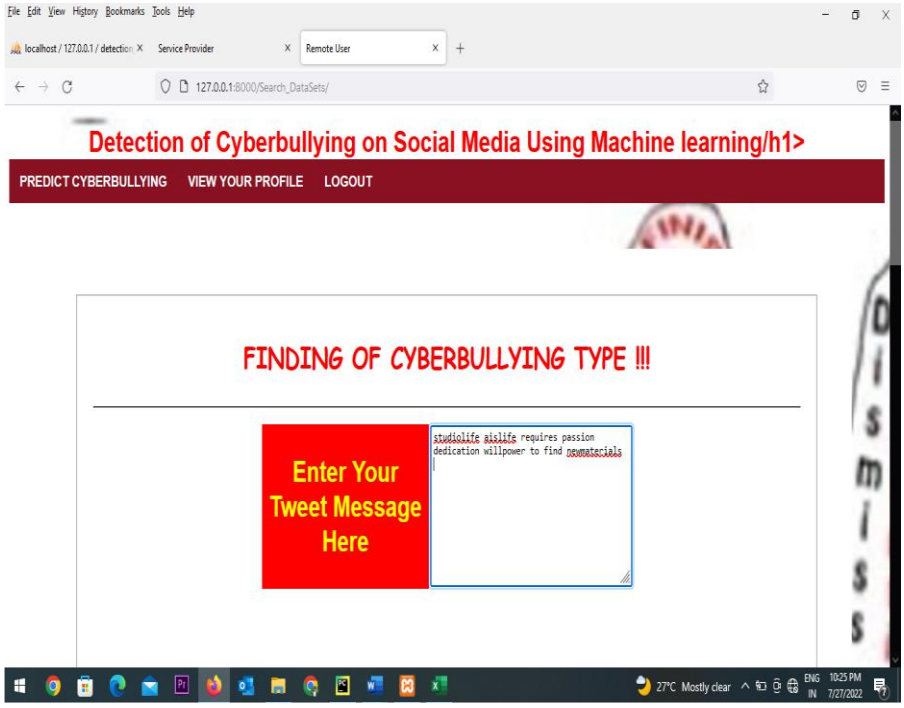

Fig.7 Enter Message For Prediction

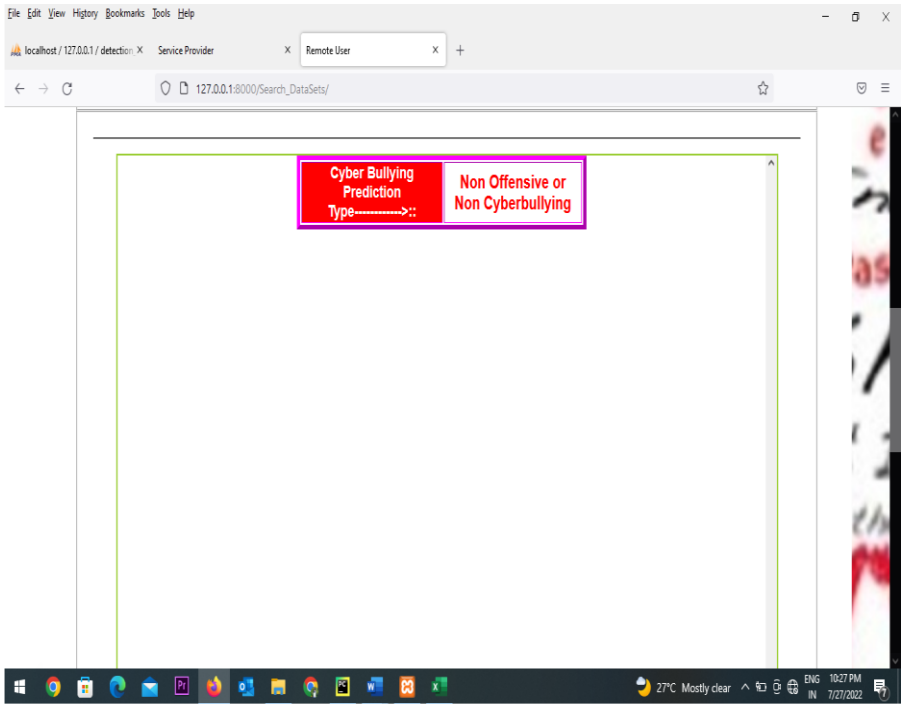

Fig.8 Prediction status

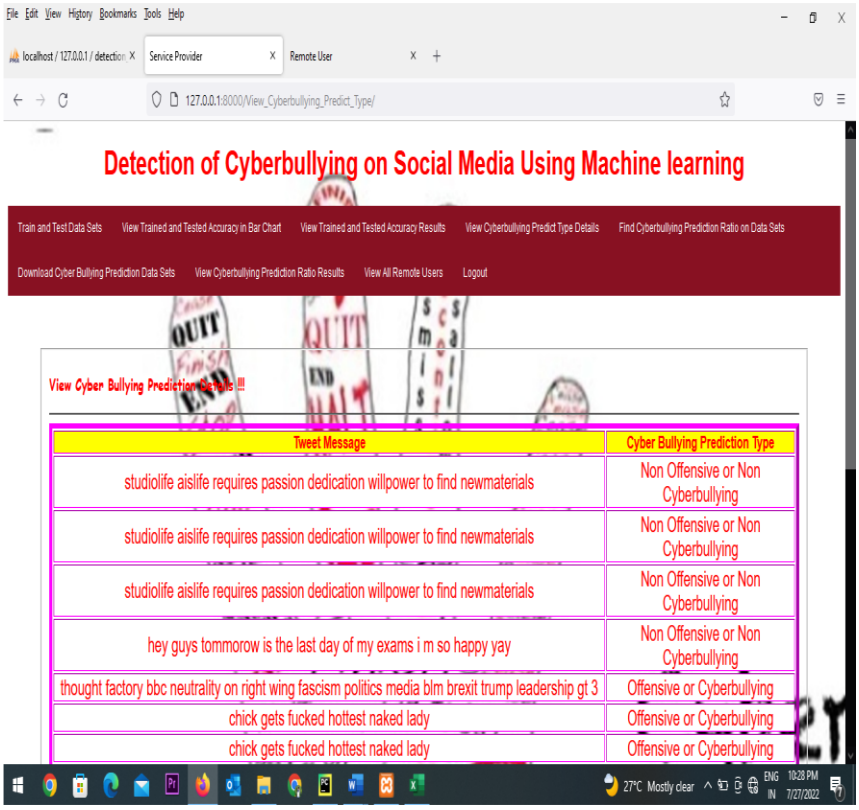

Fig.9 Cyber bullying prediction Details

V. FUTURE SCOPE AND CONCLUSION

Cyber bullying across net is dangerous and ends up in mishappenings like suicides, depression etc and so there's a requirement to manage its unfold. so cyber bullying detection is necessary on social media platforms. With avability of a lot of information Associate in Nursingd higher classified user info for numerous alternative styles of cyber attacks Cyberbullying detection is used on social media we tend tobsites to ban users making an attempt to require half in such activity during this paper we planned an design for detection of cyber bullying to combat things. we tend to mentioned the design for 2 forms of data: Hate speech information on Twitter and private attacks on Wikipedia. For Hate speech linguistic communication process techniques established effective with accuracies. 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. In Future work, Word2Vec models that use context of features proved effective in both datasets giving similar results in comparatively less features when combined with Multi Layered Perceptrons

REFERENCES

[1] I. H. Ting, W. S. Liou, D. Liberona, S. L. Wang, and G. M. T. Bermudez, "Towards the detection of cyberbullying based on social network mining techniques," in Proceedings of 4th International Conference on Behavioral, Economic, and Socio Cultural 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 machine learning for the detection of troll profiles in twitter social network: 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 cyberbullying behavior in Twitter data," 2015, doi: 10.1109/EIT.2015.7293405.

[4] R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," 2016, doi: 10.1145/2833312.2849567.

[5] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," 2019, doi: 10.1109/ICACCS.2019.8728378.