ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

DETECTION OF CYBERBULLYING ON SOCIAL MEDIA USING MACHINELEARNING

Akula Bhavana
Email: Bhavanaakula186@gmail.com
M. Tech, Department of Computer Science and Engineering.
Annamacharya Institute of Technology and Science, Hyderabad, Telangana, India.

Ramesh Babu Varugu
Email: rameshvarugu82@gmail.com
Assistant Professor & HOD, Department of CSE.
Annamacharya Institute of Technology and Science, Hyderabad, Telangana, India.

Abstract - Internet has become the biggest platform to represent our skills. Various websites allow people to use their platform to showcase their skills through videos, articlesand other information in different formats. Mostof the websites provide facility of commenting on any of uploaded information. However, there is possibility that people can use inappropriate language in their comments. Bullying comments are disrespectful, abusive, and unreasonable. The danger of online bullying and harassment affects the free flow ofthoughts by restricting the dissenting opinions of people. Sites struggle to promote discussions effectively. Leading many communities to limitor close down user comments altogether. This project mainly focuses on identification of bullying comments. The required data is taken from machine learning site 'Kaggle' and 'Github'. In this project, three different machine learning algorithms- logistic regression, support vector machine (SVM), Multinomial Naïve Bayes are used to identify the best machine learning algorithm based on our evaluation metrics for comments prediction.The main aim of the project is to predict the bullying comments and its strength like mild, strong, and moderate.

Keywords – Machine Learning, Cyberbullying, Social Media, Twitter...

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 miss 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 findimmoral 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 offCyber 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

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

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.

Cyber bullying is harassment, threatening, embarrassing or targeting someone for the purpose of having fun or even by well-planned means.

Researches on Cyber bullying Incidents show that 11.4% of 720 youngpeoples 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 hoursa 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 tolots 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.

II. EXISTING SYSTEM

Hsien 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.

P. Zhou et al suggested a B-LSTM technique based on concentration.

Banerjee et al used KNN with new embedding's to get a precision of 93%.

Deborah Hall built Bully Blocker, a mobile application that informs parents of cyberbullying activities against their child onFacebook which counted warning signs and vulnerability factors to calculate a value to measure probability of being bullied.

III. PROPOSED SYSTEM

Proposed System Cyberbullying detection is solved in this project as a binary classification problem and classifying them as Non-Bullying or bullying statements. In this project the data is collected from Github and Kaggle and three algorithms are used namely Multinomial Naïve Bayes, Logistic Regression and Support Vector Machine. The purpose of using these three algorithms is to get accurate prediction and strength of offensiveness and strength is also categorized as mild, moderate and strong.

In this project there are two types of end users, Admin and User.

• **Admin** can train and test dataset, view the accuracy of three algorithms, find predictiontype details and ratios, download the prediction datasets done by user.

• **User can login**, enter the text for prediction as cyberbullying or not.

IV. System Design & Development

Requirement analysis in system engineering and software engineering encompasses those

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

tasks that determine conditions to meet for a new product, taking account of possibly conflicting requirements of the various stakeholders such as users. The requirements should be documented, measurable, traceable, test table related to identified business needs or opportunities and defined to a level of detail sufficient for system design.

V. REQUIREMENT ANALYSIS

Requirement analysis in system engineering and software engineering encompasses those tasks that determine conditions to meet for a new product, taking account of possibly conflicting requirements of the various stake holders such as users. The requirements should be documented, measurable, traceable, test table related to identified business needs or opportunities and defined to a level of detail sufficient for system design.

✓ Data Collection

This is the very first step in classification work. Here, the system is trained with labeled data sets for supervised classification.

Data set containing comments collected from Github and Kaggle and contains three columns namely id, tweet and label.

✓ Data Preprocessing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format.

Data pre-processing is a proven method for resolving issues such as, incomplete, inconsistent, and/or lacking in certain behaviors or trends that contain errors.

Here two techniques are used for data preprocessing.

✓ Feature Extraction

Feature extraction is important for Natural Language Processing. Texta data cannot be classified by classifiers therefore they need to be converted to numerical data. Each document (tweet or comment in this case) can be written as a vector and those vectors can be used for classification. The following project studies two Feature extraction methods:

CountVectorizer: CountVectorizer is used to convert a collection of text documents to a vector of term/token counts. It also enables the pre-processing of text data prior to generating the vector representation.

TF-IDF: TF-IDF Stands for **"Term Frequency — Inverse Document Frequency"**. This is a technique to quantify words in a set of documents. Here, a score for each word is computed to signify its importance in the document and corpus. This method is a widely used technique in Information Retrieval and Text Mining.

✓ Life Cycle Model

We use waterfall model for software development process which has various phases in it. The sequential phases in Waterfall model are –

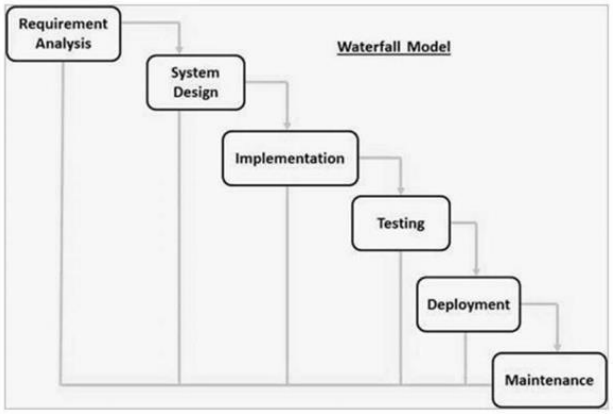
Requirement Gathering and analysis − All possible requirements of the system to be developed are captured in this phase and documented in a requirement specification document.

System Design − the requirement specifications from first phase are studied in this phase and

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

the system design is prepared. This system design helps in specifying hardware and system requirements and helps in defining the overall system architecture.

Implementation − with inputs from the system design, the system is first developed in small programs called units, which are integrated in the next phase. Each unit is developed and tested for its functionality, which is referred to as Unit Testing.



Database Design

Admin Table

```
mysql> desc auth_group;
+-------+-------------+------+-----+---------+----------------+
| Field | Type        | Null | Key | Default | Extra          |
+-------+-------------+------+-----+---------+----------------+
| id    | int(11)     | NO   | PRI | NULL    | auto_increment |
| name  | varchar(80) | NO   | UNI | NULL    |                |
+-------+-------------+------+-----+---------+----------------+
2 rows in set (0.07 sec)
```

Posts Table

```
mysql> desc posts;
+--------------+-------------+------+-----+---------+----------------+
| Field        | Type        | Null | Key | Default | Extra          |
+--------------+-------------+------+-----+---------+----------------+
| tweet_id     | int(11)     | NO   | PRI | NULL    | auto_increment |
| tweet_message| varchar(256)| NO   |     | NULL    |                |
| pred_type    | varchar(100)| YES  |     | NULL    |                |
+--------------+-------------+------+-----+---------+----------------+
3 rows in set (0.02 sec)
```

VI. RESULT

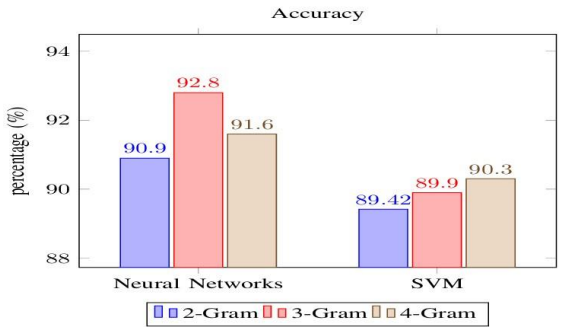
TABLE I. STATISTICS OF THE DATASET

Total number of Conversations	1608
Number of cyberbullying	804
Number of non-Cyberbullying	804
Number of distinct words	5628
Number of token	48843
Maximum Conversation size	773 Characters
Minimum Conversation size	59 Characters

After pre-processing the dataset, we follow the same step presented in Section III to extract the features. We then split the dataset into ratios (0.8,0.2) for train and test. Accuracy, recall and precision, and f-score are taken as a performance measure to evaluate the classifiers. We apply SVM as well as Neural Network (NN) as they are among the best performance classifiers in the literature. We run several experiments on different n-gram language model. In Particular, we take into consideration 2-gram, 3-gram, and 4-gram during the evaluation of the model produced by the classifiers. Table II summarizes the accuracy of both SVM and NN. The SVM classifier achieved the highest percentage using 4-Gram with accuracy 90.3% while the NN achieved highest accuracy using 3-Gram with accuracy 92.8%. It is found that the average accuracy of all n-gram models of NN achieves 91.76%, while the average accuracy of all n-gram models of SVM achieves 89.87%. Fig. 2 depicts the accuracy results of both classifiers.

TABLE II. THE ACCURACY OF SVM AND NN IN DIFFERENT LANGUAGE MODEL

Classifier	2-Gram	3-Gram	4-Gram	Average
SVM	89.42%	89.9%	**90.3%**	89.87%
Neural Network	90.9%	**92.8%**	91.6%	91.76%



ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

In addition to accuracy, Table III and Table IV show the evaluations of both classifiers in terms of precision and recall respectively for each language model. The trade-off between recall and precision is shown in Table V which represents the f score of both classifiers in the different language model. Table V summarizes the f-score of both SVM and NN. The SVM classifier achieved the highest f-measure using 4-Gram with f-score 90.3% while theNN achieved highest f-measure using 2-Gram with f-score 92.2%. It is found that the average f-score of all n-gram models of NN achieves 91.9%, while the average f-score of all n-gram models of SVM achieves 89.8%. Fig. 3 summarizes the f-score ofthe classification of the SVM and Neural Network. The results of average accuracy as well as the average f-score indicate that NN performs betterthan SVM.

TABLE III. RECALL OF SVM AND NN

Classifier	2-Gram	3-Gram	4-Gram	Average
SVM	89.42%	90.3%	**90.8%**	90.1%
Neural Network	91.6%	91.5%	**92%**	91.7%

TABLE IV. PRECISION OF SVM AND NN

Classifier	2-Gram	3-Gram	4-Gram	Average
SVM	89.42%	89.5%	**90%**	89.6%
Neural Network	**93%**	92.5%	91.7%	92.4%

TABLE V. F-SCORE OF SVM AND NN

Classifier	2-Gram	3-Gram	4-Gram	Average
SVM	89.42%	89.8%	**90.3%**	89.8%
Neural Network	**92.2%**	91.9%	91.8%	91.9%

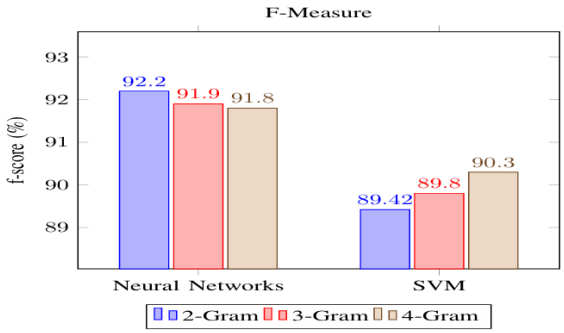


Fig. 3. Comparison between SVM and Neural Network in Terms of F-Measure

In addition to the previous experiments, we evaluate and compare our classifiers on the proposed approach with the work of [23]. In this work, they used logistic regression and SVM for classification and used the same data. Moreover, we have calculated the average accuracy, recall, precision and F score of our two classifiers. The summary of results is shown in Table VI. To compare the work, it is found that our proposed NN model outperforms all other classifiers and is ranked as the best results in terms of average accuracy and F-Score achieving accuracy 91.76% and f-score 91.9%. In Fig. 4 we are comparing between our best classifier with their best classifier in case of accuracy. Finally, here in Fig. 5 we are comparing between our best classifier with their best classifier in case of F-Measure.

TABLE VI. COMPARISON WITH RELATED WORK

	Classifier	Avg. Accuracy	Avg. Recall	Avg. Precision	Avg. F-Score
Vikas S Chavan	Logistic regression	73.76	61.47%	64.4%	62.9%
	SVM	77.65%	58.29%	70.29%	63.7%
Current Results	Neural Network	**91.76%**	**91.7%**	**92.4%**	**91.9%**
	SVM	89.87%	90.1%	89.6%	89.8%

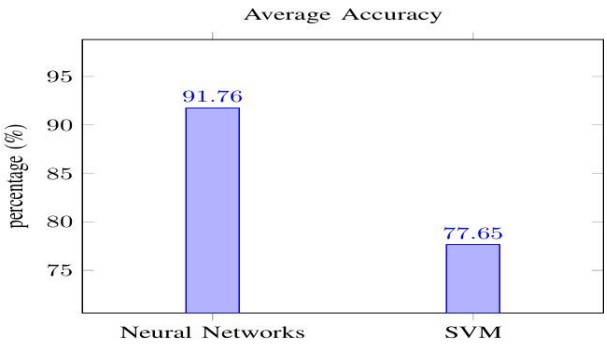


Fig. 4. Comparison between the Best Classifiers in Terms of Accuracy

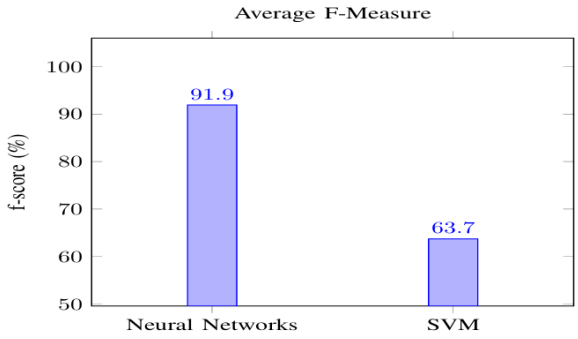


Fig. 5. Comparison between the Best Classifiers in Terms of F-Measure

VII. CONCLUSION

In this paper, we proposed an approach to detect cyber bullying using machine learning techniques. We evaluated our model on two classifiers SVMand Neural Network and we used TFIDF and sentiment analysis algorithms for features ex

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

traction. The classifications were evaluated on different n-gram language models. We achieved 92.8% accuracy using Neural Network with 3- grams and 90.3% accuracy using SVM with 4 grams while using both TFIDF and sentiment analysis together. We found that our Neural Network performed better than the SVMclassifier asit also achieves average f-score 91.9% while theSVM achieves average f-score 89.8%. Furthermore, we compared our work with another related work that used the same dataset, finding that our Neural Network outperformed their classifiers in terms of accuracy and f-score. By achieving this accuracy, our work is definitely going to improve cyber bullying detection to help people to use social media safely. However, detecting cyberbullying pattern is limited by the size of training data. Thus, a larger cyberbullying data is needed to improve the performance. Hence, deep learning techniques will be suitable in the larger data as they are proven to outperform machine learning approaches over largersize data

VIII. FUTURE SCOPE

The present model can be used as add on feature in social media websites to detect bullying statements. Further changes like identifying the user who constantly bullies on social media can be added.

IX. 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.

[6] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," 2011, doi: 10.1109/ICMLA.2011.152.

[7] J. Yadav, D. Kumar, and D. Chauhan, "Cyberbullying Detectionusing Pre-Trained BERT Model," 2020, doi:10.1109/ICESC48915.2020.9155700.

[8] M. Dadvar and K. Eckert, "Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study," arXiv. 2018.

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

ISSN: 2278-4632
Vol-14, Issue-8, August: 2024

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

[11] Z. Waseem and D. Hovy, "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," 2016, doi: 10.18653/v1/n16-2013.

[12] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," 2017.

[13] E. Wulczyn, N. Thain, and L. Dixon, "Ex machina: Personal attacks seen at scale," 2017, doi: 10.1145/3038912.3052591.

[14] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," Artif. Intell. Rev., vol. 53, no. 6, 2020, doi: 10.1007/s10462-019-09794-5.

[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013.