International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Use of Machine Learning to Identify Cyberbullying

***Gandhali Jadhav*[1]*, Mrs. J.C. Pasalkar* [2]**

[1]Student, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, Maharashtra, India
[2]Associate Professor, Department of Information Technology, All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT

It is an age of the Internet and electronic media, and social media platforms are one of the most frequently used communication mediums nowadays. But some people use these sites for malicious purpose and among those negative aspects "Cyberbullying" is prevalent. Cyberbullying is a form of bullying done through electronic means and is used to insult or harm others. Many researchers have proposed solutions and strategies to overcome this menace, but sarcasm is one aspect of it that still needs to be touched. This study aims to highlight previous researchers and to propose an approach to detect cyberbullying along with the element of sarcasm included in it.

Keywords: Cyberbullying, Hate speech, Personal attacks, Machine learning, Twitter, Wikipedia.

1. INTRODUCTION

Social media is a platform which enables users to communicate and interact with their friends online and allows them to share their photos, videos and daily updates. Nowadays, almost every person is found on social media. According to statistics nearly 2 billion users used social sites in 2015 and the figure has now increased to 3.196 billion Social media has its perks, but it has negative aspects as well. "Cyberbullying" is one of those aspects that need to be handled. Cyberbullying is a form of bullying or harassment that is done through electronic means and is common among young people and teenagers. It includes posting malicious and harmful comments or posts online and sharing personal information about someone to humiliate them. Cyberbullying is one critical issue that is prevailing throughout the world. The person being bullied falls into depression, causes self-harm and in worst cases commits suicide. Thus, Cyberbullying is a severe problem that needs to be taken care of considering the severity of damage it causes to an individual's mental well-being.

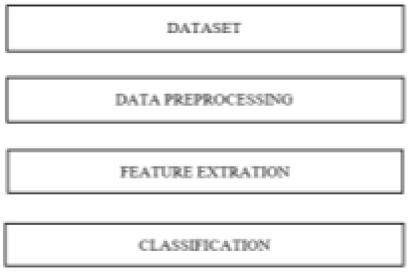
2. LITERATURE REVIEW

Many researchers have proposed mechanisms to detect cyberbullying. Reviewed machine learning algorithms for cyberbullying detection in Arabic social media networks. Conducted a survey to explore work done on detection and prevention of cyberbullying. Presented a task known as SemEval-2019 which was identifying and categorizing offensive language in social media. Conducted a study for detecting cyberbullying and cyber aggression in social media. Audited an existing algorithm using Twitter dataset. The algorithm aims to detect who is the recipient or who is the person being bullied.

J. Yadav, proposes a new approach to cyberbullying detection in social media platforms by using the BERT model with a single linear neural network layer on top as a classifier. The model is trained and evaluated on the Form spring forum and Wikipedia dataset. The proposed model gave a performance accuracy of 98% for the Form spring dataset and of 96% for the Wikipedia dataset which is relatively high from the previously used models. The proposed model gave better results for the Wikipedia dataset due to its large size g without the need for oversampling whereas the Form spring dataset needed oversampling.

3. PROPOSED METHODOLOGY

Fig describes the methodology used for solving the problem which is applied on both the datasets.



- Natural Language processing: The real-world posts or text contain various unnecessary characters or text. For example, numbers or punctuation are irrelevant to bullying detection. Before applying the machine learning algorithms to the comments, we need to clean and prepared them for the detection phase. In this phase, various processing task including removal of all irrelevant characters like stop-words, punctuation and numbers, tokenization, stemming etc. After the preprocessing, we prepare the two important features of the texts as follows: 1) Bag-of-Word: The machine learning algorithms cannot work directly with the raw text. 2) TF-IDF: This is another feature that we consider for our model. TF- IDF (Term Frequency-Inverse Document Frequency) is a statistical measure that can evaluate how relevant a word is to a document in a collection of documents
- Machine Learning: This module involves in applying various machine learning approaches like Decision Tree (DT), Random Forest, Support Vector Machine, Naive Bayes to detect the bullying message and text. The classifier with the highest accuracy is discovered for a particular public cyberbullying dataset. Next section, some common machine learning algorithms are discussed to detect cyberbullying from social media texts.

Machine Learning Algorithms

A. Support Vector Machine (SVM)

This theorem is basically used to plot a hyperplane that creates a boundary between data points in number of features (N)-dimensional space. To optimize the margin value hinge function is one of best loss function for this. Linear SVM is used in the following case which is optimum for linearly separable data In case of misclassification, i.e., our model makes a mistake in our data point's class prediction, we add the reduction with the gradient update regularization.

B. Logistic Regression

It is a classification model and not a regression model. The probabilistic function used to model the output of problem is sigmoid function in (7) T(x) is hypothesis function for our classifier, L is weights derived by classifier, C is bias derived by classifier and T is feature vector(input). Since sigmoid lies between 1 and 0 it is ideal for classification.

C. Random Forest

A random forest consists of many individual decision trees which individually predict a class forgiven query points and the class with maximum votes is the result. Decision Tree is a building block for random forest which provides a prediction by decision rules learned from feature vectors. An ensemble of these uncorrelated trees provides a more accurate decision for classification or regression.

D. Multi Layered Perceptron

Multi Layered Perceptron's are the Artificial Neural Networks containing at least 3 layers: one input, one output and at least one hidden layer. Each node has activation value calculated using an activation function in a process called forward propagation and back propagation is used to train the weight used in the neural networks. Sigmoid function is similar to the tanh function which is hyperbolic in nature between -1 and 1. Rely is defined as f(x)=max(0, x).

1. EXPERIMENTS AND RESULTS

Google colas was used for the experiments. For each classifier the following parameters were evaluated on the test sets

- Accuracy(A): is defined as no of correct predictions divided by total number of predictions. A □ True positives / Size of dataset
- Precision(P): Out of all the positive predictions by the classifier how many are positive. P = True positives / True positives + False positives
- Recall(R): Out of all the positive inputs how many were predicted positive. R = True positives / True positives + False negatives
- F-measure(F): Calculates HM (harmonic mean) of precision and helps in comparison of both precision and recall. F = 2*P*R / P+R

As machine learning algorithms use numeric data for training, so the text was first converted into the numerical form using a label encoder. After that, the dataset was divided into 80% training set, and 20% test set and then classification algorithms were applied. For classification machine learning algorithms; SVM, naïve Bayes, Random Forest and then an ensemble approach was used. The results obtained after applying algorithms on datasets are mentioned below:

Table 1-Dataset 2-Twitter

Algorithm	Accuracy
Support Vector Machine (SVM)	68%
Logistic Regression	69%
Random Forest	72%
Multi Layered Perceptron	81%

Table 2-Dataset 2-Wikipedia

Algorithm	Accuracy
Support Vector Machine (SVM)	78%
Logistic Regression	87%
Random Forest	80%
Multi Layered Perceptron	79%

Table 3--Average Accuracy

Algorithm	Accuracy
Support Vector Machine (SVM)	79.7 %
Logistic Regression	75.7%
Random Forest	78.9%
Multi Layered Perceptron	79%

The results clearly show that the used approach yielded considerably good results and by observing the average accuracy we can say that SVM and Multi Layered Perceptron classifier performed better than others.

5. 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 is vital on social media platforms. We discussed the architecture for two types of data: Hate speech Data on Twitter and Personal attacks on Wikipedia

The previous work done in this regard was also highlighted. Cyberbullying is a vast term and has different aspects. Among those aspects, sarcasm is essential. Sarcasm is a way of insulting someone and has adverse effects on the victim. As per our observation, this aspect of bullying was not considered in the previous researches. Hence this study aimed to include that aspect as well. In this study, we detected cyberbullying using machine learning algorithms. Then results were presented in a tabular format. The results depicted that SVM and Multi Layered Perceptron performed better than remaining classifiers with 79% average accuracy. The second one is Random Forest with 78% accuracy, then Logistic Regression with 76.7% accuracy. However, in this study, only textual features were considered. For future, network and contextual features can be considered.

6. RELATED WORK

Base Paper topic name	Author name	Journal Name & Publication Year	link
Detection of Cyberbullying on social media Using Machine learning	Varun Jain, Vishant Kumar, Vivek Pal	Proceedings of the Fifth International Conference on Computing Methodologies.	https://ieeexplore.ieee.org/ document/9418254

Cyberbullying Detection on Social Networks Using Machine Learning Approaches	Md Manowarul Islam, Selina Sharmin	Conference Paper April 2021	https://www.researchgate.net/publication/ 351131976_Cyberbullying_Detection_on_ Social_Networks_Using_Machine_ Learning_Approaches
Cyber Bullying Detection on social media using Machine Learning.	Aditya Desai, Shashank Kalkaska, Omkar Kumbhar, and Rashmi Dhumal	ITM Web of Conferences 40, 03038 (2021) ICACC-2021	https://thesai.org/Publications/ViewPaper ?Volume=10&Issue=5&Code= IJACSA&SerialNo=87
NLP and Machine Learning Techniques for Detecting Insulting Comments on Social Networking Platforms	Hitesh Kumar Sharma, K Kshitiz, Shailendra	2018 International Conference on Advances in Computing and Communication Engineering	https://www.itm-conferences.org /articles/itmconf

ACKNOWLEDGEMENTS

The success of any project or a research paper depends largely on the encouragement and guidelines of many others. This Research Seminar would not have been possible without their support. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

First and foremost, we wish to record our sincere gratitude to the Management of this college and to our Respected Principal Dr.P.B. Mane, for his constant support and encouragement in the preparation of this report and for the availability of library facilities needed to prepare this report. Our sincere thanks to Dr. M. A. Thalor, Head, Department of Information Technology, AISSMS IOIT, Pune who was abundantly helpful and offered invaluable assistance, support, and guidance.

We express our sincere gratitude to our Guide, Smt. J.C. Pasalkar for guiding us in investigations of this Seminar and in carrying out experimental work. Our numerous discussions were extremely helpful. We are highly indebted to him for his guidance and constant supervision as well as for providing necessary information regarding the Seminar & also for his support in completing the Seminar. We hold him in esteem for guidance, encouragement and inspiration received from him. I thank all for continuous strength we were bestowed for completion of this Seminar.

REFRENCES

[1] Chaffey Dave, "Global social media research summary 2019 | Smart Insights," 2019. [Online]. Available: https://www.smartinsights.com/social-media-marketing/socialmedia-strategy/new-global-social-media-research/. [Accessed: 28- Sep-2019].

[2] "What Is Cyberbullying | StopBullying.gov." [Online]. Available: https://www.stopbullying.gov/cyberbullying/what-is-it/index.html. [Accessed: 30-Sep-2019].

[3] C. Van Hee et al., "Automatic detection of cyberbullying in social media text," PLoS One, vol. 13, no. 10, Oct. 2018.

[4] M. Ptaszyński, G. Leliwa, M. Piech, and A. Smywiński-Pohl, "Cyberbullying Detection -- Technical Report 2/2018, Department of Computer Science AGH, University of Science and Technology," Aug. 2018.

[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 Detection using Pre-Trained BERT Model," 2020, doi: 10.1109/ICESC48915.2020.9155700