IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11463

Detection of Twitter-Cyberbullying using Python

Namrata Khade[1], Snehal Sarkate[2], Palak Kombade[3], Vaishnavi Alone[4], Vaishnavi Parate[5]

Assistant Professor Dept. of Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India[1]

UG Students, Computer Science and Engineering, Priyadarshini College of Engineering, Nagpur, Maharashtra, India[2,3,4,5]

Abstract: Our paper provides Detection of Cyberbullying using Machine Learning. In this project, we aim to build a system that tackles Cyberbullying by identifying the mean-spirited comments and also categorizing the comments as bullied one or not. The goal of this project is to show the implementation of software that will detect bullied tweets. As the social networking sites are increasing, cyberbullying is increasing day by day in everyone's daily life who is using internet access. To identify such bullying tweets in the twitter handle we are going to make a software which will help to detect such mean type of comments with the help of Machine Learning model. As developing ML model, it will automatically detect the mean-spirited comments from the comment section. For this a Machine learning model is proposed to identify or detect and prevent the bullying on social media. As Machine Learning is used, we used two classifiers such as Naïve Bayes and SVM (Support Vector Machine) for training and testing the social median contents. Twitter API is used to fetch tweets and tweets are passed to the model to detect whether the tweets are bullying or not.

Keywords: Cyberbullying detection · Machine Learning · Twitter· Tweets · Online harassment.

I. INTRODUCTION

These days, technology has become a very dominant part of our lives and most people can't live without it. The Internet comes up with a platform to share their ideas. Many people are expending a large amount of time on social media. Communication with people is no exception, as technology has swapped the way people interact with a broader manner and has given a new dimension to communication. Many people are illegally or unlawfully using these communities. Many youngsters are getting bullied these days. Bullies use various services like Twitter, Facebook, Email to bully people. Studies show that about 37% of children in India are involved in cyberbullying and nearly 14% of bullying occurs regularly. Cyberbullying affects the victim both ways emotionally and psychologically. Social media also allows bullies to harness the anonymity which satisfies their unkind deeds. Things also get more serious when bullying occurs more repeatedly over time. So, preventing it from happening will help the victim. Cyberbullying and its impact on social media: Cyberbullying is an act of threatening, harassing or bullying someone through modern ways of communicating with each other and with anybody/everybody in the world via social media apps/sites.

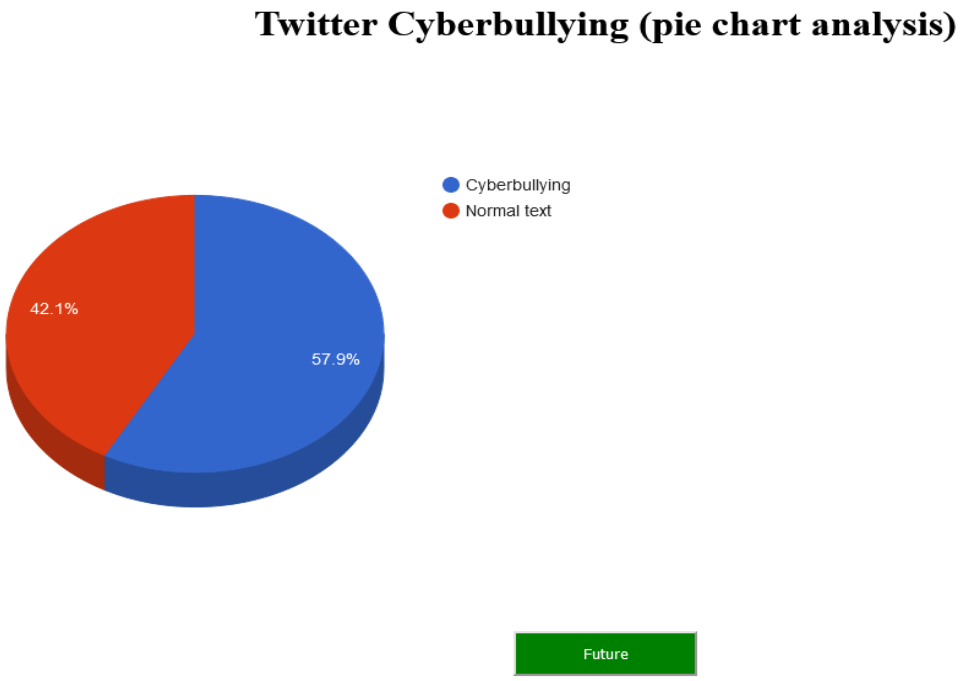
Cyberbullying is not just limited to creating a fake identity and publishing/posting some embarrassing photo or video, unpleasant rumors about someone but also giving them threats. The impacts of cyberbullying on social media are horrifying, sometimes leading to the death of some unfortunate victims. The behavior of the victims also changes due to this, which affects their Emotions, self-confidence and a sense of fear is also seen in such people. Thus, a complete solution is required for this problem. Cyberbullying needs to stop. The problem can be tackled by detecting and preventing it by using a machine learning approach, this needs to be done using a different perspective. The main purpose of our paper is to develop an ML model so it can detect and prevent social media bullying, so nobody will have to suffer from it. The proposed technique is implemented on the social media bullying dataset which was collected from various sources like Kaggle, GitHub, etc. The performance of both NB and SVM is compared to TFIDF. Twitter API is used to fetch a particular location's tweets to detect whether they are Bullying or not. Furthermore, the probability of each tweet is calculated to predict the result and the result of each tweet is stored into the database with bully's username

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11463



This is the pie chart of the bullying analysis of the tweets. Twitter Cyberbullying check the level of acceptance of the system by the user. This includes the process of training the user to the system efficiently.

II. PROPOSED SYSTEM:

- In this paper, an answer is proposed to recognize twitter cyber-bullying. The primary contrast with past examination is that we not just fostered an AI model to detect cyber-bullying content yet additionally carried out it on particular locations ongoing tweets.
- The Data Preprocessing, Data Extraction will be performed on the brought Tweets.
- Preprocessed tweets will be passed to Naïve Bayes model to calculate the probabilities of brought tweets to check whether a got tweet is harassing or not.

ADVANTAGES OF PROPOSED SYSTEM:

- The proposed strategy is implemented on the virtual entertainment harassing dataset which was gathered from various sources like Kaggle, GitHub, and so on.
- Moreover, the likelihood of each tweet is determined to foresee the outcome and the consequence of each tweet is stored into the data set with menaces username.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11463

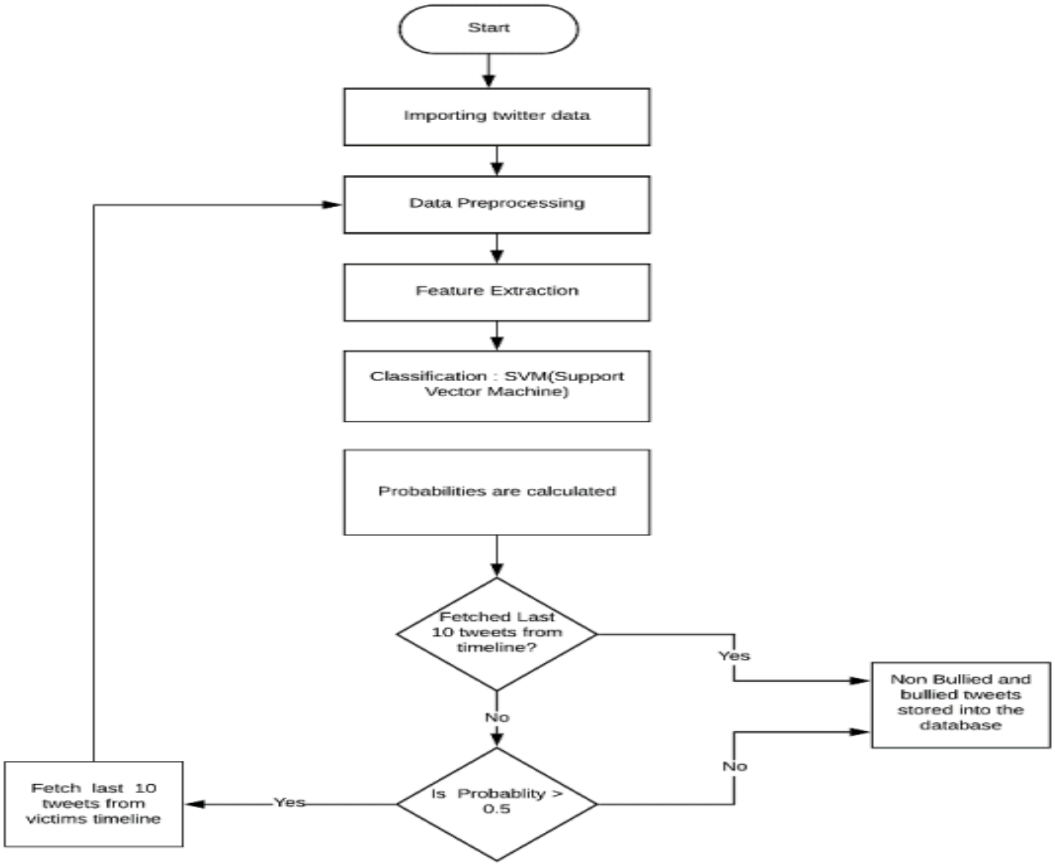


Fig 4: Flowchart of software setup

III. RESEARCH METHOD

1. Developing the Model: The entire model is divided into 3 major steps: Preprocessing, the algorithm, and feature extraction.

A. Preprocessing: The Natural Language Toolkit (NLTK) is used for the preprocessing of data. NLTK is used for tokenization of text patterns, to remove stop words from the text, etc.

• Tokenization: In tokenization, the input text is split as the separated words and words are append to the list. Then 4 different tokenizers are used to tokenize the sentences into the words:

o Whitespace Tokenizer
o WordPunct Tokenizer
o TreebankWord Tokenizer
o PunctWord Tokenizer

• Lowering Text: It lowers all the letters of the words from the tokenization list. Example: Before lowering “Hey There” after lowering “hey there”.
• Removing Stop words: This is the most important part of the preprocessing. Stop words are useless words in the data. In this stage stop words like \t, https, \u, are removed from the text.

B. Feature Extraction: In this step, the proposed model has transformed the data in a suitable form which is forwarded to the machine learning algorithms. The TFDIF vectorizer is used to extract the features of the given data. Features of the data are extracted and put them in a list of features. Also, the probability (i.e., the text is Bullying or Non-Bullying) of each text is extracted and stored in the list of features.

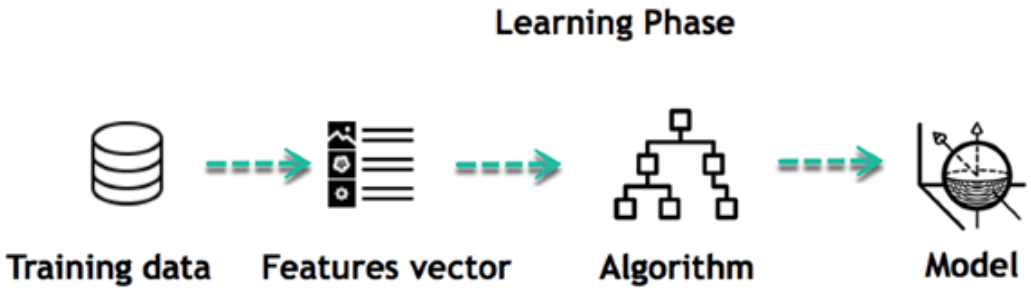
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.11463

C. Algorithm Selection: To detect social media bullying automatically, supervised Binary classification machine learning algorithms like SVM with linear kernel and Naive Bayes is used. The reason behind this is both SVM and Naive Bayes calculate the probabilities for each class (i.e., probabilities of Bullying and Non-Bullying tweets). Both SVM and Naïve Bayes algorithms are used for the classification of the two-cluster. Both the machine learning models were projected on the same dataset. But SVM perform better than Naive Bayes on similar work on the same dataset as it gives highest accuracy than Naïve Bayes.



IV. FINDINGS AND ANALYSIS

INPUT DESIGN AND OUTPUT DESIGN: -

INPUT DESIGN

The info configuration is the connection between the data framework and the client. It involves the creating determination and strategies for information arrangement and those means are important to place exchange information in to a usable structure for handling can be accomplished by investigating the PC to peruse information from a composed or printed archive or it can happen by having individuals entering the information straightforwardly into the framework.
The plan of info centers around controlling how much information required, controlling the blunders, abstaining from delay, trying not to additional means and keep the interaction basic. The information is planned in such a manner so it furnishes security and usability with holding the protection. Input Design thought about the accompanying things:

- What information ought to be given as info?
- How the information ought to be organized or coded?
- The discourse to direct the working staff in giving information.
- Techniques for getting ready info approvals and steps to follow when mistake happen.

V. RESULT

In this section, the SVM and Naive Bayes on the dataset collected from the different sources like Kaggle, GitHub, etc. is collected. After performing preprocessing and feature extraction on the dataset, for training and testing, and divided the dataset into ratios 0.45 and 0.55 respectively. Both SVM and Naive Bayes are evaluated to determine the accuracy, recall, f-score, and precision. Incidentally SVM outperformed Naive Bayes in every aspect. Table I exhibits the accuracies of both the Naive Bayes and SVM. The Support Vector Machine achieved the highest accuracy i.e., 71.25%, while Naive Bayes achieved 52.70% accuracy.

Classifier's	Accuracy (in %)
Naive bayes	52.70
Support vector machine	71.25

TABLE: The Accuracy of Support Vector Machine and Naive Bayes

VI. FUTURE SCOPE

- The field of OSN security and protection is a new and arising one, offering numerous headings to seek after.
- Security analysts can ceaselessly give improved answers for online dangers; they can likewise find new security dangers to address. We trust that to work on the current arrangements, the subsequent stage is to make cooperative energy

DOI: 10.17148/IJARCCE.2022.11463

among the different security arrangements. This will make more powerful and viable security answers for distinguishing counterfeit profiles, spammers, phishing assaults, socware, and different dangers.

- A further examination bearing for further developing OSN clients' protection is to dissect and assess the different existing security arrangements presented by OSN administrators, pinpointing their deficiencies and recommending strategies for further developing security arrangements. Research that creates procedures to more readily instruct clients about these arrangements would likewise be of worth, as would methods to make clients more mindful of existing OSN dangers.

VII. CONCLUSION

A solution is proposed for detecting and preventing Twitter cyberbullying using Machine Learning algorithms. Our model is evaluated on both Support Vector Machine and Naive Bayes, also for feature extraction, used the TFIDF vectorizer. As the results show us that the accuracy for detecting cyberbullying content has also been great for Support Vector Machine which is better than Naive Bayes. Our model will help people from the attacks of social media bullies and as to protect from bullying it is used as an effective method to be safe from bullying attacks.

VIII. REFERENCES

1. Fire M, Goldschmidt R, Elovici Y. Online Social Networks: Threats and Solutions. IEEE Commun Surv Tutor. 2014;16: 2019–2036. https://doi.org/10.1109/COMST.2014.2321628
2. Penni J. The future of online social networks (OSN): A measurement analysis using social media tools and application. Telemat Inform. 2017; 34: 498–517. https://doi.org/10.1016/j.tele.2016.10.009
3. Lauw H, Shafer JC, Agrawal R, Ntoulas A. Homophily in the Digital World: A LiveJournal Case Study. IEEE Internet Comput. 2010; 14: 15–23. https://doi.org/10.1109/MIC.2010.25
4. Rezvan M, Shekarpour S, Balasuriya L, Thirunarayan K, Shalin VL, Sheth A. A Quality Type-aware Annotated Corpus and Lexicon for Harassment Research. Proceedings of the 10th ACM Conference on Web Science. New York, NY, USA: ACM; 2018. pp. 33–36.
5. Hee CV, Jacobs G, Emmery C, Desmet B, Lefever E, Verhoeven B, et al. Automatic detection of cyber- bullying in social media text. PLOS ONE. 2018; 13: e0203794. https://doi.org/10.1371/journal.pone.0203794 PMID: 30296299
6. Hosseinmardi H, Shaosong Li, Zhili Yang, Qin Lv, Rafiq RI, Han R, et al. A Comparison of Common Users across Instagram and Ask.fm to Better Understand Cyberbullying. 2014 IEEE Fourth Interna-tional Conference on Big Data and Cloud Computing. 2014. pp. 355–362.
7. Citron DK. Addressing Cyber Harassment: An Overview of Hate Crimes in Cyberspace. the Internet.2015; 6: 12.
8. Wall D. What are Cybercrimes? Crim Justice Matters. 2004; 58: 20–21. https://doi.org/10.1080/09627250408553239
9. Abu-Nimeh S, Chen T, Alzubi O. Malicious and Spam Posts in Online Social Networks. Computer. 2011; 44: 23–28. https://doi.org/10.1109/MC.2011.222
10. Doerr B, Fouz M, Friedrich T. Why Rumors Spread So Quickly in Social Networks. Commun ACM. 2012; 55: 70–75. https://doi.org/10.1145/2184319.2184338

IX. BIOGRAPHY

Mrs. Namrata S. Khade is working as a Asst. Professor at Priyadarshini College of Engineering. She is having 12 years of experience in the field of teaching to engineering students. She completed her Engineering in 2007 and Master in Engineering in 2013.She is a member of IEEE, ISTE and CSI. She is having more than 30 research published in International Journals and Conferences. Her interests include distributed parallel computation, System Programming, Computer Graphics and Wireless Sensor Network.

