Grenze International Journal of Engineering and Technology, June Issue

Cyber Bullying Predictive Analysis on Twitter Data with Multi Model Supervised Technique

S.Giridharan[1], M. Dhileep[2], R. Mehdir Muhammed[3] and P.V. Ranjith[4]
[1-4]Department of Cyber Security, SRM Valliammai Engineering College
Email: mailtoogiri@gmail.com, {jeevam701269@gmail.com, mehdirmuhammed2002@gmail.com, ranpvs@gmail.com}

***Abstract*— A multi-model supervised technique is employed to predict incidences of cyberbullying on Twitter in a thorough manner. The suggested approach makes use of sentiment analysis, natural language processing, and machine learning techniques to detect and prevent cyberbullying activities. The technology helps to create a safer online environment by precisely classifying potentially hazardous information by analyzing Twitter data. Combining several models improves the robustness and accuracy of predictions. The research uses cutting-edge analytical methods to identify harmful interactions and promote pleasant online interactions, which helps to avoid cyberbullying proactively.**

***Index Terms*— Cyberbullying, machine learning, twitter data.**

I. INTRODUCTION

In an era dominated by social media, the prevalence of cyberbullying poses significant challenges to the well-being of online communities. The system's primary objective is to accurately classify potentially harmful content, thereby contributing to the creation of a safer online environment. The synergy of multiple models not only enhances predictive accuracy but also fortifies the system's robustness. This research represents a significant step towards proactive cyberbullying prevention strategies, employing advanced analytical tools to identify harmful interactions and foster positive online engagement. By embracing a multifaceted approach, our study underscores the importance of early intervention and the promotion of respectful digital discourse, emphasizing the role of cutting-edge.

II. RELATED WORKS

[1] Akib Mohi Ud Din Khanday, Syed Tanzeel Rabani, Qamar Rayees Khan, Showkat Hassan Malik carried out the process of cyberbullying detection with traditional machine learning technique where the implemented machine learning has only two kinds of classifier mechanisms leading to acceptable range of process outcomes. [2] Arvind Kumar Gautam & Abhishek Bansal made an extensive approach to detect cyberbullying happening through mail-based communication. Carrying out cybercrimes such as e-mail bombing, spamming, phishing, spoofing in order to defame or bully a individual have also been taken into account by the authors. [3] K. S. Alam, S. Bhowmik and P. R. K. Prosun gathered information on sensex about the victims falling prey to cyberbullying which even leads to diminishing their fame among the society hence ending their life due to humiliation happened online. The included offensive language, cyber bullying content and other kinds of improper contents affecting the dignity of an individual in the cyberspace. [4] Giri. S, Banerjee. S included multiple languages and emojis which

Grenze ID: 01.GIJET.10.2.196

are frequently used for bullying an individual along with the rarely used ones. This helped the authors to detect and find more accurate and useful output which can be used for future references in improving the cyber bullying detection models. [5] Sneha Chinivar, Roopa M.S., Arunalatha J.S., Venugopal K.R. found the growing cyber bullying actions happening in the social space like Facebook Twitter Instagram including hate speech carried out by targeting an individual in the cyberspace. The authors made multiple researches for finding the bullying comments and hate speech that can be found under the comments section of a particular post posted by an individual.

III. ARCHITECTURE FRAMEWORK

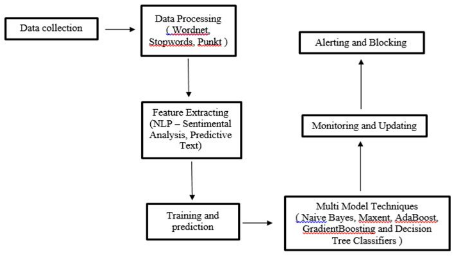


Figure 1: Cyberbullying prediction framework

In figure 1, initially, we will be collecting unprocessed data from twitter application as common without classifying the categories (bullying and non-bullying). The collected data consists of words and sentences used in real practical world. The data will be considered raw. Then the unprocessed data undergoes processing for better and accurate error correction. Here we have used python as our primary programming language, we have imported few modules like stopwords, wordnet, punkt. A set of commonly used words in a language are referred as stopwords. Here we use stopwords to eliminate widely used words that hold very few information. Wordnet in NLP is used to gather information about a particular word, sentence, relation, etc. Punkt splits and divides a piece of text into a list of sentences by using an unsupervised algorithm to build a model for abbreviation words, collocations, and words that start sentences. Once the data completes it's processing state, it undergoes feature extraction where the data is classified as bullying and non- bullying based on the process happened previously. A web application will be created and connected with the dataset for testing it with random text. Later, the module will be monitored for the presence of any kind of error or issue arises, the module will be updated and let down for functioning again.

IV. APPROACH

A. Multi Model Supervised Technique

We are implementing multiple supervised techniques for gaining accuracy and exact guess about the input content based on its nature. Naïve Bayes Classifier is the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. MaxEnt classifier is employed here for better performance in sentimental analysis field as the dataset includes emojis and statements having sentimental data. Adaptive Boosting, or AdaBoost, is a machine learning ELT used for regression and classification issues. Gradient boosting is a potent boosting procedure that turns multiple weak learners into powerful learners.

B. Dataset collection

The data is collected from real world comments that are posted twitter application under comments section of a post. The reason behind using real-time dataset is that it falls under the category of usual happening in cyberspace where every individual praises and bully's other intended individual. The dataset comprises of two sections (bullying and non- bullying) making the process of understanding more easier. It includes multiple languages that are widely used by every individual of different geographical region for wider usage.

V. THE PROPOSED SCHEME

A. Data Collection

The data is collected directly from twitter application. The data collected will have enormous number of comments, emojis, symbols. The data is collected manually from real- time user accounts. While collecting the data, setting up a parameter on how to carry out the research of collecting the comments was the challenging part. The collected data is unprocessed, and it did contain many unfinished comments along with some unwanted punctuations, symbols, random signs, etc.

B. Data Processing

The data undergoes the process of processing where the collected data will be corrected based on the parameters which helps to avoid false positives which will arise while undergoing feature extraction process. During data processing, here we induce stopwords to eliminate widely used words that hold very few information. Wordnet is used to gather information about a particular word, sentence, relation, etc. Wordnet gathers information about a particular word and sentence, how they are related to each other, knowing their synonym, modifying the synonym according to it's placement. Punkt splits and divides a piece of text into a list of sentences by using an unsupervised algorithm to build a model for abbreviation words, collocations, and words that start sentences.

C. Feature extracting

In feature extracting, with the help of sentimental analysis (emojis in the data), predictive text (processed) the feature can be classified based on its nature. The extracted dataset will be updated in the form of spreadsheet. The spreadsheet contains of both bullying and non- bullying sections with respective words and sentences. Even predictive text can be identified with the help of NLPTK. Altogether, feature extraction process completes with classifying the contents based on their nature as bullying and non-bullying.

D. Training and prediction

Naïve Bayes Classifier is the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. MaxEnt classifier is employed here for better performance in sentimental analysis field as the dataset includes emojis and statements having sentimental data. Adaptive Boosting, or AdaBoost, is a machine learning ELT used for regression and classification issues. Gradient boosting is a potent boosting procedure that turns multiple weak learners into powerful learners.

E. Monitoring and updating

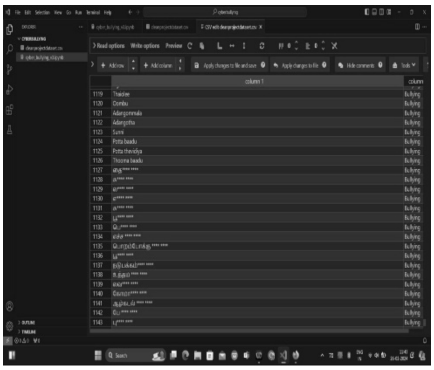
Here, a web application will be created with the help of HTML, CSS and along with the backend operational coding for the functioning of the model. The HTML framework includes the option to enter the text which can be classified as bullying or non-bullying. The CSS provides a beautiful graphical user interface for the user who is checking about the characteristics of the contents.

F. Alerting and blocking

This being the final stage of the model, the cyberbullying content will be monitored at the backend. This works only if the backend source code is connected well with the web application and only if the text entered in the text field is present in the spreadsheet provided as the dataset.

VI. EXPERIMENTAL RESULTS

A. Dataset collection



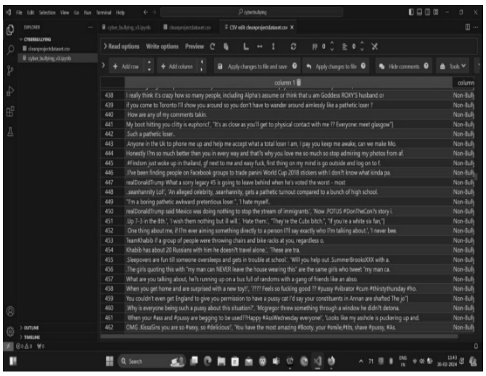


Figure 2 & 3: Dataset classified as bullying and non-bullying in spreadsheet.

In figure 2 & 3, the dataset has been collected from twitter application which are used in real-time by every individual for bullying and non-bullying has been collected from twitter application. The dataset has been collected manually and updated every time at regular interval of duration for having better accuracy.

B. Prediction

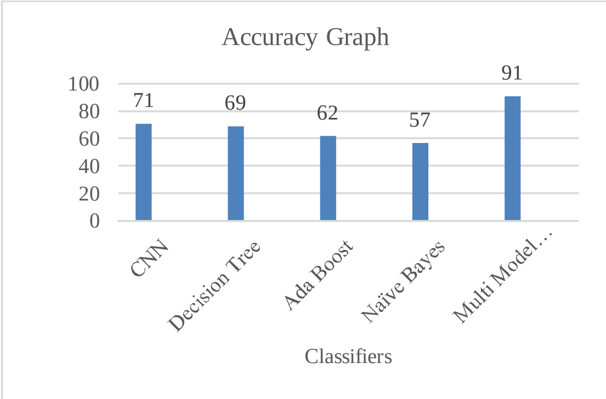


Figure 4: Plot of predicted texts used in dataset.

In figure 4, the prediction is carried out with the created dataset which displays the prediction results through plotted bar charts. It helps in better understanding of the public audience. The bar chart also displays the different multiple classifiers used here for the prediction purpose. The increase in dataset with more languages and comments, will help in improving the accuracy and predictivity of the classified texts. In future, the model can be updated with more possible ML techniques with extensive datasets.

VII. Discussion

Researchers often discuss cyberbullying as a reflection of broader social issues, including the impact of online culture and the anonymity provided by the internet. Understanding the social dynamics that contribute to cyberbullying is crucial for developing effective prevention and intervention strategies. Discussions often center around the role of technology in both facilitating and combatting cyberbullying. Social media platforms and other online spaces play a significant role in the spread of harmful content, but they can also be used for awareness campaigns and educational efforts. There is a discussion about the legal and ethical dimensions of cyberbullying. This includes debates about the responsibility of online platforms, the limits of free speech, and the development of legal frameworks to address online harassment. The importance of education and prevention programs is a common theme. Researchers discuss the role of schools, parents, and communities in raising awareness about cyberbullying, promoting digital literacy, and fostering a culture of respect online. Discussions often emphasize the need for mental health support for both victims and perpetrators of cyberbullying.

VII. Conclusion & Future Work

Future research should explore how cyberbullying manifests on these platforms and how their unique features may contribute to or mitigate online harassment. Research can delve deeper into the dynamics of cyberbullying, considering factors such as power imbalances, group dynamics, and the role of bystanders. Understanding the nuances of these interactions can inform more targeted prevention and intervention strategies. Longitudinal studies tracking individuals over time can provide insights into the long-term effects of cyberbullying. This includes examining the persistence of mental health issues, changes in online behavior, and the effectiveness of interventions over an extended period.

References

[1] Akib Mohi Ud Din Khanday, Syed Tanzeel Rabani, et al., “Detecting twitter hate speech in COVID-19 era using machine learning and ensemble learning techniques”, International Journal of Information Management Data Insights, Volume 2, Issue 2, 2022, 100120, ISSN 2667-0968, https://doi.org/10.1016/j.jjimei.2022.100120.

[2] Arvind Kumar Gautam & Abhishek Bansal (2023) “Email-Based Cyberstalking Detection On Textual Data Using Multi-Model Soft Voting Technique Of Machine Learning Approach”, Journal of Computer Information Systems, DOI: 10.1080/08874417.2022.2155267 .

[3] K. S. Alam, S. Bhowmik, et al., "Cyberbullying Detection: An Ensemble Based Machine Learning Approach," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 710-715, DOI: 10.1109/ICICV50876.2021.9388499.

[4] Giri, S., Banerjee, S. “Performance analysis of annotation detection techniques for cyberbullying messages using word-embedded deep neural networks”. Soc. Netw. Anal. Min.13, 23 (2023). https://doi.org/10.1007/s13278-022-01023-2

[5] Sneha Chinivar, Roopa M.S., et al., “Online offensive behaviour in socialmedia: Detection approaches, comprehensive review and future directions”, Entertainment Computing, Volume 45, 2023, 100544, ISSN 1875-9521, https://doi.org/10.1016/j.entcom.2022.100544.

[6] Muneer, A. Fati, S.M. “A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter”. Future Internet 2020, 12, 187. https://doi.org/10.3390/fi12110187.

[7] A. Amamra, J.-M. Robert, et al., “Generative versus discriminative classifiers for android anomaly- based detection system using system calls filtering and abstraction process,” Security. Communication. Network., vol. 9, no. 16, pp. 3483–3495, 2020.

[8] V. Lakovic, ‘‘Crisis management of Android botnet detection using adaptive neuro-fuzzy inference system,’’ Data Science, vol. 7, no. 2, pp. 347–355, Jun. 2020.

[9] P. Meshram and R. Thool, “A survey paper on vulnerabilities in android OS and security of android devices,” in Proc. IEEE Global Conf. Wireless Computing. Network., 2021, pp. 174–178.

[10] Z. Xu, K. Bai, et al., “Taplogger: Inferring user inputs on smartphone touchscreens using onboard motion sensors”, in ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2021.