Utilizing Machine Learning for Detecting Cyber Bullying in Social Media

Saba Yousha*[1], Sajjad Ali Memon[2], Shahnawaz Talpur[3]
[1]Department of Computer Information & Engineering, Mehran University of Engineering and Technology Jamshoro SINDH, Pakistan.
[2]Department of Telecommunication Engineering, Mehran University of Engineering and Technology. Jamshoro SINDH, Pakistan.
[3]Department of Computer System Engineering, Mehran University of Engineering and Technology, Jamshoro, SINDH, Pakistan.
*Correspondence: Saba Yousha, sabayusha@gmail.com

Citation | Yousha. S, Memon. S. A, Talpur. S, "Utilizing Machine Learning for Detecting Cyber Bullying in Social Media", IJIST, Vol. 5 Issue. 4 pp 760-772, Nov 2023
Received| Nov 29, 2023, **Revised|** Dec 05, 2023, **Accepted|** Dec 19, 2023, **Published|** Dec 31, 2023.

The widespread dominance of the Internet and Electronic Media has made Social Media platforms a primary mode of communication. Unfortunately, these platforms have also become breeding grounds for harmful behavior, notably "Cyber Bullying," which involves using technology to inflict disrespect and harm on others. Despite various efforts by researchers to address this issue, the detection of such behavior remains crucial in combating this menace. This study aims to emphasize an effective approach for detecting cyberbullying on Social Media platforms. The findings indicate that the SVM (Support Vector Machine) classifier outperforms other classifiers in this context. We acquired tweet data from Twitter and used significant machine learning techniques to classify and forecast whether tweets are "offensive" or "non-offensive" and after that, using the Support Vector Machine's Algorithm, a machine learning-model is prepared to detect Cyber Bullying on Social Media Platform. This research provide promising results to use ML techniques for detection of Cyber Bullying.

Keywords: Cyberbullying, Support Vector Machine (SVM), Machine Learning, Social Media Platforms, Detection.

Introduction:

Social Media serves as a dynamic platform enabling individuals to share their photos, videos, and daily updates, to remain updated with friends. Several research studies have proposed various methods to identify and address cyberbullying. To understand the efforts made in detecting and stopping cyberbullying, a poll was conducted. Additionally, research was carried out to specifically determine instances of Cyber Bullying and Cyber aggression on Social Media platforms. The findings revealed that approximately 20% of students admitted to being victims of cyberbullying. This behavior has been linked to suicidal tendencies and an increased risk of depressive symptoms [1].

Its widespread use has brought both benefits and challenges, notably the pressing issue of "Cyberbullying." Particularly prevalent among teenagers, Cyber Bullying is characterized as the intentional and persistent use of Information Technology to harm others. This form of harassment encompasses various actions, including divulging private information to cause embarrassment, posting hurtful comments, or sharing damaging content online. The consequences are severe, often leading to depression, self-harm, and tragically, even suicide. Addressing cyberbullying remains critical [2]. While numerous scholars have proposed machine learning algorithms, many overlook the intricacies of social media in their detection methods [3]. Safeguarding communities against escalating cyber threats poses a significant challenge in today's digital era where people increasingly live and interact online. Youngsters are particularly vulnerable targets, facing offensive comments or disrespectful content aimed at tarnishing their reputations [4].

Cyber Criminals exploit social media as a platform for various data breaches, including Cyber Bullying, by leveraging the information available. This online misconduct, whether via phone or the internet, is a growing global concern. Platforms like Twitter serve as spaces for people from diverse backgrounds to express themselves publicly. Sharing of current content and facilitating the exchange of thoughts and ideas are integral aspects of social media [5]. Cyberbullying refers to the use of technology, primarily online platforms, and mobile phones, to harass, threaten, or humiliate another individual. This form of behavior can manifest in various ways, such as sending cruel or threatening text messages, or emails, or posting hurtful content online. The impact of cyberbullying can severely harm the mental health and overall well-being of the victim. Over recent years, cyberbullying has become a growing concern, especially among young people. Efforts have been made to understand and prevent this behavior to mitigate its negative effects [6].

Authors advise against sharing personal information such as home addresses and cell phone numbers online. They stress the importance of swift action to safeguard oneself [7]. While social media offers certain benefits, it also comes with drawbacks, one being the pervasive problem of "cyberbullying" that requires urgent attention. This type of harassment occurs online and is perpetrated by individuals of various ages, including both adults and teenagers. Cyberbullying involves the unauthorized disclosure of private information to shame individuals, along with the propagation of malicious and harmful comments or posts on the internet [8].

The literature review provides an overview of recent studies focusing on strategies for detecting and preventing cybercrime. Researchers have examined user behavior in interactions and public posts on various platforms [9]. Persistent harassment and the influx of negative messages create a toxic atmosphere, leaving individuals feeling unsafe and distrustful. Moreover, the impact of cyberbullying isn't solely psychological; it can significantly affect physical health too. Stress and anxiety induced by bullying online can result in various physical symptoms like frequent headaches, insomnia, loss of appetite, and other health issues, compounding the distress caused by the harassment [10].

Several studies have concentrated on employing Machine Learning techniques to detect Cyber Bullying. For instance, a supervised Machine Learning System utilizing a bag-of-words strategy was developed to evaluate the emotional and contextual factors of statements [11]. Another study utilized a detection graph model and assessment algorithm to differentiate multiple perpetrators and victims, aiming to identify and rate victims and attackers. Future plans involve delving deeper into hidden forms of bullying and advanced trends in this area [12][13]. Cyberbullying stands as a significant challenge in today's society. However, scouring the vast expanse of the Internet to uncover instances of cyberbullying is akin to searching for a needle in a haystack. Regrettably, this leaves cyberbullying victims vulnerable to the harmful effects of negative words, which can lead to feelings of depression, self-harm, and tragically, even suicide [14]. Social Media has become deeply woven into our daily routines, allowing users to foster connections and maintain relationships without physical interaction. With over a billion users utilizing this technology as a primary mode of communication and a source of real-time information, research efforts are primarily focused on addressing the mental impact of social media. These studies are backed by statistical data and offer preventive suggestions to mitigate the challenges associated with its use [15]. The rise and widespread adoption of social media platforms like Twitter, Instagram, Snapchat, Facebook, TikTok, and Tumblr, among others, have provided individuals with an open space to freely express their opinions and thoughts. These platforms, gaining immense popularity, especially among teenagers, have become integral parts of our everyday existence. However, alongside their increased use, these social media platforms also come with certain drawbacks. Young teenagers. In particular, are often exposed to various behavioral and psychological risks due to their extensive engagement with these platforms [16].

In our study, we utilized an innovative approach to detect instances of cyberbullying. We developed a technique specifically tailored for tracking cyberbullying on the Twitter platform. Our proposed methodology involves an area model that incorporates various features from tweets, such as structure, actions, users, and content [17]. These features were used to create an AI classifier capable of distinguishing between Cyber Bullying and Non-Cyber Bullying tweets [18]. Our approach to detect cyberbullying on these platforms involves analyzing a sample of tweets. Our findings indicate that hashtags play a significant role in identifying instances of Cyber Bullying. Utilizing tweet content alongside behavioral features enhances the performance of detection methods. This research area warrants focused attention due to the prevalence of bullies leveraging platforms like Twitter to target others with insults and derogatory remarks [19][20]. Handling instances of bullying on social media requires careful consideration and proactive measures. Our primary objective was to pinpoint instances of cyberbullying within Twitter posts. We established a method to categorize tweets as "bullying" by employing a supervised Machine Learning method, particularly the SVM (Support Vector Machine) algorithm. Our analysis delved into understanding people's behaviors using Machine Learning Algorithms in the context of Cyber Bullying detection [21].

Objectives:

The term 'Cyber Bullying' is broad and holds diverse meanings. Social Media is crucial for identifying bullying. On Twitter criticizing someone might have a negative impact on the victim. This study used a Support Vector Machine Learning Algorithm to detect Cyber Bullying. The findings of this Research suggest that social traits may unveil online bullying. The Paper consists of the following Objectives.

- To extract data from Twitter and to Train the Model using SVM.
- To Evaluate SVM results with bullying and harassment hashtags.
- To compare results with the previous work.

Novelty Statement:

Offensive online content and cyber harassment often intertwine in social interactions, leading to serious consequences. Cyberbullying and online violence can take various forms, from hurtful comments and spreading rumors to sharing offensive images or threats [22]. The impact on individuals can be profound, affecting mental health, Low self-esteem, and sometimes even leading to tragic outcomes. To overcome all the mentioned issues automated systems equipped with machine learning algorithms can aid in the identification and detection of potentially harmful content. These systems can analyze text and user behavior to flag and address instances of cyberbullying more efficiently.

Material and Methods:

This section outlines the adopted methodology as shown in Figure 1 used in our research. The results achieved are showcased in the figure below, illustrating the identification of instances of cyberbullying within the collected dataset.

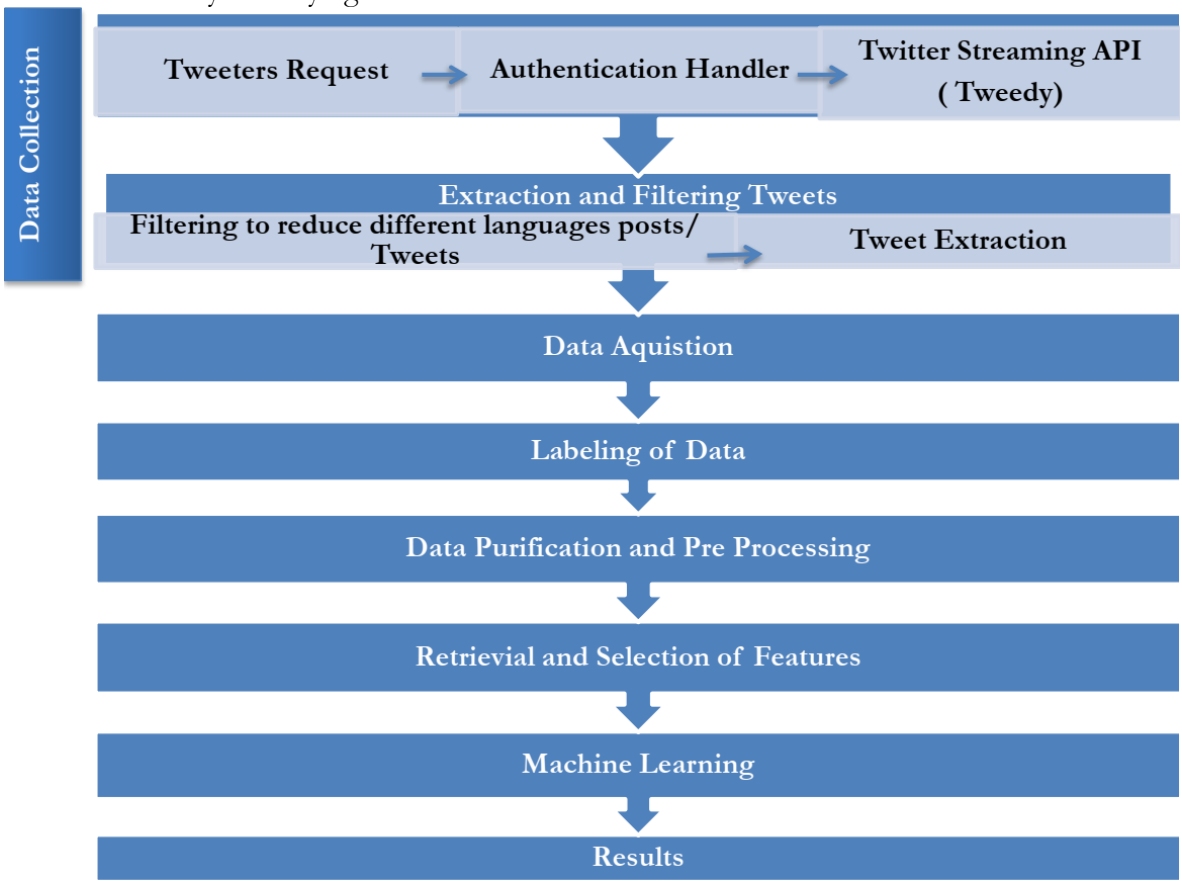


Figure 1. Flow Chart of Methodology

The collection of data involved the use of search terms such as "Cyber Bullying," "Datasets," and "Social Media," among others. Initially, the data was gathered, and certain features that required contextual information were extracted. Various indicators including similarity measures, profanity, and other relevant markers were obtained from Twitter to facilitate the identification of instances of cyberbullying [11][23]. The primary goal of tweet extraction was to convert unstructured data from tweets into structured information for analysis and processing purposes [24].

The rise in cyberbullying and the sharing of suicidal ideation online has highlighted a critical necessity for effective ideation detection to prevent these harmful behaviors. Given the challenges in identifying such content even for humans, the task becomes significantly more complex when attempting automatic detection. However, certain consumers cause violence on

these sites; the main issue in this context is "cyberbullying." The goal is to detect cyberbullying on the social media platforms. The data retrieval process utilized a Twint Scraping tool, allowing data extraction through the Twitter API from specific Twitter accounts. Since training data algorithms are trained on numeric input, the text was first converted into a numerical format and then a label encoder was used. After that, the data set was divided into 80% training and 20% testing, and the SVM Technique was used to run a machine learning algorithm. Many researchers have developed preventative strategies and tactics to tackle this problem, but identification is still necessary to defeat this threat.

Data Acquisition:

Digital Data Acquisition (DDA) or Digital Acquisition System (DAS) involves capturing signals originating from real-world physical occurrences and converting them into a digital format that computer systems and software can process and manipulate. During this process, data is gathered from physical output responses and then transformed electronically into an equivalent digital format without losing any information, making it accessible to data collectors and analysis tools. The conversion of analog signals, often received from various sensors, into digital data is a crucial step in data collection. This transformation involves converting incoming analog signals into digital data that a computer can interpret and analyze. To achieve this, analog signals need to be sampled at discrete time intervals and quantized into predefined values to be stored in the computer's memory as discrete points of information represented by binary integers. Typically, in the process of Digital Data Acquisition, a Digital-to-Analog (D/A) conversion component is employed. This component is often present on a DAQ card integrated within the computer or connected to it through interfaces like Universal Serial Bus (USB) ports. The D/A conversion helps in the conversion of analog signals into digital data that can be effectively processed and analyzed by the computer system.

Labeling of Data:

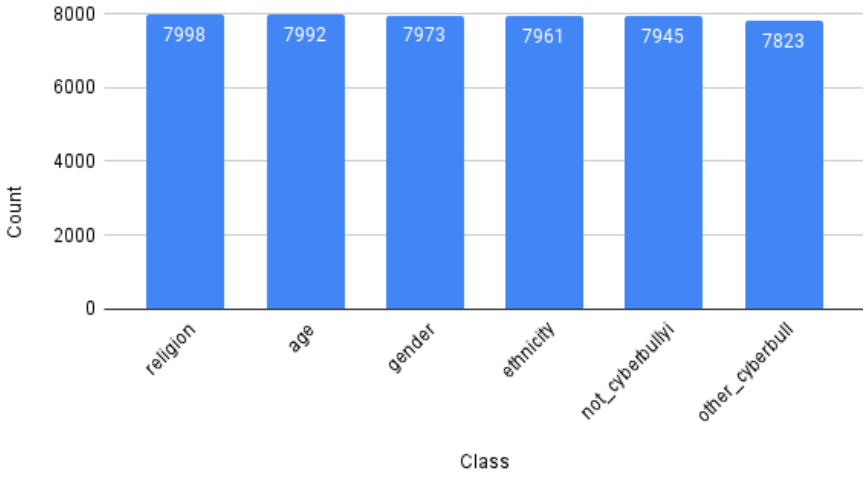
A Tweet is a concise text message limited to 140 characters and shared on Twitter, a widely used microblogging platform. These messages often contain hashtags (e.g., "#win", "#contest", "#Giveaway") that signify the content or theme of the Tweet. Usernames, such as "@office", "@tech", and "@Miss", start with the symbol "@" and represent specific accounts or individuals on Twitter. Data labeling involves marking or categorizing a collection of samples with one or more labels. This process typically takes unlabeled data and supplements it with descriptive tags. These labels can describe various aspects of the data, such as the topic of a news story, the overall sentiment of a Tweet, the language used in group discussions, or even specific elements like identifying objects in an image or determining actions portrayed. In Machine Learning, data labeling is crucial, particularly for supervised learning, where both input and output data are classified and labeled. This labeled data serves as the foundation for training Machine Learning models. To establish reliable learning patterns, a substantial volume of labeled data is often required. These labels assist the model in structuring the data in a way that aligns with the intended outcomes. The process of labeling data is an essential step in preparing the data for Machine Learning. It involves assigning relevant attributes or classifications to the data received from the output, thereby facilitating the training and development of accurate models.

Data Analysis:

The analysis conducted focuses on the assessment of online harassment detection using a Support Vector Machine (SVM) classification algorithm applied to a dataset collected from Twitter. The dataset demonstrates exceptional balance, with nearly equal instances for each class shown in Figure 2 and the Graph according to Cyber Bullying Classes shown in Figure 3. Various metrics such as recall, precision, F-measure, accuracy, and specificity were utilized to evaluate the performance of the model in detecting cyberbullying. SVM was chosen

as the model for its effectiveness in handling social media data related to cyberbullying. The parameters used in the model were based on the initial settings from standard models considered in the original articles. Python 3.7.4 was the programming language employed for the tests, utilizing libraries including NLTK, Pandas, Tweepy, SK-Learn, and others, implemented within research settings. The experimental assessments were conducted on a personal computer equipped with an Intel Core-i5 CPU, Windows 10, and 8 Gigabytes of RAM.

The dataset was preprocessed and split into training and testing datasets using the NLTK Python library. The evaluation was performed across three scenarios, with results indicating an accuracy of 83.26%, precision of 83.66%, and recall of 83.5%. Additionally, cross-validation was employed to gauge the standard deviation of the assessment measures, providing insights into the consistency of the model's performance across different subsets of the data. These metrics were specifically chosen to demonstrate the efficiency of the approach in categorizing tweets related to cyberbullying, highlighting the model's peak performance.



Top 10 most common words chart shows the frequencies fluctuation among top 10 most common words

Figure 2. Cyber Bullying on the Basis of Classes.

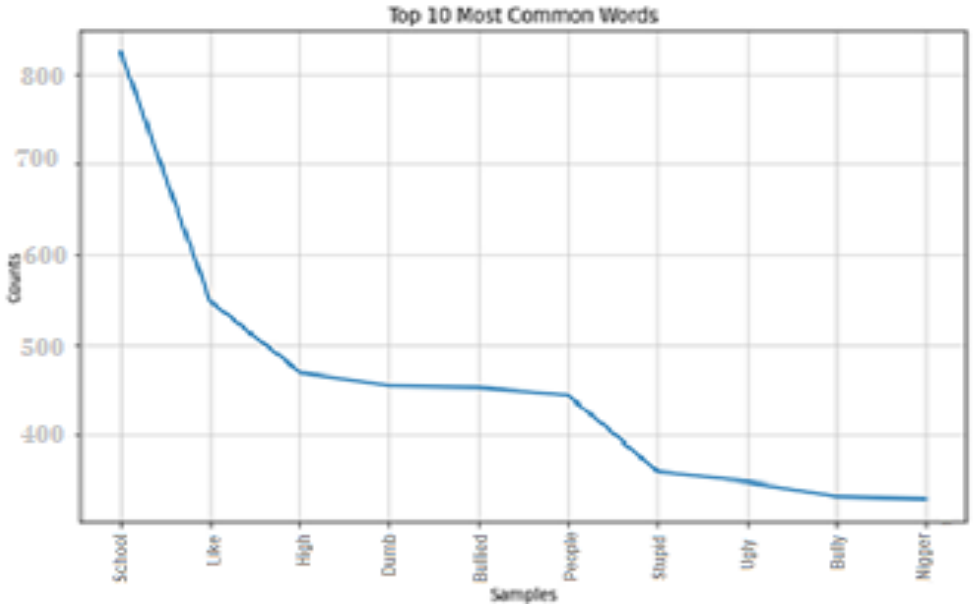


Figure 3. Graph of Cyber Bullying Classes

Word count chart Top 10 words shown in Figure 4 the count of the most common words in the dataset.

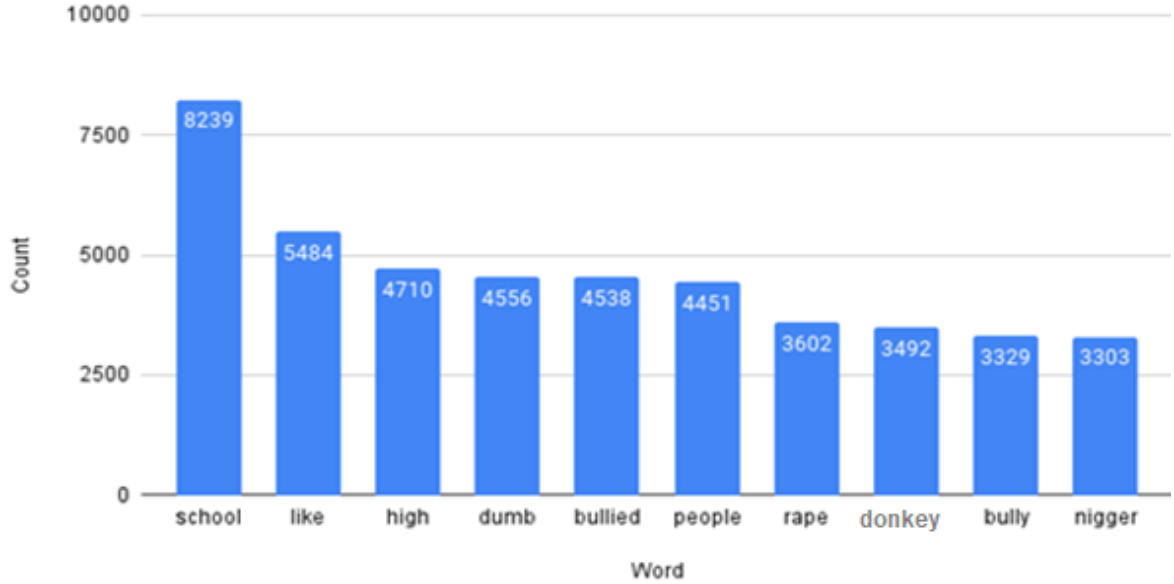


Figure 4. Keywords of Cyber Bullying

The following chart shows in Figure 5 a) the unigram, b) the bigram, and c) the trigram frequency fluctuation chart based on the count between different top 10 grams.

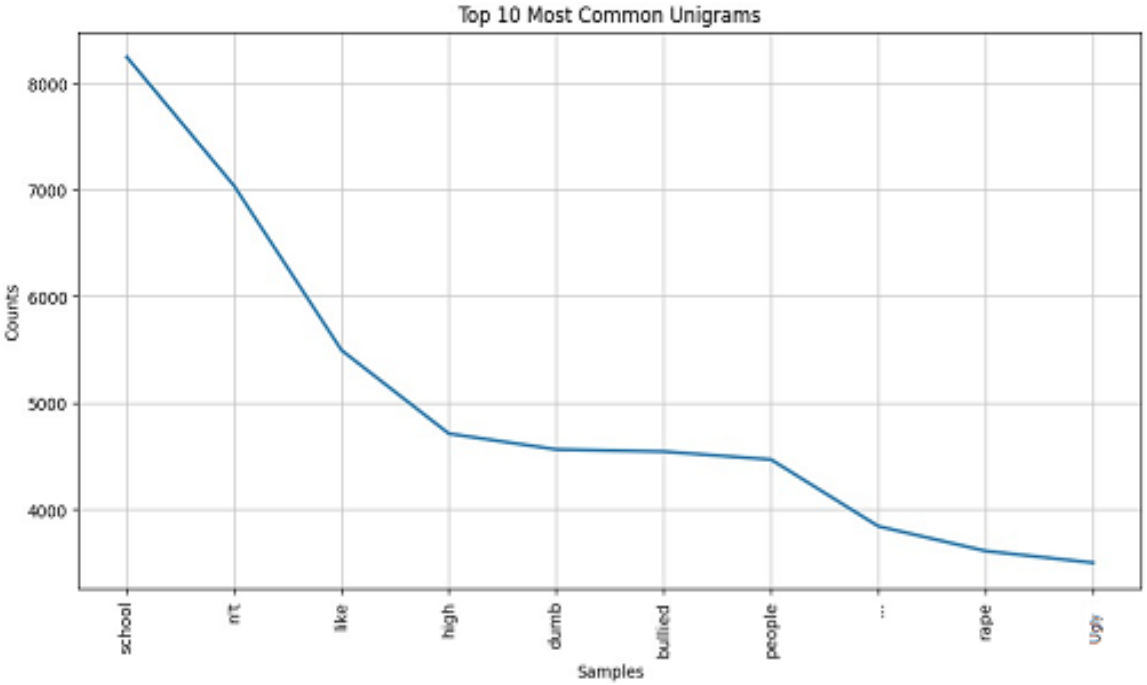


Figure 5. a) Unigram Graph

Retrieval and Selection of Features:

Feature selection is a critical component of the testing process when working with features in a machine-learning context. Its primary goal is to create a predictive model by reducing the number of input variables or features. This technique helps in decreasing the complexity of the model by eliminating unnecessary or redundant aspects, thereby focusing on the most crucial characteristics that are essential for the machine learning algorithm. In supervised learning, feature selection plays a crucial role in determining the most beneficial set of patterns or features that can be utilized to construct an effective framework. By employing

feature selection techniques, we aim to enhance the accuracy of the machine learning process. This is achieved by concentrating on the most relevant and influential factors while excluding excessive or irrelevant ones. Consequently, feature selection not only reduces the computational burden but also improves the algorithm's capability to accurately predict outcomes. There are three primary advantages of feature selection:

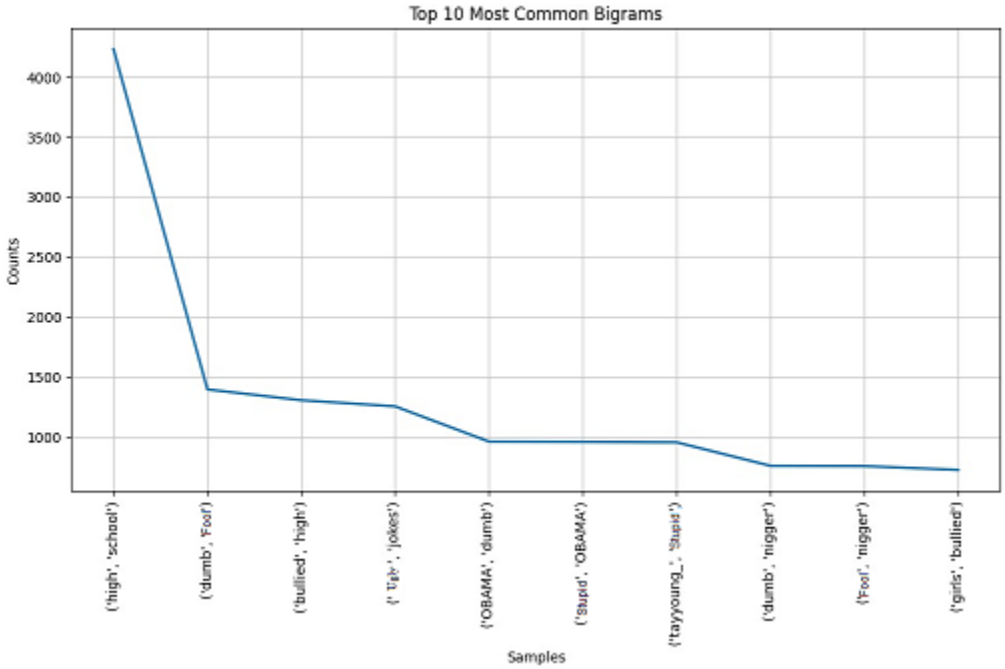


Figure 5. b) Bigram Graph

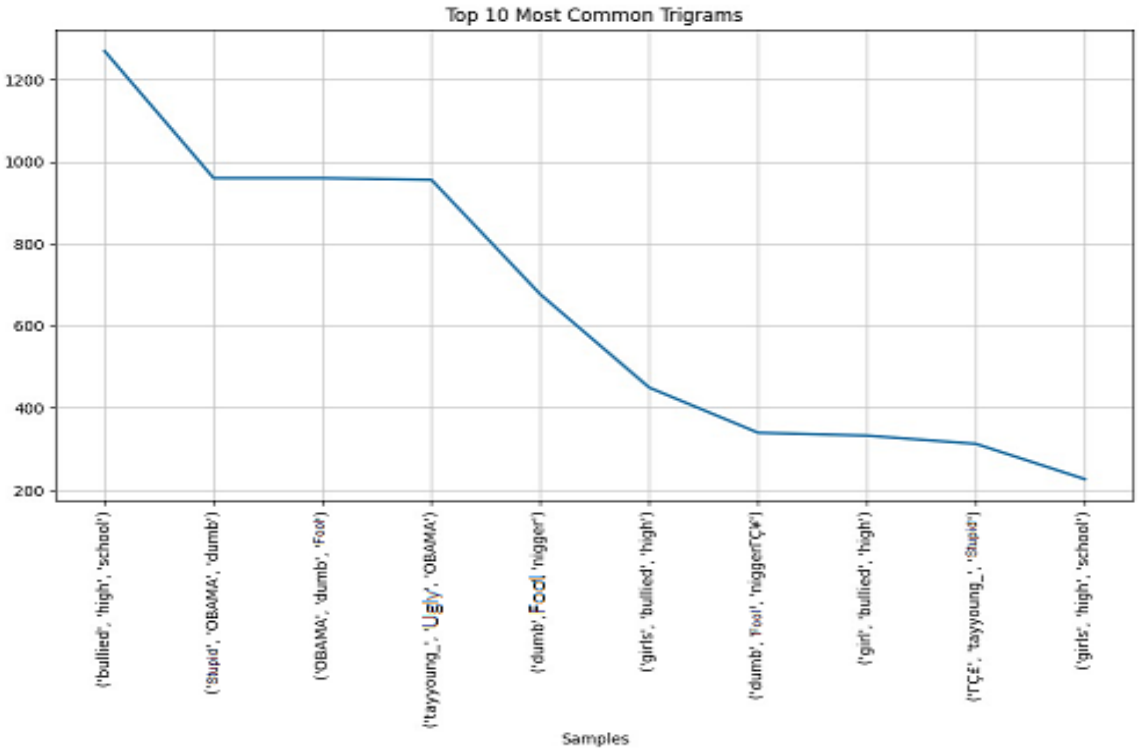


Figure 5. c) Trigram Graph

Improved Model Performance:

By focusing on the most significant features, feature selection enhances the model's predictive accuracy and generalization by reducing overfitting.

Reduced Overhead:

Feature selection decreases the computational and storage requirements, especially in cases where the dataset has a large number of features. This leads to more efficient and faster model training.

Enhanced Interpretability:

Removing irrelevant or redundant features aids in simplifying the model, making it easier to interpret and understand the relationships between the input variables and the predicted outcomes. By leveraging feature selection, machine learning models can be refined and optimized, leading to more efficient and accurate predictions while also enhancing the understanding of the underlying data patterns.

Lowest Class Fitting:

Reducing duplicate data or noise within classes can indeed enhance the quality of model predictions. By minimizing redundant or duplicated information within each class, the model can focus on more representative and essential features, thus reducing the risk of basing decisions on irrelevant or misleading data.

Enhanced Accuracy:

When the model is built on a refined and more relevant set of features, it becomes more accurate in making predictions. By selecting the most critical and impactful features, the model can better capture the relationships within the data, leading to improved modeling evidence and a more accurate representation of the underlying patterns.

Minimized Training Time:

Removing redundant or irrelevant features can significantly reduce the complexity of the model, leading to faster training times. With fewer features to process and consider during training, the algorithm can run more efficiently, leading to quicker model training and evaluation.

Machine Learning:

Machine Learning is a subset of Artificial Intelligence that empowers computers to learn from data and experiences, allowing them to recognize patterns and make predictions autonomously, minimizing human intervention. This approach is particularly useful when handling vast amounts of data, enabling the extraction of valuable insights and trends. In the context of extracting tweets from Twitter. In Figure 6 Machine Learning methods excel in discovering trends and learning continuously from data. Unlike conventional models driven by predefined equations, Machine Learning algorithms directly learn from the data itself, adapting and evolving based on the information they receive. Supervised Machine Learning, a prominent approach, involves training computers on labeled datasets. These labeled datasets contain input-output pairs, enabling the machine to learn from the relationships between inputs and their corresponding outputs. Through this training process, the machine gains the ability to make predictions or classifications based on new, unseen data. The application of the Support Vector Machine (SVM) Algorithm in your scenario is notable. SVM aims to establish an optimal decision boundary or hyperplane that effectively separates data points in a multi-dimensional space, facilitating classification tasks. By identifying this boundary, SVM enables the rapid and accurate classification of new data points, which is beneficial for tasks like tweet classification or sentiment analysis. In essence, employing Machine Learning techniques, particularly supervised learning with algorithms like SVM, allows for the extraction of valuable insights from Twitter data by enabling the system to learn patterns and make predictions based on the learned information.

The text data was converted into numerical form using encoding techniques like label encoding, as Machine Learning algorithms typically require numeric input. The dataset was then divided into 80% for training and 20% for testing purposes. The SVM Machine Learning

Algorithm was applied to classify tweets that were identified as potential instances of cyberbullying, showcasing the precision achieved in diagnosing cyberbullying within the Twitter dataset.

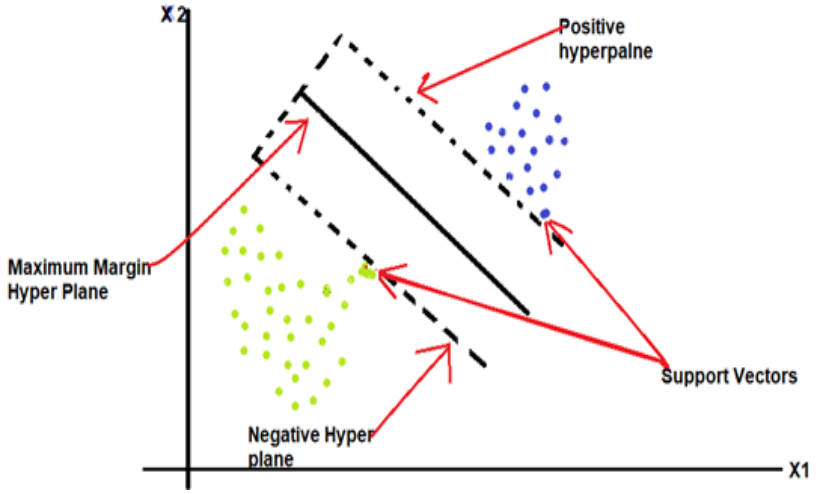


Figure 6. Support Vector Machine

The SVM classifier underwent trials for each input data scenario. Throughout this process, various steps were considered, including class count, word count, data collection, labeling, purification, and retrieval. Ultimately, a trained SVM classifier was obtained as a result of these steps. These sections seem to outline the methodology, evaluation, and outcomes of using Machine Learning, specifically SVM, for cyberbullying detection in Twitter data. The evaluation metrics like accuracy, precision, recall, and confusion metrics were employed to assess the model's performance across different data scenarios, ultimately resulting in a trained SVM classifier for cyberbullying detection.

Results:

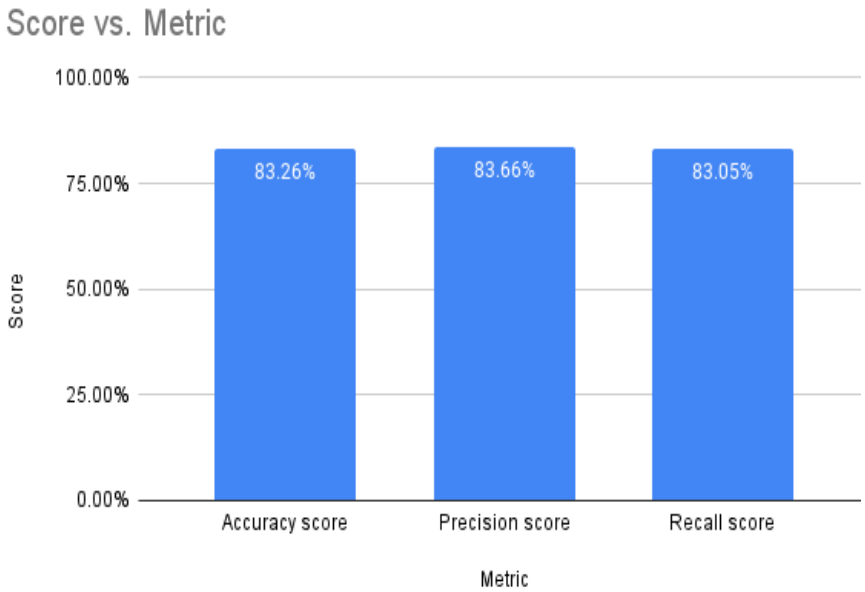


Figure 7. Model Performance on the basis of Accuracy, Precision and Recall

The classifier's performance was evaluated by comparing its testing results with other established machine learning models considered as industry standards. The evaluation involved diverse input data scenarios, validating the prediction results of cyberbullying with accuracy scores of 83.26%, precision scores of 83.66%, and recall scores of 83.05% shown in Figure 7.

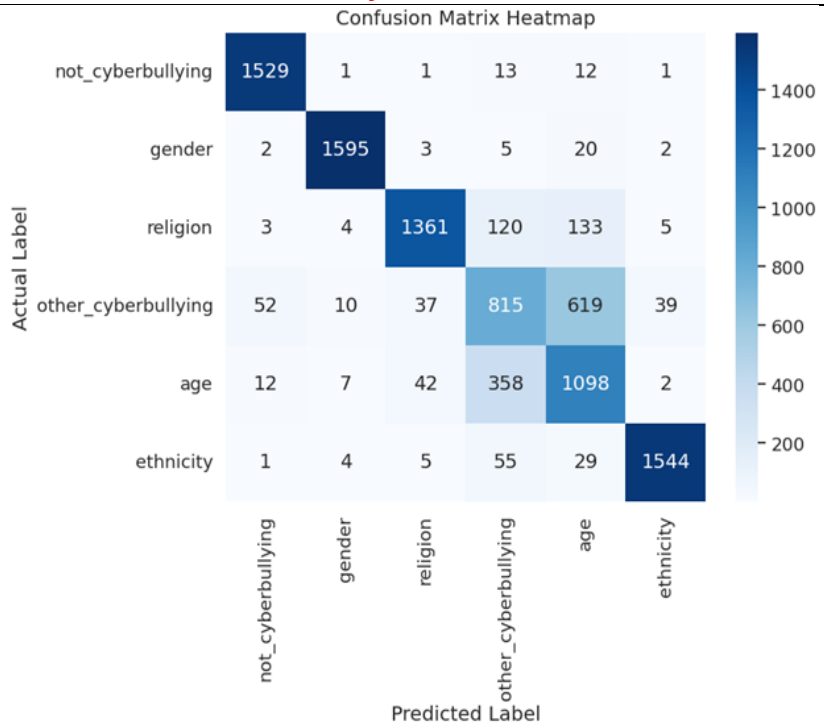


Figure 8. Confusion Matrix of Model

Confusion metrics were used in the evaluation process to assess the model's performance shown in Figure 8. A confusion matrix is a matrix representation that summarizes the predictions made by a model. It displays the number of correct and incorrect predictions for each class, aiding in the understanding of which classes the model is confusing with other classes. This matrix provides a clear breakdown of how well the model performs for each class and helps identify specific instances where the model's predictions might be inaccurate or confused with similar classes.

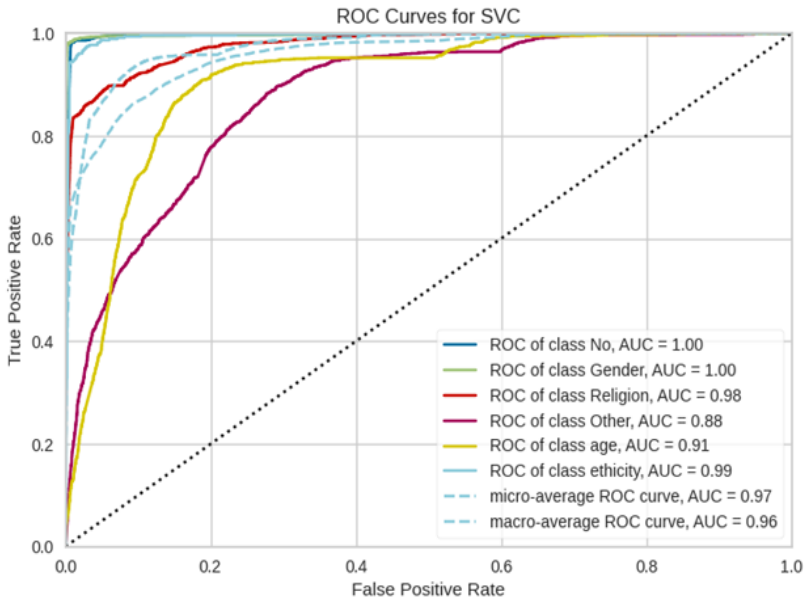


Figure 9. ROC Curve of SVM Model

The Receiver Operating Characteristic Curve (ROC curve) displays an algorithm for classification performs across all categorization levels. Two parameters are shown on this curve in Figure 9: True Positive Rate and False positive rate. Classification algorithms are employed in machine learning techniques to obtain a prediction on the input stream in order

to classify items for further investigation. In numerous instances, the susceptibility or true positive rate (detection/recognition rate) of the classification algorithms is just as crucial as the accuracy of the methods.

Conclusion:

This particular study focused on implementing Machine Learning techniques to detect cyberbullying, a term that encompasses various forms of online harassment and negativity. The concept of "Cyber Bullying" is multifaceted and can have different interpretations. Notably, social media plays a crucial role in identifying instances of bullying. On platforms like Twitter, criticism directed at someone could significantly impact the victim. Previous studies often overlooked the role of Twitter in understanding bullying behavior, prompting our research to include this element. By utilizing Machine Learning methods, we aimed to identify instances of cyberbullying, and as mentioned, the outcomes of our study have been presented earlier. Our findings suggest that certain social characteristics might serve as indicators of potential online bullying. This implies that understanding the social context in which communication occurs is just as critical as understanding the content of the communication itself. In essence, recognizing the social dynamics surrounding interactions on platforms like Twitter can provide valuable insights into potential instances of cyberbullying.

Acknowledgment:

I am deeply grateful to Allah Tallah, the most Gracious and most Merciful, for granting me the opportunity and ability to pursue my studies, as well as for His continuous blessings and gifts.

References:

[1] A. Ali and A. M. Syed, "Cyberbullying Detection using Machine Learning," Pakistan J. Eng. Technol., vol. 3, no. 2, pp. 45–50, 2020, doi: 10.51846/VOL3ISS2PP45-50.

[2] C. Sharma, R. Ramakrishnan, A. Pendse, P. Chimurkar, and K. T. Talele, "CYBER-BULLYING DETECTION VIA TEXT MINING AND MACHINE LEARNING," 2021 12th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2021, 2021, doi: 10.1109/ICCCNT51525.2021.9579625.

[3] Q. Huang, V. K. Singh, and P. K. Atrey, "Cyber bullying detection using social and textual analysis," SAM 2014 - Proc. 3rd Int. Work. Soc. Multimedia, Work. MM 2014, pp. 3–6, Nov. 2014, doi: 10.1145/2661126.2661133.

[4] N. Singh and S. K. Sharma, "Review of Machine Learning methods for Identification of Cyberbullying in Social Media," Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021, pp. 284–288, Mar. 2021, doi: 10.1109/ICAIS50930.2021.9395797.

[5] S. Neelakandan et al., "Deep Learning Approaches for Cyberbullying Detection and Classification on Social Media," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/2163458.

[6] D. Harish, M. Manimaran, V. P. Jayashakthi, and M. Alamelu, "Automatic Detection of Cyberbullying on Social Media Using Machine Learning," 2nd Int. Conf. Adv. Electr. Electron. Commun. Comput. Autom. ICAECA 2023, 2023, doi: 10.1109/ICAECA56562.2023.10201149.

[7] B. A. H. Murshed, J. Abawajy, S. Mallappa, M. A. N. Saif, and H. D. E. Al-Ariki, "DEA-RNN: A Hybrid Deep Learning Approach for Cyberbullying Detection in Twitter Social Media Platform," IEEE Access, vol. 10, pp. 25857–25871, 2022, doi: 10.1109/ACCESS.2022.3153675.

[8] S. B. Shanto, M. J. Islam, and M. A. Samad, "Cyberbullying Detection using Deep Learning Techniques on Bangla Facebook Comments," Proc. 2023 Int. Conf. Intell. Syst. Adv. Comput. Commun. ISACC 2023, 2023, doi: 10.1109/ISACC56298.2023.10083690.

[9] V. Jain, A. K. Saxena, A. Senthil, A. Jain, and A. Jain, "Cyber-Bullying Detection in Social Media Platform using Machine Learning," Proc. 2021 10th Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2021, pp. 401–405, 2021, doi: 10.1109/SMART52563.2021.9676194.

[10] N. Islam, R. Haque, P. K. Pareek, M. B. Islam, I. H. Sajeeb, and M. H. Ratul, "Deep Learning for Multi-Labeled Cyberbully Detection: Enhancing Online Safety," 2023 Int. Conf. Data

Sci. Netw. Secur. ICDSNS 2023, 2023, doi: 10.1109/ICDSNS58469.2023.10245135.
[11] M. M. Islam, M. A. Uddin, L. Islam, A. Akter, S. Sharmin, and U. K. Acharjee, "Cyberbullying Detection on Social Networks Using Machine Learning Approaches," 2020 IEEE Asia-Pacific Conf. Comput. Sci. Data Eng. CSDE 2020, Dec. 2020, doi: 10.1109/CSDE50874.2020.9411601.
[12] M. S. Nikhila, A. Bhalla, and P. Singh, "Text Imbalance Handling and Classification for Cross- platform Cyber-crime Detection using Deep Learning," 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225402.
[13] S. Pericherla and E. Ilavarasan, "Performance analysis of Word Embeddings for Cyberbullying Detection," IOP Conf. Ser. Mater. Sci. Eng., vol. 1085, no. 1, p. 012008, Feb. 2021, doi: 10.1088/1757-899X/1085/1/012008.
[14] E. Idrizi and M. Hamiti, "Classification of Text, Image and Audio Messages Used for Cyberbulling on Social Medias," 2023 46th ICT Electron. Conv. MIPRO 2023 - Proc., pp. 797–802, 2023, doi: 10.23919/MIPRO57284.2023.10159835.
[15] M. Alotaibi, B. Alotaibi, and A. Razaque, "A Multichannel Deep Learning Framework for Cyberbullying Detection on Social Media," Electron. 2021, Vol. 10, Page 2664, vol. 10, no. 21, p. 2664, Oct. 2021, doi: 10.3390/ELECTRONICS10212664.
[16] R. Beniwal, S. Jha, S. Mehta, and R. Dhiman, "Cyberbullying Detection using Deep Learning Models in Bengali Language," 2023 3rd Int. Conf. Intell. Technol. CONIT 2023, 2023, doi: 10.1109/CONIT59222.2023.10205775.
[17] C. S. Wu and U. Bhandary, "Detection of Hate Speech in Videos Using Machine Learning," Proc. - 2020 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2020, pp. 585–590, Dec. 2020, doi: 10.1109/CSCI51800.2020.00104.
[18] S. M. Kargutkar and V. Chitre, "A Study of Cyberbullying Detection Using Machine Learning Techniques," Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020, pp. 734–739, Mar. 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000137.
[19] M. Behzadi, I. G. Harris, and A. Derakhshan, "Rapid Cyber-bullying detection method using Compact BERT Models," Proc. - 2021 IEEE 15th Int. Conf. Semant. Comput. ICSC 2021, pp. 199–202, Jan. 2021, doi: 10.1109/ICSC50631.2021.00042.
[20] R. R. Dalvi, S. Baliram Chavan, and A. Halbe, "Detecting A Twitter Cyberbullying Using Machine Learning," Proc. Int. Conf. Intell. Comput. Control Syst. ICICCS 2020, pp. 297–301, May 2020, doi: 10.1109/ICICCS48265.2020.9120893.
[21] R. Shah and S. K. J. Somaiya, "Machine Learning based Approach for Detection of Cyberbullying Tweets," Int. J. Comput. Appl., vol. 175, no. 37, pp. 975–8887, 2020.
[22] A. Muneer and S. M. Fati, "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter," Futur. Internet 2020, Vol. 12, Page 187, vol. 12, no. 11, p. 187, Oct. 2020, doi: 10.3390/FI12110187.
[23] R. Kumar and A. Bhat, "A study of machine learning-based models for detection, control, and mitigation of cyberbullying in online social media," Int. J. Inf. Secur., vol. 21, no. 6, pp. 1409–1431, Dec. 2022, doi: 10.1007/S10207-022-00600-Y/METRICS.
[24] K. Wang, Q. Xiong, C. Wu, M. Gao, and Y. Yu, "Multi-modal cyberbullying detection on social networks," Proc. Int. Jt. Conf. Neural Networks, Jul. 2020, doi: 10.1109/IJCNN48605.2020.9206663.