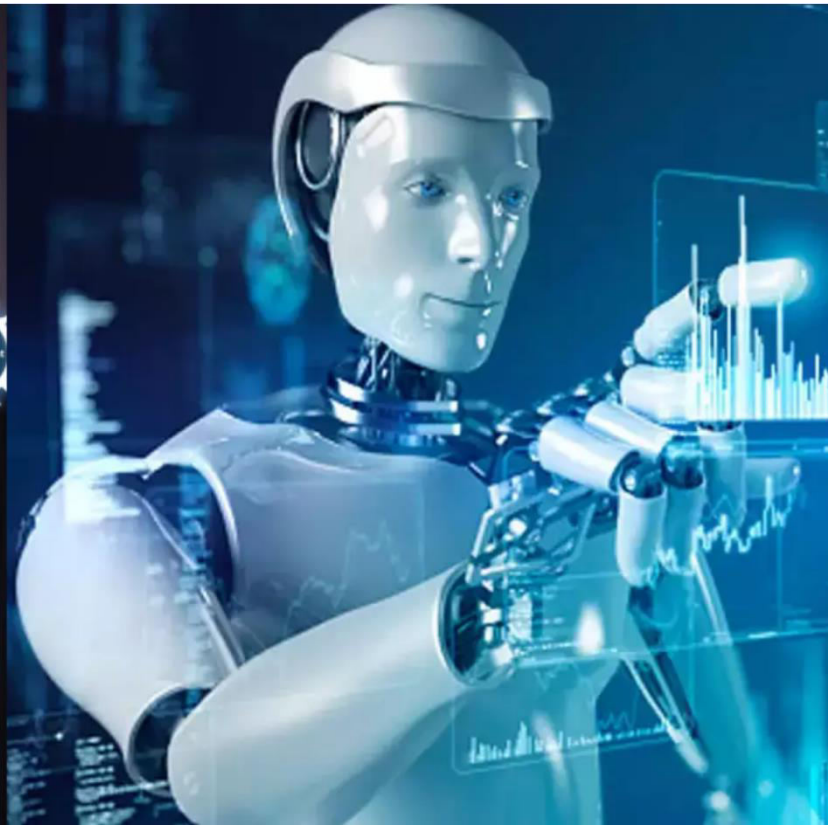
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404493|

Detection of Cyberbullying using NLP and Machine Learning in Social Networks for BI-Language

Nikitha Gs, Amritasri Shenoyy, K Chaturya, Latha Jc, Janani Shree M

Department of CSE, MVJ College of Engineering Bangalore, India

ABSTRACT: Cyberbullying, a contemporary menace within the realm of social media, has emerged alongside the surge in social media usage, leading to the exploitation of freedom of expression. Statistics indicate that a substantial 36.5% of individuals believe they have experienced cyberbullying at some point in their lives. This figure has more than doubled since 2007, and there has been a noticeable increase from 2018 to 2023, indicating an unfavorable trend. Although existing solutions in the market aim to mitigate this issue, they often come with usage limitations or inefficient algorithms.

KEYWORDS: Cyberbullying, Machine Learning, Natural language processing, Hinglish Languages, Social Media.

I. INTRODUCTION

In the dynamic realm of social media, the emergence of cyberbullying presents itself as a prominent and contemporary challenge. The exponential growth in the use of social media platforms has unfortunately transformed these digital spaces, originally intended for free expression, into fertile grounds for the misuse and exploitation of online interactions. Startling statistics indicate that a significant 36.5% of individuals believe they have encountered cyberbullying at some point in their lives, reflecting a notable increase over previous years.

This escalating trend raises concerns about the profound impact of cyberbullying on individuals and society as a whole. Notably, the prevalence of cyberbullying has more than doubled since 2007, with a discernible uptick observed from 2018 to 2023. In response to this burgeoning challenge, various solutions have been introduced to alleviate the detrimental effects of cyberbullying in the digital realm. However, these solutions are not without limitations, often constrained by usage restrictions or hindered by suboptimal algorithms. This paper embarks on a critical exploration, aiming to address the inherent limitations of current cyberbullying detection mechanisms. The overarching objective is to delve into novel approaches that can fundamentally enhance our understanding of cyberbullying and enable the automatic detection of such incidents across tweets, comments, and messages on diverse social media platforms. To achieve this goal, we leverage real-time Twitter data, encompassing headlines, comments, and text messages from trending posts. The foundation of our investigation is a meticulously designed labeling study for cyberbullying.

II. RELATED WORK

The research focuses on developing a technique [1] to detect online abusive and bullying messages using natural language processing and machine learning. It addresses the negative impacts of social media, such as cyberbullying and online harassment, which can lead to serious mental and physical distress. Two distinct features, Bag-of-Words (BoW) and term frequency-inverse document frequency (TFIDF), are utilized to analyze the accuracy of four machine learning algorithms in identifying bullying text or messages on social media. The research may focus primarily on textual data, potentially overlooking other forms of online abuse, such as cyberbullying through images or videos. The effectiveness of the developed technique may be specific to certain languages or cultural contexts, potentially limiting its applicability in diverse online communities. There may be ethical concerns regarding the potential misclassification of non-abusive messages as abusive, leading to censorship or infringement of free speechFurthermore, the coordination packet is assumed to be small enough to be transmitted within slot duration.

A proposed deep learning framework [3] aims to detect cyberbullying in real-time Twitter tweets and social media posts in English, Hindi, and Hinglish. Recent studies show that deep neural network-based approaches are more effective than conventional techniques at identifying cyberbullying content. The system is currently designed to recognize cyberbullying posts only in English, Hindi, and Hinglish, potentially limiting its effectiveness for other

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404493|

languages. The system may struggle to accurately interpret nuanced or sarcastic language used in online communication, which could lead to false positives or negatives in identifying cyberbullying content. The effectiveness of the system may be influenced by the quality and diversity of the training data, potentially leading to biases in its detections. There may be concerns about the privacy of individuals' social media posts, as the system involves evaluating public online content for cyberbullying.

III. METHODOLOGY

1. Natural Language: Processing In this stage, we procured real-time tweets from Twitter, extracted Whatsapp chats, and gathered YouTube comments in both English and Hinglish. The real-time data includes various unnecessary characters. Therefore, as a preparatory step for the detection phase, data cleaning is imperative before applying machine learning algorithms. During the preprocessing stage, we eliminate hashtags, stopwords, numeric data, and hexadecimal patterns, in addition to converting the text to lowercase. This process is facilitated by utilizing NumPy and vectorized functions. We manually compiled a list of stopwords for both English and Hinglish and employed it to filter out these words from the cleaned data. The presence of these extraneous words can adversely impact the accuracy and predictions of the model.

2. Machine Learning In this subsequent phase, we employed diverse machine learning approaches, including Linear SVC, Decision Tree, Naive Bayes, Bagging Classifier, Logistic Regression, Random Forest, Multinomial Naive Bayes, K Neighbors Classifier, and Adaboost Classifier, to train the model and assess the accuracy based on our literature survey findings. We also computed F1 scores for evaluation purposes and refined accuracy through iterative stages. Our aim was to identify the optimal pairing between feature selection methods like TF-IDF and count vectorizer and machine learning models.

IV. ALGORITHMS

Following the steps of preprocessing and feature selection, we conducted a study involving various machine learning models, identifying eight models for comparative analysis based on factors such as popularity, ease of use, training, and prediction time. The classifiers used in the study are outlined as follows:

1. Support Vector Machine (SVM): SVM is a classification algorithm aiming to fit data and determine the best-fitting hyperplane that categorizes or divides data into different classes. It relies on extreme vectors or points, known as support vectors, to create hyperplanes, which influence the orientation and position of the hyperplane. SVM excels in higher-dimensional data, avoiding dimensionality problems, and is less prone to over fitting.

2. K-Nearest Neighbors (KNN): KNN is a straightforward text classification algorithm that categorizes new data by comparing it with all available data using a similarity measure. It establishes classification rules in a tree-shaped form, where internal nodes denote attribute conditions, branches denote conditions for outcomes, and leaf nodes represent class labels. Distance is used to classify a new sample based on its nearest neighbors.

3. Logistic Regression: Logistic regression is a supervised machine learning algorithm used to predict categorical dependent variables by considering a set of independent variables. It predicts whether data belongs to class '1' and provides probabilities rather than exact values.

4. Random Forest Classifier: The Random Forest classifier comprises numerous decision trees, each contributing a class prediction. The class with the most votes becomes the model's prediction. Random Forest is fast, scalable, robust to noise, interpretable, and avoids overfitting, but its real-time prediction capability slows with an increasing number of trees.

5. Bagging Classifier: Bagging classifier, an ensemble machine learning algorithm, mitigates model overfitting by combining the strengths of multiple models. It reduces errors by training each model individually and combining them in an averaging process.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

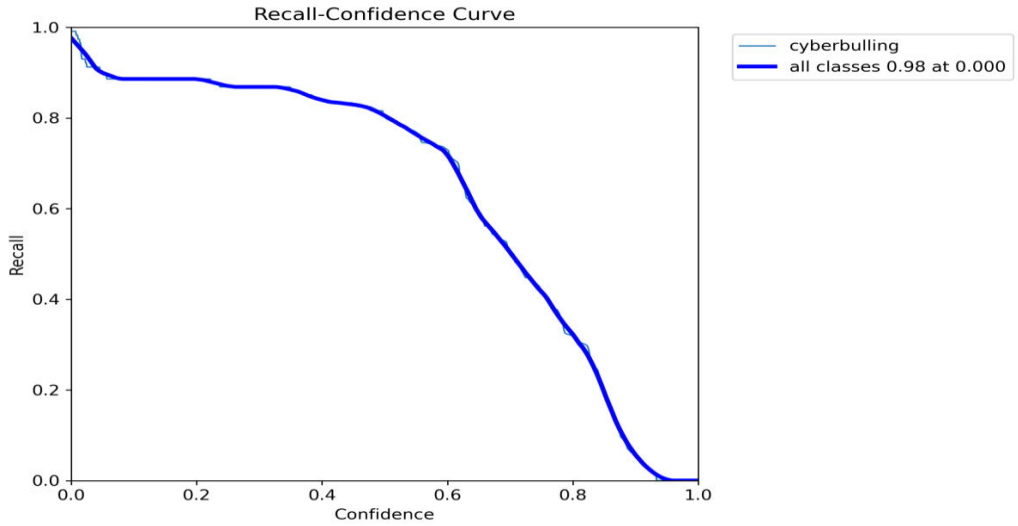
|DOI: 10.15680/IJIRSET.2025.1404493|

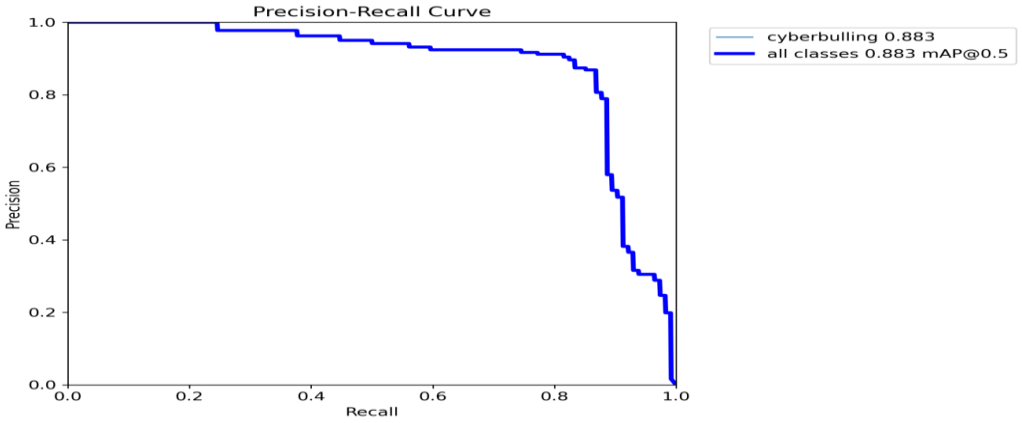
6. Stochastic Gradient Descent (SGD): Classifier SGD is an optimization algorithm used for discriminative learning of linear classifiers under convex loss functions such as SVM and logistic regression. The SGD classifier is a linear model optimized by the SGD optimization method.

7. Adaboost Classifier: Adaboost is an ensemble-boosting classifier that builds a strong learner from the mistakes of weaker models. It employs one level decision trees as weak learners, sequentially adding them to the ensemble. The algorithm corrects predictions by placing more focus on training examples where previous models make predictable errors.

8. Multinomial NB Classifier: The Multinomial NB classifier is a probabilistic machine learning algorithm commonly used in Natural Language Processing (NLP). It predicts the tag of a text, such as a newspaper article or piece of mail, based on Bayes' theorem. The algorithm calculates the probability of a given sample and returns a tag with a high probability, assuming that each feature being classified is not related to any other feature.

V. RESULT AND DISCUSSION





International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404493|

VI. CONCLUSION

Hence, we have successfully managed to extract, clean, and visualize the data using various Python libraries. Our implementation incorporated several natural language processing techniques, including tokenization, lemmatization, and vectorization (feature extraction). Through a thorough review of research papers in this domain, we found that, in feature extraction, count vectorizer and TF-IDF yield notably higher accuracy compared to word2vec and bag of words.

vectorizer marginally outperforms TF-IDF in terms of accuracy. Our exploration also involved identifying and applying various algorithms in our project, such as Support Vector Machine, Logistic Regression, K Nearest Neighbor, Random Forest, Bagging Classifier, Decision Tree Classifier, SGDC classifier, Multinomial Classifier, and AdaBoost Classifier. Following the training of our models, we consolidated the outcomes into a comprehensive plot depicting Accuracy and F1 score for each algorithm.

Upon scrutinizing the results, we observed that Linear SVC and SGD (stochastic gradient classifier) exhibit comparatively superior performance in classifying and predicting bullying messages in Hinglish languages. Notably, these algorithms demonstrate efficient training and prediction times compared to others in the set.

REFERENCES

1. Md Manowarul Islam, Md Ashraf Uddin, Linta Islam, Arnisha Akter, Selina Sharmin, Uzzal Kumar Acharjee, “Cyberbullying Detection on Social Networks Using Machine Learning Approaches”, IEEE-2020.
2. John Hani, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer, Ammar Mohammed, “Social Media Cyberbullying Detection using Machine Learning”, international Journal of Advanced Computer Science and Applications (IJACSA), Volume 10 Issue 5, 2019.
3. Mitushi Raj, Samridhi Singh, Kanishka Solanki, and Ramani Selvanambi, “An Application to Detect Cyberbullying Using Machine Learning and Deep Learning Techniques”, SN Computer Science. 2022
4. Andrea Perera, Pumudu Fernando, “Accurate Cyberbullying Detection and Prevention on Social Media”, Procedia Computer Science Volume 181, 2021
5. Bandari Saichandana, Pille Kamakshi, “Classification of Cyberbullying Detection in Social Networking with Audio using Machine Learning Approach”, International Journal on Recent and Innovation Trends in Computing and Communication, jul 13, 2023
6. Ninad Mehendale, Karan Shah, Chaitanya Phadtare, Keval Rajpara, “Cyber Bullying Detection for Hindi-English Language Using Machine Learning”, SSRN: may 2022
7. Sudheer Panyaram, Muniraju Hullurappa, “Data-Driven Approaches to Equitable Green Innovation Bridging Sustainability and Inclusivity,” in Advancing Social Equity Through Accessible Green Innovation, IGI Global, USA, pp. 139-152, 2025.
8. Amgad Muneer and Suliman Mohamed Fati, “A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter”, Future Internet 2020, 12(11), 187;
9. Nikita Singh, Aishwarya Sinhasane, Sagar Patil and Seetharaman Balasubramanian, “Cyberbullying Detection in Social Networks: A Survey”, (ICCIP) 2020,
10. Tofayet Sultan, Nusrat Jahan, Ritu Basak, Mohammed Shaheen Alam Jony, Rashidul Hasan Nabil, “Machine Learning in Cyberbullying Detection from Social-MediaImage or Screenshot with Optical Character Recognition”, I.J. Intelligent Systems and Applications, 2023,
11. Vaibhav Jain, Ashendra Kumar Saxena, Athithan Senthil, Abhishek Jain and Arpit Jain, “Cyber-Bullying Detection in Social Media Platform using Machine Learning”, IEEE 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details