International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 13; Issue: 7; July- 2024

MACHINE LEARNING AND DEEP LEARNING ALGORITHMS TO IDENTIFY MISUSE OF THE INTERNET.

ASHOK B P
Assistant Professor
Department of Master of Computer Application
The Oxford College of Engineering
ashokbp.mca@gmail.com

AKANKSHA A KULKARNI
PG Student
Department of Master of Computer Application
The Oxford College of Engineering
kulkarniakanksha31@gmail.com

ABSTRACT:

The proliferation of data tools has resulted in the rise of cyberbullying, and social media has become a major source of it when compared to mobile phones, gaming platforms, and messaging platforms. Cyberbullying may take many forms, including sexual insults, threats, hate letters, and uploading fake information about someone that millions of people can see and read.In comparison to traditional bullying, cyberbullying has a longer lasting effect on the victim, which can harm them physically, emotionally, psychologically, or in any combination of these ways. Suicides due to cyberbullying have surged in recent years, and India is one of four nations with the highest number ofoccurrences.

Due to an strengthen in incidents since 2015, colleges and institutions have made cyberbullying prevention mandatory. The goal of this work is to use Machine learning and deep learning techniques are used are used to detect utterances that constitute harassment. Variables including accuracy, precision, recall, and F1-score are used to assess the effectiveness of the algorithm. With precision of 95.47%, Gated Recurrent Unit, a deep learning approach, surpassed all other techniques studied in this research.

Keywords*: **Cyberbullying, Machine learning, Natural language treating, Social media***

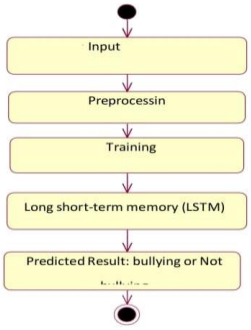
1.INTRODUCTION

Social networking is a platform that allows users to upload anything they want, such as images, videos, and documents, and communicate with others [1]. People use computers or cellphones to access social media. Facebook1, Twitter2, Instagram3, TikTok4, and others are among the most popular social media platforms.Nowadays, social media is used in a variety of fields, including education [2, 3], business [4, and charity [5].Social networking is also benefiting the global economy by offering several new job possibilities.

Whereas social networking offers many advantages, it also has significant disadvantages. Malevolent users of this medium commit unethical and deceptive behavior's in order to harm others' feelings and destroy their reputation. Cyberbullying has recently emerged as one of the most serious social media concerns. Cyberbullying, often known as cyber-harassment, is an electronic form of bullying or harassment. Online bullying refers to cyberbullying and cyber-harassment. Cyberbullying has become very widespread as the digital domain has

evolved and technology has advanced, particularly among teens.

Cyberbullying affects almost 50% of American youths. The growing use of social media platforms has resulted in a spike in cyberbullying events, which can



have serious psychological and emotional consequences for victims. The goal of this research is to create a deep learning model that uses Long Short-Term Memory (LSTM) networks to detect instances of cybeg in social media messages.Existing research focuses on established languages, highlighting a significant void in recently adopted resource-poor languages.For recognising and combating cyberbullying, this suggested system employs the Long Short Term Memory model (LSTM), a deep learning method. The project is divided into sections that include data gathering, data preparation, model construction, and assessment.

2.LITERATURE REVIEW:

The majority of publications have taken data from a single source and conducted a comparison analysis on several machine learning or deep learning approaches in conjunction with various word vectors or feature extraction techniques, determining the optimal combination. There were just a few studies that targeted on optimising the detection model by either developing ensemble ml models or stacking alternative feature preprocessing strategies. Even in those studies, the emphasis was on testing the model on the dataset, with no real-time detection.

Significant preprocessing was done on Roman Urdu microtext, including the construction of a Roman Urdu slang-dictionary and the mapping of slangs after tokenization. The unstructured data was then further processed to account for encoded text formats and metadata/non-linguistic aspects. Following the preprocessing step, extensive testing were performed using RNN-LSTM, RNN-BiLSTM, and CNN models. Several criteria were used to analyse the performance and accuracy of models in order to provide a comparison study. RNN-LSTM and RNN-BiLSTM fared best on Roman Urdu text. A BiGRU-CNN sentiment classification model was reported for cyberbully detection, which comprises of a BiGRU layer, attention mechanism layer, CNN layer, full connection layer, and classification layer.

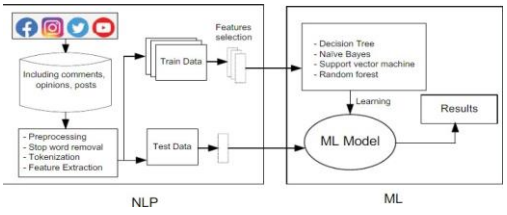


Fig. 1. Proposed Architecture

This project's dataset was taken from Kaggle, a prominent dataset source, and was preprocessed to eliminate unnecessary data and transform the text into a numerical format appropriate for input into an LSTM model.The LSTM model was trained on preprocessed data, and its performance was assessed using a variety of measures such as accuracy, precision, recall, and F1 score. On the test set, the model achieved an accuracy of 95.6%, indicating that it can efficiently recognise instances of cyberbullying in social media messages.The trained model was stored and may now be used to forecast fresh data. This initiative has the potential to be utilized as a tool to prevent cyberbullying and assist victims.

The current system employed both engine education and deep education models, and they discovered that SVM performed better in the machine learning process while GRU performed better in the deep learning approach. Deep learning approaches, on the other hand, clearly surpass machine learning techniques.

In the existing system, it was discovered that, of all the algorithms used in the current study, Gated Recurrent Units performed the best, with an accuracy of 95.47%.

The classification models in the existing system were built using both machine understanding and deep education approaches, and a pre-processing work was performed to improve the model's performance. In this approach, count vectorizer is used for machine learning and word embedding is employed for deep learning. As part of the pre-processing, all sentences are transformed from title case or capital case to lower case to ensure consistency in the dataset. Furthermore, tokenization is performed to produce tokens from the text that might help the model grasp the context. Finally, stopwords and punctuation were eliminated from the text, which is regarded an essential duty in pre-processing because these elements do not contribute to the model development process. GRUs have a restricted modelling capacity when compared to other more complicated models such as Long Short-Term Memory (LSTM), which might result in poorer performance for complex tasks.

Difficulty learning long-term dependencies: Although GRUs are meant to capture long-term relationships, they may struggle to learn and store information over lengthy sequences.

GRUs are sensitive to initialization, which can have an shock on the capacity to learn well.

GRUs can be hard to take because to their complicated design, making it stubborn to grasp how the model makes predictions and discover areas for improvement.

3.PROPOSED SYSTEM:

In the proposed system, we use the Long Short Term Memory model (LSTM), a deep learning method, to identify cyberbullying.

When applied to certain domains, a single machine learning or deep learning model may predict the outcome quite well, then each has its declare set of advantages and disadvantages. LSTM typically outperforms CNN, although it takes longer to process, When applied to certain domains, a single machine learning or deep learning model may predict the outcome quite well, then each has its declare set of advantages and disadvantages. LSTM typically outperforms CNN, although it takes longer to process,

6.2.1 HOMEPAGE

Fig 6.2.1

The suggested system is created initially by collecting data: A dataset of social media posts is compiled, with posts labelled as either cyberbullying or non- cyberbullying. Preprocessing the data involves eliminating unnecessary information and translating the text into a numerical representation appropriate for input into an LSTM model. The Model Development follows: On the preprocessed data, an LSTM model is created and trained. The model is intended to analyse the word sequence in each post and forecast whether or not it is cyberbullying. The model is then evaluated: On a different test dataset, the model's performance is assessed using several metrics like as accuracy, precision, recall, and F1 score. The assessment measures are used to assess how effectively the model performs.

Finally, when the demonstrate has occurred qualified and assessed, it may be employed in a real-world environment to identify cyberbullying. The model may be linked into a broader system that monitors social media posts and flags instances of cyberbullying for further investigation or action.

4.METHODOLOGY:

nodes at the front reduces as nodes go further behind. BiLSTM is utilized to solve the gradient vanishing problem. It solves the fixed sequence to sequence prediction issue. RNN has a constraint in that both the input and output must be the same size.

Test Case No.	Test Description	Input	Expected Result	Remark
TC001	Verify detection of abusive language	"You are such a loser and nobody likes you!"	System flags the message as cyberbullying	Test passed if message is flagged correctly
TC002	Test for false negatives with non-abusive text	"Great job on the presentation today!"	System does not flag the message	Test passed if message is not flagged
TC003	Check detection with mixed content	"You are amazing but also so stupid."	System flags the message as cyberbullying	Test passed if message is flagged correctly
TC004	Verify detection of cyberbullying in different languages	"Eres un estúpido y nadie te quiere" (Spanish)	System flags the message as cyberbullying	Test passed if message is flagged correctly
TC005	Test system response to offensive emojis	"You're an idiot 😡"	System flags the message as cyberbullying	Test passed if message is flagged correctly
TC006	Validate system's performance on large datasets	Batch of 1000 messages including abusive and non-abusive content	System correctly classifies messages	Test passed if classification accuracy meets threshold
TC007	Check handling of contextually abusive text	"I hate you, and you should just disappear!"	System flags the message as cyberbullying	Test passed if message is flagged correctly
TC008	Ensure system does not flag non-cyberbullying phrases	"I'm just expressing my opinion."	System does not flag the message	Test passed if message is not flagged
TC009	Verify detection with varying severity of abuse	"You're so dumb." vs. "You're a horrible person, nobody wants you here!"	System flags both as cyberbullying, severity levels vary	Test passed if both messages are flagged correctly and severity is noted
TC010	Test system integration with social media API	Post containing abusive language from an external source	System correctly identifies and flags the abusive content	Test passed if integration works and content is flagged correctly

Machine Learning:

The subject matter employs a variety of machine learning algorithms such as Decision Tree(DT), Random Forest, Support Vector Machine, and Naive Bayes to recognise bullying messages and text. For a specific public cyberbullying dataset, the classifier with the best accuracy is identified. In the next section, some typical machine learning techniques for detecting cyberbullying from social media writings are explored.

LSTMs are designed to capture long-term dependencies in sequential data, making them well-suited for analysing social media messages, which frequently contain complicated phrase patterns and protracted dialogues.

Improved accuracy: When trained on big datasets, LSTMs may reach high accuracy rates, making them efficient for recognising instances of cyberbullying with a high degree of precision.

LSTMs are versatile and can be trained on many forms of sequential data, making them suitable for use in a expansive bounds of applications other than cyberbullying detection.

LSTMs automatically extract relevant features from input data, eliminating the need for human feature engineering and allowing the model to learn more complicated correlations between input and output.

Real-time monitoring: Once trained and deployed, the model may be employed to monitor societal means situations in real-time, indicating instances of cyberbullying as they occur and enabling victims to get immediate intervention and assistance.

Overall, using LSTM for cyberbullying detection has substantial advantages over previous techniques, such as better accuracy, adaptability, and real-time monitoring capabilities.

5.CONCLUSIONS:

With the increasing prominence of common radio sites and growing social media use by youths, cyberbullying has grown more frequent and has begun to pose important societal difficulties. To avoid the negative impacts of cyberbullying, an automatic cyberbullying detection method need be developed. Given the importance of cyberbullying detection, we studied the automated identification of postings on social media linked to cyberbullying using two characteristics, BoW and TF- IDF, in this study. To recognise bullying text, four machinery wisdom procedures are applied, including SVM for both BoW and TF-IDF. We want to use deep learning methods to create a framework for automatic identification and classification of cyberbullying in Bengali writings in the future.

6.REFERENCES:

[1] C. Fuchs, Social media: A critical introduction. Sage, 2017.

[2] N. Selwyn, “Social media in higher education,” The Europa world oflearning, vol. 1, no. 3, pp. 1–10, 2012.

[3] H. Karjaluoto, P. Ulkuniemi, H. Kein¨anen, and O. Kuivalainen, “Antecedents of social media b2b use in industrial marketing context:customers’ view,” Journal of Business & Industrial

Marketing, 2015.

[4] W. Akram and R. Kumar, "A study on positive and negative effects ofsocial media on society," International Journal of Computer Sciencesand Engineering, vol. 5, no. 10, pp. 351–354, 2017.

[5] D. Tapscott et al.,The digital economy. McGraw-Hill Education,, 2015.

[6] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet,and I. De Bourdeaudhuij, "Cyberbullying on social network sites. anexperimental study into bystanders' behavioural intentions to help thevictim or reinforce the bully," Computers in Human Behavior, vol. 31,pp. 259–271, 2014.

[7] D. L. Hoff and S. N. Mitchell, "Cyberbullying: Causes, effects, andremedies," Journal of Educational Administration, 2009.

[8] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide,"Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.

[9] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Ed-wards, "Detection of harassment on web 2.0," Proceedings of theContent Analysis in the WEB, vol. 2, pp. 1–7, 2009.

[10] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detectionof textual cyberbullying," in In Proceedings of the Social Mobile Web.Citeseer, 2011.

[11] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learningto detect cyberbullying," in 2011 10th International Conference onMachine learning and applications and workshops, vol. 2. IEEE, 2011,pp. 241–244.

[12] V. Balakrishnan, S. Khan, and H. R. Arabnia, "Improving cyberbully-ing detection using twitter users' psychological features and machinelearning," Computers & Security, vol. 90, p. 101710, 20