iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53951

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Machine Learning and Deep Learning Algorithm for Online Bullying Identification

Swaroop Gowda D[1], Ravi Dandu[2]
[1]*School of CSA, Reva University, Bengaluru, India*
[2]*Associate Professor, School of CSA, Reva University, Bengaluru, India*

Abstract: ***The popularity of information technologies has led to bullying in cyberbullying and social media has become its main place compared to mobile phones, gaming platforms and messaging. Cyberbullying can take many forms, including sexual harassment, threats, hate mail, and posting false information about a person that millions of people can see and read. Compared to bullying, cyberbullying has a longer-term impact on victims, which can affect them physically, emotionally, psychologically or in combination. Suicides due to cyberbullying have been on the rise in recent years, and India is among the four countries with the highest incidence.***
Due to the increasing number of incidents since 2015, universities and schools have made it mandatory to prevent cyberbullying. The goal of this project is to identify cyberbullying conversation using machine learning and deep learning. Evaluate model performance using metrics such as accuracy, precision, recall, and F1 score. Gated Recurrent Units, a deep learning method, outperformed all methods in this study with 95.47% accuracy.
Cyberbullying, Machine learning, Natural language processing, Social media.
Keywords: Social media, Machine learning, Cyberbullying, Natural language processing.

I. INTRODUCTION

Social Media is a platform that allows users to upload anything they want such as photos, videos, documents, texts and communicate with others [1]. People use computers or mobile phones to access social media. Facebook1, Twitter2, Instagram3, TikTok4 etc. are some of the most popular social media platforms.

Today, social media is used in many ways, including education [2, 3], business [4], and charity [5]. Social networking is also beneficial to the global economy by providing many new jobs.

Although social networks have many advantages, they also have disadvantages. Bad users of this environment engage in dishonest and fraudulent acts to harm others and damage their reputation. Cyberbullying has recently become one of the most pressing social problems.

Cyberbullying, often referred to as cyber harassment, is an electronic form of bullying or harassment. Cyberbullying refers to cyberbullying and online harassment. With the growth of the digital world and the advancement of technology, cyberbullying has become common, especially among young people.

Cyberbullying affects nearly 50% of American teens. Bullying affects victims physically and mentally.

Due to the trauma of cyberbullying, people become victims of self-harm, such as suicide. For this reason, detecting and preventing cyberbullying is important for the protection of young people. Here, we present a machine learning-based cyberbullying detection model that can determine whether the communication is related to cyberbullying. We explore various machine learning methods in cyberbullying detection models, such as Naive Bayes, Support Vector Machines, Decision Trees, and Random Forests. Testing is done using two sets of data from Twitter and Facebook messages and posts.

We perform the performance analysis using two independent functions: BoW and TF-IDF. The findings show that TF-IDF features outperform BoW in terms of accuracy, while SVM outperforms the method used in this article.

The increasing use of social media has led to the emergence of cyberbullying, which can be devastating for victims. The purpose of this research is to develop a deep learning model using a short-term ad hoc (LSTM) network to identify bullying patterns in social media. Current research has focused on established languages, highlighting major differences in recently adopted languages.

Suggested use of short-term memory model (LSTM), a deep learning method, to identify and combat cyberbullying. The project is devoted to data collection, data preparation, and design and site evaluation.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

II. LITERATURE REVIEW

Most publications take data from a single source and make comparisons between different machine learning and deep learning with different languages or layers to determine the best combination. Only a few studies focus on visualizing good models by developing machine learning models or stacking proxy feature pre-processing techniques. Even in these studies, the goal is to test the model of datasets without real-time detection.

Significant progress has been made on Roman Urdu micro text, including the creation of the Roman Urdu slang dictionary and the slang map after tokenization. Unstructured data is then processed to account for textual content and metadata/non-textual content. After preliminary steps, extensive testing is done using RNN-LSTM, RNN-BiLSTM and CNN models. Various methods were used to validate the effectiveness and accuracy of the models to provide a comparative study. RNN-LSTM and RNN-BiLSTM perform best in Roman Urdu text.

The BiGRU-CNN classification theory model has been published for cyberbullying detection, which consists of BiGRU layer, layered layer, CNN layer, fully connected layer and layered layer.

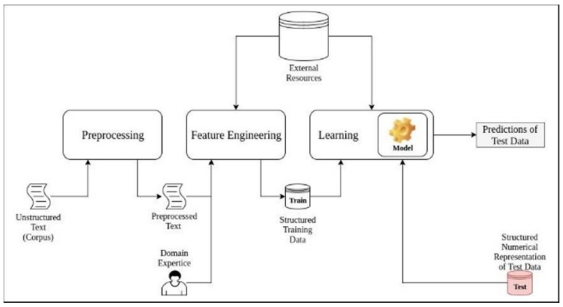


Figure 1. Proposed Architecture

III. EXISTING SYSTEMS

The data for this project is from the well-known data source Kaggle, pre-processed to remove redundant data and convert the text to numbers suitable for entry into the LSTM model. LSTM models are first trained on the data and their performance is evaluated using various criteria such as accuracy, precision, recall, and F1 score.

On the test set, the model achieved 95.6% accuracy, showing that it can identify the nature of cyberbullying in social media messages. The training model is stored and can now be used to predict new data. The initiative can be used as a tool to prevent cyberbullying and help victims.

Existing methods use both machine learning and deep learning models, and they found that DVMs outperformed machine learning, while GRUs performed better in deep learning.

On the other hand, deep learning clearly outperforms machine learning techniques.

In the current system, Gated Recurrent Units was found to be the most efficient among all algorithms used in this study, with an accuracy rate of 95.47%.

The classification model in the current system is based on machine learning and deep learning, and preliminary studies have been made to improve the performance of the model one.

This method uses computational victimizers for machine learning and word embedding's for deep learning. As part of pre-processing, all sentences are capitalized or changed from uppercase to lowercase to ensure consistency in data. In addition, tokenization is used to generate markers from the text that can help the model understand the content. Finally, the stop words and punctuation marks, which were thought to have an important role in the first one, since they did not contribute to the structure model, were removed from the text. Compared to other complex models such as Long Short Term Memory (LSTM), the

GRU has modelling capability that can lead to poor performance of complex tasks.

Long-term learning is difficult: Although GRUs are designed to capture long-term relationships, long-term data can be difficult to learn and retain.

GRUs are sensitive to baseline, which will affect their training.

GRUs can be difficult to interpret due to their nature, making it difficult to understand what the model predicts and see areas for improvement.

Difficulty with noisy data: GRUs can be affected by noisy data, which affects their ability to produce accurate predictions.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

IV. RECOMMENDED SYSTEM

In the proposed system, we use the deep learning Long Term Memory Model (LSTM) to describe cyberbullying.

The purpose of this proposal is to design and develop a cyberbullying system that can be used to identify and analyse the nature of online bullying by social media users. The plan will be the best model for detecting cyberbullying posts on social media platforms, thus overcoming many shortcomings in the detection process in the past. In addition, the proposed system receives effective training models and word usage patterns that make this system unique. This technology facilitates the analysis of cyberbullying rates on various social media sites to implement the necessary protection.

❖ Planning process that begins with data collection: collecting data on social media posts that are not labelled as harassment.

Data pre-processing involves removing redundant data and converting the text into a representation suitable for entry into the LSTM model. The design is as follows: On the previous data, the LSTM model is built and trained. The model is designed to analyse the order in each post and predict whether it is cyberbullying. Then evaluate the model: evaluate the performance of the model using various parameters such as accuracy, precision, recall and F1 score in different tests. Evaluation is used to evaluate the performance of the model.

Finally, after the model has been trained and evaluated, it can be used to detect cyberbullying in a real environment. The model could be extended to a broader system that monitors social media and flags instances of cyberbullying for further investigation or action.

V. METHODOLOGY

A single machine learning or deep learning model can very well predict results when applied to certain things, but each model has its own advantages and disadvantages. LSTMs generally outperform CNNs despite requiring longer processing time, fewer hyper parameters, and less maintenance. Meanwhile, LSTM is more accurate for longer sentences, but takes more time to analyse. Because RNNs suffer losses while processing sequences, their knowledge of previous nodes decreases as nodes move back. BiLSTM is used to solve the vanishing gradient problem. It solves the array-to-array prediction problem. RNNs have the limitation that the input and output must be the same size.

A. Natural Language Processing

The actual message or message contains many characters or text.

For example, numbers and grammar have little effect on bullying. Before we can apply machine learning techniques to them, we need to clean up the analytics and prepare it for discovery. At this stage all special characters like stop words, symbols and numbers are removed, tokenization, rooting etc. many operations are performed.

B. Machine Learning

This topic uses different machine learning algorithms such as decision trees (DT), random forests, support vector machines, and pure Bayes to analyse offensive words and text. For a given public cyberbullying dataset, the best performing classifier is determined. The next section will examine some machine learning techniques for detecting cyberbullying from social media. The LSTM is designed to capture long-term dependencies in linked data, making it ideal for analysing social media, which often includes complex sentences and long conversations.

C. Improved Accuracy

LSTMs can achieve highly accurate results when trained on large datasets, making them effective in identifying cyberbullying incidents with high accuracy. The LSTM is versatile and can learn many types of data connections, making it suitable for many applications beyond cyberbullying detection. The LSTM automatically extracts relevant features from the input data without special engineering training, allowing the model to learn more about the relationship between input and output.

D. Real-time Monitoring

Once trained and implemented, the model can be used to monitor social media in real-time, report cyberbullying incidents as soon as they occur, and respond and assist victims immediately.

Overall, cyberbullying detection using LSTM has advantages over previous methods such as better accuracy, flexibility, and real-time monitoring capabilities.

©IJRASET: All Rights are Reserved

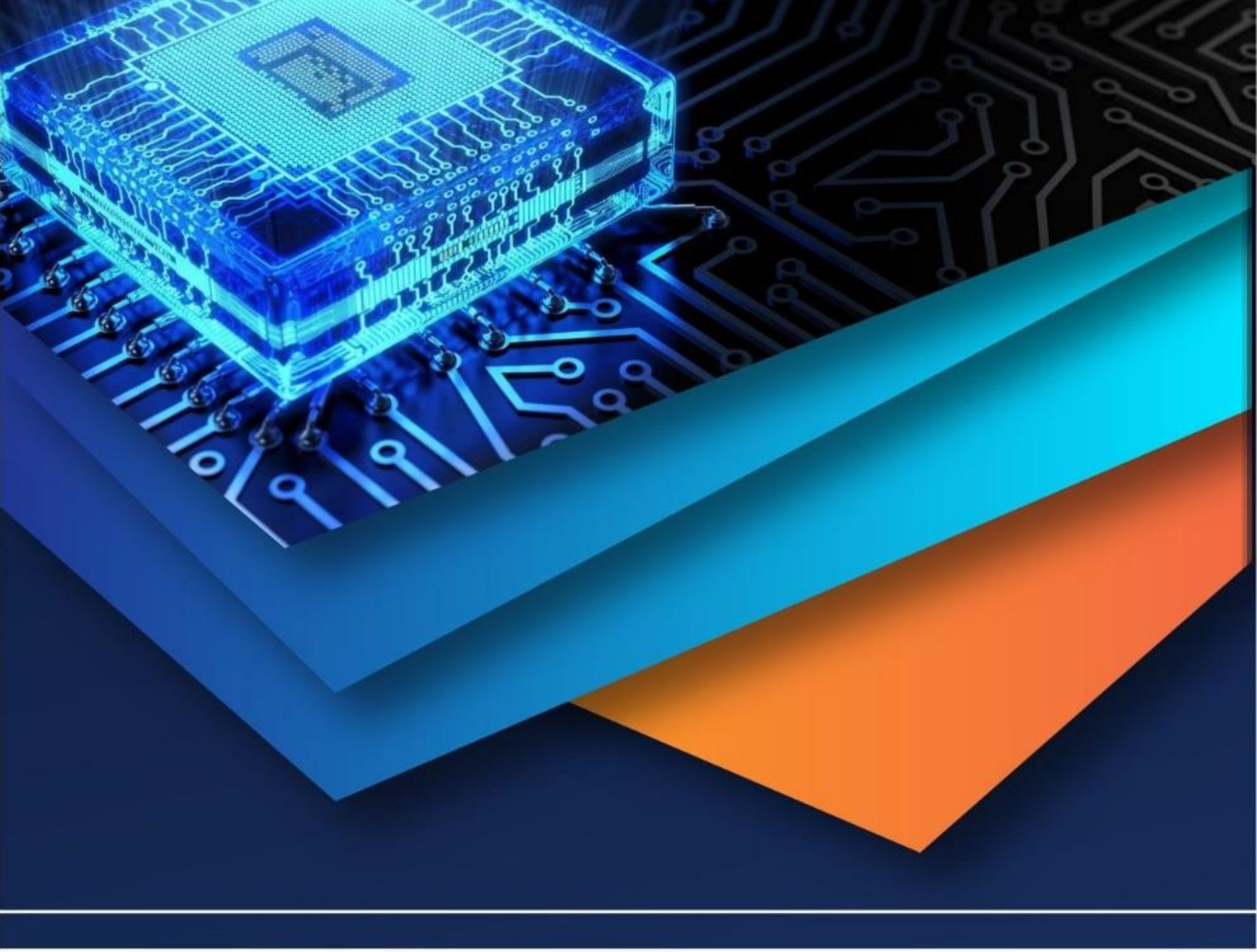
VI. CONCLUSIONS

With the increase in social networking sites and the use of social media by young people, cyberbullying has become more common and has started to cause problems in society. An automated cyberbullying system should be developed to avoid the negative effects of cyberbullying.

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, and remedies," 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] C. Fuchs, Social media: A critical introduction. Sage, 2017.
[13] N. Selwyn, "Social media in higher education," The Europa world oflearning, vol. 1, no. 3, pp. 1–10, 2012.
[14] H. Karjaluoto, P. Ulkuniemi, H. Keinanen, and O. Kuivalainen, "An-tecedents of social media b2b use in industrial marketing context: customers' view," Journal of Business & Industrial Marketing, 2015.
[15] W. Akram and R. Kumar, "A study on positive and negative effects of social media on society," International Journal of Computer Sciences and Engineering, vol. 5, no. 10, pp. 351–354, 2017.
[16] D. Tapscott et al., The digital economy. McGraw-Hill Education,, 2015.
[17] S. Bastiaensens, H. Vandebosch, K. Poels, K. Van Cleemput, A. Desmet, and I. De Bourdeaudhuij, "Cyberbullying on social network sites. An experimental study into bystanders' behavioural intentions to help the victim or reinforce the bully," Computers in Human Behavior, vol. 31, pp. 259–271, 2014.
[18] D. L. Hoff and S. N. Mitchell, "Cyberbullying: Causes, effects, and remedies," Journal of Educational Administration, 2009.
[19] S. Hinduja and J. W. Patchin, "Bullying, cyberbullying, and suicide,"Archives of suicide research, vol. 14, no. 3, pp. 206–221, 2010.
[20] D. Yin, Z. Xue, L. Hong, B. D. Davison, A. Kontostathis, and L. Ed-wards, "Detection of harassment on web 2.0," Proceedings of the Content Analysis in the WEB, vol. 2, pp. 1–7, 2009.
[21] K. Dinakar, R. Reichart, and H. Lieberman, "Modeling the detection of textual cyberbullying," in In Proceedings of the Social Mobile Web. Citeseer, 2011.
[22] K. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," in 2011 10th International Conference on Machine learning and applications and workshops, vol. 2. IEEE, 2011, pp. 241–244.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)