iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51749

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 V May 2023- Available at www.ijraset.com

Cyberbullying Detection using Machine Learning Algorithms

Bhagyashree Kadam[1], Rohan Borhade[2], Anisha Bardeskar[3], Vaishnavi Shelke[4], Pranjal Bhosale[5]
[1]*Assistant Professor,* [2, 3, 4, 5]*B.E Students, Department of Information Technology, Sinhgad Institute of Technology and Science, Pune, India*

Abstract: ***In today's digital age, the internet has become an integral part of our lives, resulting in an exponential increase in the volume of data generated. However, this surge in online activity has also led to the emergence of cyberbullying as a major concern in web 4.0. Cyberbullying refers to the use of technology to intimidate, harass or threaten an individual, and is considered a form of cybercrime. Given the lack of available datasets, anonymous identities of perpetrators and the privacy of victims, previous research in cyberbullying detection has been limited. To address this issue, a new approach based on text mining and machine learning algorithms is proposed to proactively detect bullying text. Unlike previous research, which only considered textual features, the current study extracts three types of features: textual, behavioural and demographic features. Textual features include specific words commonly used in cyberbullying, which may indicate the presence of bullying behaviour. Behavioural features are based on personality traits and are extracted to determine the likelihood of a user engaging in bullying behaviour in the future. Demographic features, such as age, gender and location, are also extracted from the dataset. Overall, this text mining approach using machine learning algorithms can effectively detect cyberbullying, providing a valuable tool to combat this growing concern in the cyber world.***

Keywords: Cyberbullying Detection, Decision Tree Algorithm, Support Vector Machine (SVM), Random Forest Algorithm, Multi Layer Perceptron.

I. INTRODUCTION

In the current digital era, cyberbullying has grown to be a significant problem. It can have a significant effect on people and even have tragic results like despair, anxiety, and even suicide. Social media platforms rapid expansion has made it simpler for cyberbullies to target their victims while remaining anonymous. Machine learning presents a viable solution to the urgent problem of detecting and stopping cyberbullying. In this study, we describe a machine-learning method for spotting online harassment on Twitter. Four classification algorithms Decision Tree, Random Forest, Support Vector Machine, and Multi Layer Perceptron were trained and assessed using a dataset of hate speech. The dataset was split into a 70/30 training/testing split, and the evaluation criteria employed were accuracy, precision, recall, and F1-score. The findings indicate that the Multi Layer Perceptron had the maximum recall while the Decision Tree method had the highest accuracy, precision, and F1-score.

The suggested methodology may help cyber-investigators and researchers find cases of cyberbullying in fresh tweets. Deep learning methods will be used in our upcoming work to extract pronunciation features and enhance the model's accuracy.

II. METHODOLOGIES

A. Decision Tree Algorithm

The decision tree classifier is a versatile tool that can be applied to both classification and regression problems. It serves a dual purpose by not only representing the decision-making process but also making decisions itself.

The decision tree is constructed as a tree structure, with each internal node representing a condition and each leaf node representing a final decision. In classification tasks, the decision tree returns the predicted class for a given target variable, while in regression tasks, it provides the predicted value for a given input.

B. Random Forest Algorithm

Ensemble learning method known as Random decision forests, is commonly used for classification, regression, and other tasks. This method involves constructing a large number of decision trees during training, which collectively form a "forest" of trees. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean prediction of the individual trees is returned [2].

©IJRASET: All Rights are Reserved

C. Support Vector Machine

Support Vector Machines (SVMs) were used to identify and classify the cyber bullying posts. Researcher achieved an average accuracy of 97.11% on the conversational from spring dataset with modifications using SVM with poly kernel. [3]

Mangaonkaret al presented a collective approach to classifying a tweet as ”bullying” or ”non-bullying” .Supervised machine learning algorithms such as SVM,NB, logistic regression, are utilized. .Hee. et al developed a model for detection of cyberbully content automatically from ASKfm - Dutch social media using linear SVM .Van Hee et al detected cyber bully content by evaluating the texts composed by stander, victim , attacker in online social media by using LSVM model which beats the keyword-based baseline and word n-gram techniques for detection. [4]

D. Multi Layer Perceptron

An MLP (multilayer perceptron) is an artificial neural network that operates in a feed forward manner, producing a set of outputs based on a given set of inputs. An MLP consists of at least three layers of the nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a chain rule [5].

E. Other Evaluation Formulas

1) *Precision:* The total number of correctly identified true bullying posts out of retrieved bullying posts.
   Precision = (relevant cyberbullying texts) (retrieved cyberbully texts) / (retrieved cyberbully texts).
2) *Recall: Recall* = Number of correctly identified bullying cases from total number of true bullying cases.
   Recall = (relevant cyberbully texts) (retrieved cyberbully texts) / (relevant cyberbully texts).
3) *F1-score:* F-measure = It is the harmonic mean of precision and recall.
   F-measure = 2 x [(Precision x Recall) / (Precision + Recall)].

III.RESULT

In order to detect Cyberbullying using Machine Learning, we implemented above given classification algorithms on hate speech dataset [1] and accuracy, precision, recall and F1-score were considered as evaluation parameters. Dataset was divided in training and testing data in proportion 7:3 and following results were obtained.

TABLE I
RESULTS OF CLASSIFICATION ALGORITHMS

Algorithms	Evaluation Parameters			
	Accuracy	Precision	Recall	F1-score
Decision Tree Algorithm	95.39 %	0.99	0.95	0.97
Random Forest Algorithm	96.01 %	0.99	0.96	0.96
Support Vector Machine	96.14 %	0.98	0.96	0.96
Multi Layer Perceptron	95.77 %	0.98	0.96	0.97

Here we can see that Support Vector Machine is giving maximum accuracy for given dataset followed by Random Forest, Multi Layer Perceptron and Decision Tree Algorithm. In terms of Precision, Decision Tree and Random Forest algorithms give highest value of 0.99. Random Forest Algorithm, Support Vector Machine and Multi Layer Perceptron gives best Recall value ie. 0.96 and F1-score is found to be best of Decision Tree Algorithm and Multi Layer Perceptron.

IV.SUMMARY

Cyberbullying Detection implements our coded, machine learning algorithms, in finding a negative comment from the messages it receives by a user. The algorithm first gives the message a value and then based on our pre trained data, it decides if the comment is harsh enough to be transformed or not. Algorithms are used for accuracy better than previous one. Decision tree, Random forest, SVM, Multilayer perceptron among all these algorithms decision tree gives highest accuracy.

.

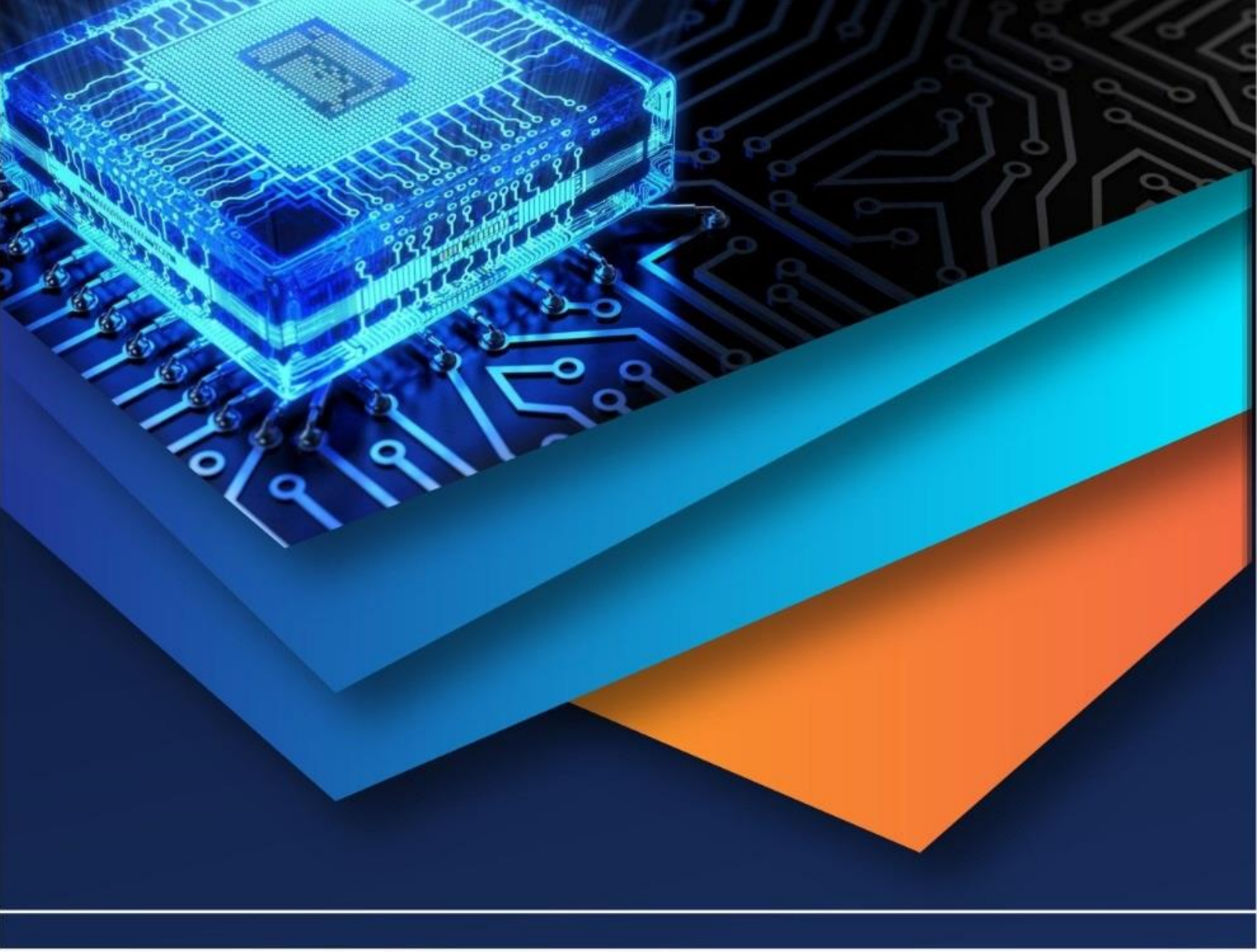
V. FUTURE SCOPE

The further plan of action is to use deep learning techniques on the extracted pronunciation features to detect cyber-bullying in Twitter. The developed model will be capable of detecting instances of cyber-bullying in new tweets. The process of multilingual data which is a technology that incorporates linguistics, computer science and AI would be used. In future there must be techniques like audio and image processing carried on for betterment.

REFERENCES

[1] Public Source. (2023). Cyberbully Classification [Data set]. Kaggle.https://doi.org/10.34740/KAGGLE/DSV/5493187
[2] Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016.
[3] Nikhila, M. S., Bhalla, A. Singh, P. (2020), 'Text imbalance handling and classification for cross platform cyber-crime detection using deep learning', 2020 11thInternational Conference on Computing, Communication and Networking Technologies (ICC-CNT) pp. 1–7.
[4] Singh, N. Sharma, S. K. (2021), Review of machine learning methods for identification of cyberbullying in social media, in '2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)', pp.284–288.
[5] Leibniz, Gottfried Wilhelm Freiherr von (1920). The Early Mathematical Manuscripts of Leibniz: Translated from the Latin Texts Published by Carl Immanuel Gerhardt with Critical and Historical Notes (Leibniz published the chain rule in a 1676 memoir). Open court publishing Company.

©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)