IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203080 |

Analysis of Cyberbullying Using Machine Learning

Shalini R J, Sowmya S, Vivitha V, Mr. S. Velmurgan, B.E., M.E.,

UG Students, Dept. of Information Technology, R.M.D. Engineering College, Tiruvallur, Tamilnadu, India

Assistant Professor, Dept. of Information Technology, R.M.D. Engineering College, Tiruvallur, Tamilnadu, India

ABSTRACT: In recent years, smart city services have moved the existence of people from the physical to the virtual world (cyberspace), e.g. online banking, e-commerce, telemedicine, etc. Online content is a crucial part of smart city and sustainable management of it is a primary task of current society. Along with the benefits of smart cities, the problems of the physical world are also shifted to the cyber world, like harassment in the covid 19 Pandemic situation. Especially for young people, cyber bullying even could lead them to do self-harm and suicide. Here, we use the machine learning algorithm namely Logistic Regression, Random Forest, Decision Tree, Support vector machine, K-Nearest Neighbor to categorize the threat in cyber bullying detection of the system. Experimental results show the better performance of the system.

KEYWORDS: Logistic Regression; Random Forest; Decision Tree; Support vector machine; K-Nearest Neighbor

I. INTRODUCTION

Cyberbullying, a pervasive issue in the digital age, continues to pose significant challenges to online communities, especially amongst adolescents and young adults. As the prevalence of cyberbullying grows, so does the urgency to develop effective tools for detection and mitigation. Machine learning (ML) offers promising avenues for combating cyberbullying by automating the identification of harmful online behavior. This analysis delves into the application of various ML algorithms—Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN)—to detect and analyze cyberbullying instances. Each algorithm offers unique strengths in handling the complexities of textual data, thereby providing valuable insights into identifying and addressing cyberbullying behavior. In this analysis, we evaluate the performance of these ML algorithms on datasets containing diverse instances of cyberbullying behavior. By comparing their accuracy, precision, recall, and F1 scores, we aim to discern the strengths and limitations of each algorithm in addressing the multifaceted challenges posed by cyberbullying.

II. RELATED WORK

There are several works on machine learning-based cyberbullying detection. A supervised machine learning algorithm was proposed using a bag-of-words approach to detect the sentiment and contextual features of a sentence [1]. This algorithm shows barely 61.9% of accuracy. Massachusetts Institute of Technology conducted a project called Ruminati [2] employing support vector machine to detect cyberbullying of youtube comments. The researcher combined detection with common sense reasoning by adding social parameters. The result of this project was improved to 66.7% accuracy for applying probabilistic modelling. Reynolds et al. [3] proposed a language-based cyberbullying detection method which shows 78.5% of accuracy. The authors used the decision tree and instance-based trainer to achieve this accuracy. To improve cyberbullying detection, the author of the paper [4] has used personalities, emotion and sentiment as the feature. Several deep learning-based models were also introduced to detect the cyberbullying. Deep Neural Network-based model is applied for cyberbullying detection by using real-world data [5]. The authors first analyze cyberbullying systematically then used transfer learning to do the detection task. Badjatiya et al. [6] has presented a

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203080 |

method using deep neural network architectures for detecting hate speech. A convolutional neural network-based model has been proposed to detect cyberbullying [7]. The authors employed word embedding where similar words have similar embedding. In a multimodal context, Cheng et al. [8] research the novel issue of cyberbullying identification by collaboratively exploiting social media data. This challenge, however, is difficult due to the complex combination of both cross-modal associations among multiple methods and structural correlations between various social media sessions, and the complex attribute information of different modalities. They propose XBully, a novel cyberbullying identification system to overcome these challenges, which first reformulates multi-modal social media data as a heterogeneous network and then tries to learn node embedding representations on it. Many literatures on cyberbullying have concentrated on text analysis over the past few decades. Cyberbullying, however, is becoming multi-objective, multi-channel, and multi-form. The variety of bullying data on social platforms can't be met by conventional text analytical techniques. Wang et al. [9] suggested a multi-modal identification system that integrates multi-modal information such as image, video, comments, time on social media to cope with the latest type of cyberbullying. In particular, they not only extract textual characteristics, but also apply hierarchical attention networks to capture the social network session function and encode various media information including video, image. The authors model the multi-modal cyberbullying detection system to address the latest type of cyberbullying on the basis of these characteristics.

III. PROPOSED SYSTEM

Early research on cyberbullying was primarily based on statistical analysis and investigation, which concentrated on the concept, statistical techniques, and effects of cyberbullying. These studies improved the factuality of cyberbullying and encouraged researchers to pay more attention to cyberbullying from the perspective of severity. Cyberbullying detection has been considered task comes around machine learning. Here we collect the datasets as in the CSV file and then it get into pre-processed. Then the dataset split into train as like of 80% and test as like of 20%. Then we train the model with the train data. Thus, we classify the cyber bullying threats by the machine learning algorithms and compared with each other for better accuracy. Then finally we can predict the cyber bullying threats with better accuracy metrics.

IV. BULLYING DETECTION MODEL

The cyberbullying detection framework which consists of two major parts. The first part is called NLP (Natural Language Processing) and the second part is named as ML (Machine learning). In the first phase, datasets containing bullying texts, messages or post are collected and prepared for the machine learning algorithms using natural language processing. The processed datasets are then used to train the machine learning algorithms for detecting any harassing or bullying message on social media including Facebook and Twitter.

A. Methodology

• Natural Language processing: The real-world posts or text contain various unnecessary characters or text. For example, numbers or punctuation are irrelevant to bullying detection. Before applying the machine learning algorithms to the comments, we need to clean and prepared them for the detection phase. In this phase, various processing task including removal of all irrelevant characters like stop-words, punctuation and numbers, tokenization, stemming etc. After the preprocessing, we prepare the two important features of the texts as follows:

1) Bag-of-Word: The machine learning algorithms cannot work directly with the raw text. So before applying the algorithms we must convert them to vectors or numbers. So, the processed data is converted to Bag-of-Words (BoW) for the next phase.

2) TF-IDF: This is another feature that we consider for our model. TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical measure that can evaluate how relevant a word is to a document in a collection of documents. In bag of words, every word is given equal importance while in TF-IDF the words that occur more frequently should be given more importance as they are more useful for classification.

• Machine Learning: This module involves in applying various machine learning approaches like Decision Tree (DT), Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbor to detect the bullying

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203080 |

message and text. The classifier with the highest accuracy is discovered for a particular public cyberbullying dataset. Next section, some common machine learning algorithms are discussed to detect cyberbullying from social media texts.

B. Machine Learning Algorithms

In this section, we discussed the basic mechanisms of several machine learning algorithms. We presented Decision Tree (DT), Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbor in each subsection.

1) Decision Tree: The decision tree classifier can be used in both classification and regression. It can help represent the decision as well as make a decision. The decision tree is a tree like structure where each internal node represents a condition, and each leaf node represents a decision. A classification tree returns the class where the target falls. A regression tree yields the predicted value for an addressed input.

2) Random Forest: Random Forest classifier is consisting of multiple decision tree classifiers. Each tree gives a class prediction individually. The maximum number of the predicted class is our final result. This classifier is a supervised learning model which provides accurate result because several decision trees are merged to make the outcome. Instead of relying on one decision tree, the random forest takes the prediction from each generated tree and based on the majority votes of predictions, and it decides the final output.

3) Support Vector Machine: Support Vector Machine (SVM) is a supervised machine learning algorithm which can be applied in both classification and regression alike a decision tree. It can distinguish the classes uniquely in n-dimensional space. Thus, SVM produces a more accurate result than other algorithms in less time. In practice, SVM constructs set a of hyperplanes in a infinite-dimensional space and SVM is implemented with kernel which transforms an input data space into the required form. For example, Linear Kernel uses the normal dot product of any two instances as follows:

$$K(x, xi) = sum(x * xi) \quad (1)$$

4) Logistic Regression: Logistic regression aims to solve classification problems. It does this by predicting categorical outcomes, unlike linear regression that predicts a continuous outcome. Logistic regression aims to solve classification problems. It is one of the simplest and commonly used Machine Learning algorithms for two-class classification. It is easy to implement and can be used as the baseline for any binary classification problem. Its basic fundamental concepts are also constructive in deep learning.

5)K-Nearest Neighbor: This algorithm is used to solve the classification model problems. K-nearest neighbor or K-NN algorithm basically creates an imaginary boundary to classify the data. When new data points come in, the algorithm will try to predict that to the nearest of the boundary line.

V. RESULTS

In this module, we finally can predict the correct cyber bullying threats by testing the testing data. Finally, the performance metrics are considered in terms of accuracy of the system. This dataset is built from the user comments on different posts. We compare the various parameters of the machine learning algorithms based on the two important features vectors BoW and TF-IDF. Figure 1 show the precision and accuracy results and from the graph it is clear that LR outperforms the other algorithm. The results also indicate that, TF-IDF provides better accuracy than BOW feature. This is because rather than taking almost all word into vectors, TF-IDF takes the most frequent words and maintain better performances.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203080 |

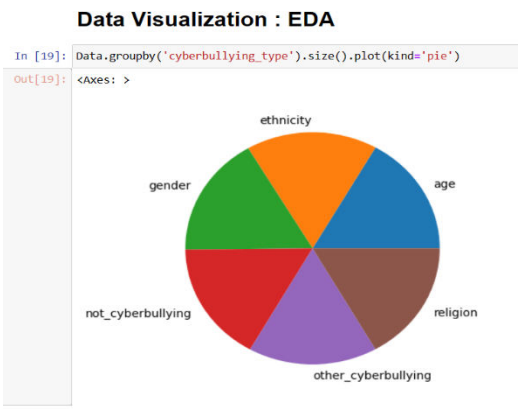

Fig 1. Data Visualization

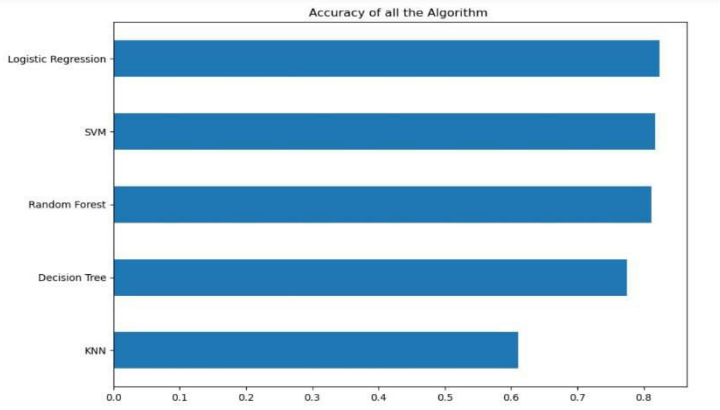

Fig 2. Accuracy of all the algorithm

VI. CONCLUSION AND FUTURE WORK

In this project, we proposed an approach to detect cyber bullying using machine learning techniques. We evaluated our model on a classifier machine algorithm for features extraction. We achieved better accuracy. We found that our algorithm performed better than the existing machine learning classifier as it also achieves average f-score at high. Future enforcement analysis of cyberbullying is a research area that aims to find effective ways to prevent, detect, and respond to cyberbullying incidents. Some of the topics that future enforcement analysis of cyberbullying may cover are: Developing advanced machine learning and natural language processing techniques to automatically identify and flag cyberbullying content on social media and other online platforms. Designing and implementing innovative solutions that involve education, awareness, intervention, and support for cyberbullying victims.

REFERENCES

[1] Peter K Smith, Jess Mahdavi, Manuel Carvalho, Sonja Fisher, Shanette Russell, and Neil Tippett. Cyberbullying: Its nature and impact in secondary school pupils. Journal of child psychology and psychiatry, 49(4):376–385, 2008.
[2] Kelly Reynolds, April Kontostathis, and Lynne Edwards. Using machine learning to detect cyberbullying. In 2011 10th International Conference on Machine learning and applications and workshops, volume 2, pages 241–244. IEEE, 2011.
[3] B Nandhini and JI Sheeba. Cyberbullying detection and classification using information retrieval algorithm. In Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015), page 20. ACM, 2015.
[4] B Sri Nandhini and JI Sheeba. Online social network bullying detection using intelligence techniques. Procedia Computer Science, 45:485–492, 2015. Conference on, pages 242– 247. IEEE, 2017.
[5] V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, “Detection of Cyberbullying Using Deep Neural Network,” 2019, doi: 10.1109/ICACCS.2019.8728378.
[6] K. Reynolds, A. Kontostathis, and L. Edwards, “Using machine learning to detect cyberbullying,” 2011, doi: 10.1109/ICMLA.2011.152.
[7] J. Yadav, D. Kumar, and D. Chauhan, “Cyberbullying Detection using Pre-Trained BERT Model,” 2020, doi: 10.1109/ICESC48915.2020.9155700.
[8] M. Dadvar and K. Eckert, “Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study,” arXiv. 2018.
[9] S. Agrawal and A. Awekar, “Deep learning for detecting cyberbullying across multiple social media platforms,” arXiv. 2018.
[10] Y. N. Silva, C. Rich, and D. Hall, “BullyBlocker: Towards the identification of cyberbullying in social networking sites,” 2016, doi: 10.1109/ASONAM.2016.7752420.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details