International Journal of Innovative Research in Engineering
Volume 4, Issue 3 (May-June 2023), PP: 284-289.
www.theijire.com ***ISSN No: 2582-8746***

Cyber bullying Detection on Social Media Using Machine Learning

Gowthami.S[1], Meneka.S[2], Nilani.K[3], Priyadharshini.G[4], Radha Bhuvaneshwari.S[5], Roshini.P[6]
[1,2]*Assistant Professor, Department Of Computer Science & Engineering, Vivekanandha College Of Technology For Women, Namakkal, Tamil Nadu, India.*
[3,4,5,6]*UG Scholar, Department Of Computer Science & Engineering, Vivekanandha College Of Technology For Women, Namakkal, Tamil Nadu, India.*

How to cite this paper:
Gowthami.S[1], Meneka.S[2], Nilani.K[3], Priyadharshini.G[4], Radha Bhuvaneshwari.S[5], Roshini.P[6], "Cyber bullying Detection on Social Media Using Machine Learning", IJIRE-V4I03-284-289.

https://www.doi.org/10.59256/ijire.2023040390

Abstract: *Cyber bullying has become a growing concern in today's society, with more and more people turning to the internet to harass and intimidate others. Digital forensics is an essential tool for investigating cyber bullying activities, as it allows for the collection and analysis of digital evidence. However, traditional digital forensics techniques can be time-consuming and require a significant amount of human effort. In this paper, we propose the use of machine learning algorithms to aid in the investigation of cyber bullying activities. By training these algorithms on a dataset of known cyber bullying incidents, we can create a predictive model that can automatically classify new instances of cyber bullying. This can significantly reduce the time and effort required for investigations, allowing for a more efficient response to cyber bullying incidents. The challenges associated with using machine learning for cyber bullying detection, including the need for high-quality training data and the potential for bias in the algorithms. We also explore the various types of digital evidence that can be used in cyber bullying investigations, such as social media posts, emails, and instant messages. We present a case study in which we apply our proposed approach to a real-world cyber bullying incident. Our results show that the machine learning algorithm was able to accurately identify the cyber bullying activity with a high level of precision, demonstrating the potential of this approach for improving the efficiency and effectiveness of cyber bullying investigations.*

I.INTRODUCTION

Digital forensics refers to the process of collecting, analysing, and preserving electronic data in order to investigate and prevent cybercrimes. One of the major cybercrimes that digital forensics can help investigate is cyber bullying, which involves the use of digital technologies to harass, intimidate, or embarrass individuals or groups.

In recent years, cyber bullying has become a growing concern, particularly among young people who spend a significant amount of time on social media platforms and other digital communication channels. To combat cyber bullying, digital forensics experts have started using machine learning techniques to identify patterns of behaviour and detect cyber bullying activities. Machine learning algorithms can analyse large volumes of data and identify suspicious patterns of behaviour that may indicate cyber bullying. For example, machine learning models can be trained to identify abusive language, threatening messages, and other indicators of cyber bullying in online communication channels. By leveraging machine learning algorithms, digital forensics experts can quickly and accurately identify cyber bullying activities, which can help law enforcement agencies and other stakeholders take appropriate action to prevent further harm. Ultimately, the use of machine learning in digital forensics can help create a safer and more secure online environment for everyone.

II.LITERATURE SURVEY

1) The study reviewed the existing literature for various machine learning algorithms and identified Light GBM as the most efficient. A model for detecting bullying tweets for real time tweets was developed.
2) The advent of social media, particularly Twitter, raises many issues due to a misunderstanding regarding the concept of freedom of speech. One of these issues is cyber bullying, which is a critical global issue that affects both individual victims and societies.
3) Cyber bullying and cyber aggression are increasingly worrisome phenomena affecting people across all demographics. More than half of young social media users worldwide have been exposed to such prolonged and/or coordinated digital harassment
4) In this paper, a holistic multi-dimensional feature set is developed which takes into account individual-based, social network-based, episode based and linguistic content-based cyber bullying features.
5) Social media networks like Face book and Twitter create a great platform to share public views, opinions, and feelings by text message, image, and video. The public is very much interested to use these networks because of the comfortable Graphical User Interface (GUI) by a single click and taps to share content from their electric gadgets, gizmos, and mostly by their smart phones.

6) This study aims to highlight previous researchers and to propose an approach to detect cyber bullying along with the element of sarcasm included in it. The results proved that SVM classifier performed better than other classifiers.
7) Prior to the innovation of information communication technologies (ICT), social interactions evolved within small cultural boundaries such as geo spatial locations. The recent developments of communication technologies have considerably transcended the temporal and spatial limitations of traditional communications.
8) In this article, we aim to explore different approaches that take into account the time in the detection of cyberbullying in social networks. We follow a supervised learning method with two different specific early detection models, named threshold and dual.
9) Cyber bullying has greatly affected people's daily lives. We conduct this study to detect whether online comments contain cyber bullying behaviors and classify cyber bullying to alleviate this problem. This paper uses an improved information gain algorithm for feature selection, and the bidirectional LSTM neural network is used for classification.
10) This research detects cyber bullying for the Malay language using supervised machine learning (ML) and Natural Language Processing (NLP). Due to the high number of cyber bullying cases in Malaysia over the years and the belief that there is an increased number of unreported cyber bullying cases, there needs an intelligent way to detect cyber bullying on social media.

III.EXISTING SYSTEM

Experts believe that every government should take this issue seriously and work to find a solution. In 2016, an incident known as the Blue Whale Challenge resulted in a large number of juvenile suicides in Russia and other countries. In recent years, individuals have expressed and shared their thoughts freely over the Internet. Yet, due to the characteristics of social media, it appears that harmful usage of social media is occurring. If we can create useful tools for detecting cyber bullying on social media, we can reduce cyber bullying. As a result, in this study, we offer a method for detecting cyber bullying based on social network analysis and data mining. The method will look at three basic strategies for discovering cyber bullying: keyword matching, opinion mining, and social network analysis.

Cyber bullying is a recurrent act of harassing, humiliating, threatening, or bothering someone using electronic devices and online social networking websites. Cyber bullying is more destructive than conventional bullying because it has the capacity to spread shame to an infinite online audience. According to UNICEF and an Indonesian Ministry of Communication and Information study, 58% of 435 teenagers are unaware about cyber bullying. Some of them may have even been bullies, but since they did not understand cyber bullying, they were unable to see the detrimental consequences of their actions. Bullies may fail to see the consequences of their acts because they may not witness quick reactions from their victims. Our study attempted to discover cyber bullying actors using texts and user credibility analysis and to inform them about the dangers of cyber bullying. We gathered information from Twitter. Because the data was unlabelled, we created a web-based labelling tool to categorise tweets as cyber bullying or non-cyber bullying. The programme provided us with 301 cyber bullying tweets, 399 non-cyber bullying tweets, 2,053 bad terms, and 129 curse words. Following that, we used SVM and KNN to learn about and recognise cyber bullying texts. SVM has the greatest f1-score (67%), according to the data. We also conducted a user credibility study and discovered 257 Normal Users, 45 Harmful Bullying Actors, 53 Bullying Actors, and 6 Potential Bullying Actors.

Disadvantage

- Machine learning algorithms are only as good as the data they are trained on. If the data used to train the algorithm is incomplete, biased, or otherwise flawed, the algorithm may produce inaccurate or unreliable results.
- The require significant computational resources and may not be feasible for smaller organizations or investigations. In addition, there may be technical challenges in collecting and analysing digital evidence, particularly if the evidence has been deleted or encrypted.

IV.PROPOSED SYSTEM

A proposed system for using digital forensics and machine learning to investigate cyber bullying activities. To investigate cyber bullying activities, digital forensics techniques can be applied to gather evidence from electronic devices such as smart phones, computers, and social media accounts. Machine learning algorithms can be used to analyse the collected data and identify patterns and trends in the behaviour of the cyber bully. Collecting data on cyber bullying activities from various sources, such as social media platforms, messaging apps, and email accounts. Cleaning and preparing the data for analysis, which may involve removing irrelevant or duplicate data, standardizing the data format, and converting the data into a machine-readable format. Identifying key features and patterns in the data, such as the frequency and nature of the cyber bullying messages, the identity of the sender, and the social network of the victim. Using the extracted features to train machine learning algorithms to detect cyber bullying activities and to identify potential cyber bullies. Testing the machine learning models on a separate dataset to evaluate their accuracy and performance. Interpreting the results of the analysis and making informed decisions based on the evidence gathered. This may involve identifying the sources of cyber bullying, gathering additional evidence, and taking appropriate actions to stop the cyber bullying. It is important to note that the proposed system would need to be designed with careful consideration of ethical and legal issues, such as privacy, consent, and due process. It is also important to involve human expertise and judgment throughout the investigation process to ensure that the results of the analysis are interpreted and applied in a fair and ethical manner.

Advantage

- The identify patterns and trends in cyber bullying activities that may not be immediately apparent to human investigators. This can help to identify cyber bullies who may be using multiple accounts or disguising their identities.
- The can be used to identify potential cyber bullying incidents before they occur, which can help to prevent or mitigate the impact of cyber bullying.
- It used to identify potential cyber bullying incidents before they occur, which can help to prevent or mitigate the impact of cyber bullying.

V.ALGORITHM USED

- ➢ Logistic Regression
- ➢ Random Forest
- ➢ ADABOOST
- ➢ Decision Tree

Logistic Regression

Cyber bullying is a growing concern in today's digital age, with a significant number of people being subjected to it. As a result, there is a need for effective methods to investigate and prevent cyber bullying activities. Digital forensics is one such method that involves the analysis of digital devices and data to gather evidence in a legal investigation. Machine learning, on the other hand, is a technique that enables computers to learn from data and improve their performance over time. Logistic regression is one such algorithm that is widely used in machine learning for classification tasks. Combining digital forensics and machine learning techniques can provide a powerful tool for investigating cyber bullying activities. This approach can help in identifying patterns and trends in the data that may not be apparent to human analysts. In this context, logistic regression can be used to classify cyber bullying-related activities based on various features such as the content of the messages, the sender's profile, and the frequency of the messages. By training the model on a large dataset of cyber bullying activities, the algorithm can learn to accurately identify and classify new instances of cyber bullying. The combination of digital forensics and machine learning can be a powerful approach to investigate and prevent cyber bullying activities, which can have severe psychological and emotional consequences for the victims.

Random Forest

Cyber bullying is a serious problem that has become more prevalent with the rise of digital communication technologies. Digital forensics is one approach that can be used to investigate cyber bullying activities, which involves the analysis of digital devices and data to gather evidence in a legal investigation. Machine learning is another tool that can be used to analyse large amounts of data and identify patterns and trends that may not be apparent to human analysts. One popular machine learning algorithm is the Random Forest algorithm, which is a decision tree-based approach that can be used for both classification and regression tasks. In the context of investigating cyber bullying activities, the Random Forest algorithm can be trained on a large dataset of cyber bullying-related activities, including features such as the content of the messages, the sender's profile, and the frequency of the messages. By training the model on this dataset, the algorithm can learn to accurately classify new instances of cyber bullying. The Random Forest algorithm is that it can handle high-dimensional datasets and can identify important features that contribute to the classification accuracy. This information can be useful for understanding the underlying factors that contribute to cyber bullying activities and can help inform prevention strategies. The Random Forest algorithm, can be a powerful approach to investigate and prevent cyber bullying activities. By identifying patterns and trends in the data, this approach can help law enforcement agencies and other stakeholders to take proactive measures to prevent cyber bullying and protect the well-being of victims.

Adaboost

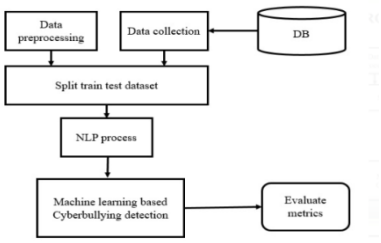
Cyber bullying is a significant problem that has become more prevalent with the widespread use of digital communication technologies. Digital forensics is one approach that can be used to investigate cyber bullying activities, which involves the analysis of digital devices and data to gather evidence in a legal investigation. Machine learning is another tool that can be used to analyse large amounts of data and identify patterns and trends that may not be apparent to human analysts. One popular machine learning algorithm is AdaBoost, which is an ensemble method that combines multiple weak classifiers to create a stronger classifier. In the context of investigating cyber bullying activities, AdaBoost can be trained on a large dataset of cyber bullying-related activities, including features such as the content of the messages, the sender's profile, and the frequency of the messages. By training the model on this dataset, the algorithm can learn to accurately classify new instances of cyber bullying. AdaBoost is that it can improve the classification accuracy by combining multiple weak classifiers. This can be useful for identifying complex patterns and trends in the data that may be difficult to identify with a single classifier. Where the number of positive instances (cyber bullying activities) is much smaller than the number of negative instances (non-cyber bullying activities). This is important because cyber bullying activities are often rare events, making it difficult to collect a large dataset of positive instances. The combining digital forensics and machine learning techniques, such as AdaBoost, can be a powerful approach to investigate and prevent cyber bullying activities. By identifying patterns and trends in the data, this approach can help law enforcement agencies and other stakeholders to take proactive measures to prevent cyber bullying and protect the well-being of victims.

Decision Tree

Cyber bullying is a growing problem in today's digital age, with a significant number of people being subjected to it. Digital forensics is one approach that can be used to investigate cyber bullying activities, which involves the analysis of digital devices and data to gather evidence in a legal investigation. Machine learning is another approach that can be used to analyse large amounts of data and identify patterns and trends that may not be apparent to human analysts. One popular machine learning algorithm is the Decision Tree algorithm, which is a tree-based approach that can be used for both classification and regression tasks.

In the context of investigating cyber bullying activities, the Decision Tree algorithm can be trained on a large dataset of cyber bullying-related activities, including features such as the content of the messages, the sender's profile, and the frequency of the messages. By training the model on this dataset, the algorithm can learn to accurately classify new instances of cyber bullying. One advantage of using the Decision Tree algorithm is that it can handle both categorical and numerical data, making it useful for analysing a wide range of features that may be relevant to cyber bullying activities. Additionally, decision trees can be visualized, which can help investigators to understand how the algorithm is making its classification decisions. The digital forensics and machine learning techniques, such as the Decision Tree algorithm, can be a powerful approach to investigate and prevent cyber bullying activities. By identifying patterns and trends in the data, this approach can help law enforcement agencies and other stakeholders to take proactive measures to prevent cyber bullying and protect the well-being of victims.

Architecture Diagram



VI.MODULE LIST

- Data Collection
- Data Pre-processing
- NLP processing
- Model Selection

Module Description

Data Collection

- Data for our project is collected form the website called Kaggle.
- The dataset contains 16858 records.

Data Preprocessing

- The dataset contains only text data which contains some special cases.
- We need to remove those special case.

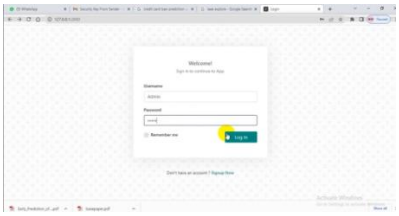
NLP Processing

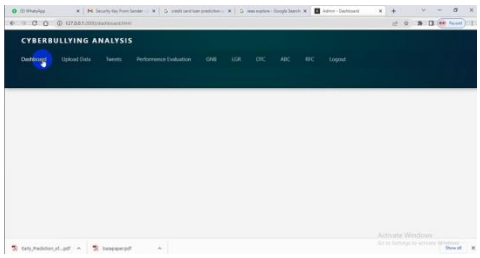
The text data not usable to build the machine learning model. Hence, we need to convert into the vector format. In this process, the NLP toolkit is used to perform the vectorization operations.

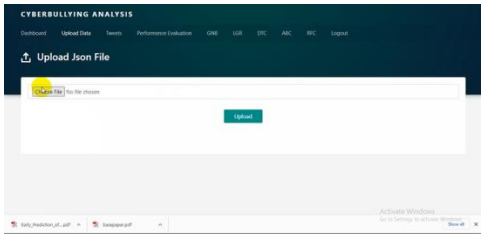
Model Selection

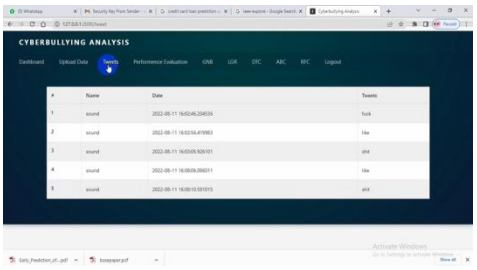
- ADABoost and Random Forest based model is compared in the application in order to build the efficient machine learning models
- Based on the better performance with cross validation, the model will be selected.

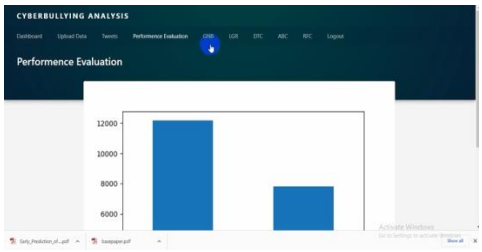
Result

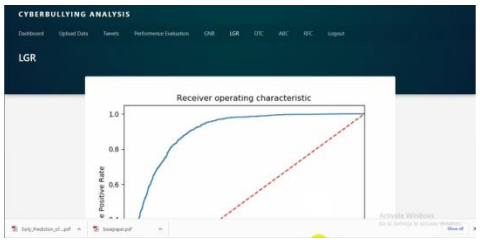












VII.CONCLUSION

In conclusion, the use of digital forensics and machine learning techniques can be very effective in investigating cyber bullying activities. By collecting electronic data from devices and social media accounts, digital forensics experts can identify evidence of cyber bullying and extract relevant features for training a machine learning model. Machine learning algorithms can then be used to analyse the data and identify patterns and trends in the cyber bully's behaviour. This approach can be particularly effective in cases where the cyber bully is using anonymous accounts or trying to hide their identity. The proposed system for investigating cyber bullying activities using digital forensics and machine learning has the potential to greatly improve the ability of law enforcement and school officials to address cyber bullying and protect victims. It is important to continue to develop and refine these techniques to stay ahead of cyber bullies and keep up with the constantly evolving digital landscape.

Reference

1) *A &. F. S. M. Muneer, "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter", Future Internet, vol. 12, no. 11, 2020.*
2) *A. Agarwal, "Information technology vis-a-vis human rights: an analytical and legal approach", Int'l JL Mgmt. & Human, vol. 5, no. 2, 2022.*

3) *C. Van Hee, G. Jacobs, C. Emmery, B. Desmet, E. Lefever, B. Verhoeven, et al., "Automatic detection of cyberbullying in social media text", PLoS ONE, vol. 13, no. 10, 2018.*
4) *H. Ahmad Ghazali, A. Abu Samah, S. Z. Omar, H. Abdullah, A. Ahmad and H. A. Mohamed Shaffril, "Predictors of Cyberbullying among Malaysian Youth", Journal of Cognitive Sciences and Human Development, vol. 6, no. 1, pp. 67-80, 2020.*
5) *H. Margono, "Analysis of the Indonesian Cyberbullying through Data Mining: The Effective Identification of Cyberbullying through Characteristics of Messages", Dissertation, 2019.*
6) *H. Rosa, N. Pereira, R. Ribeiro, P. C. Ferreira, J. P. Carvalho, S. Oliveira, et al., "Automatic cyberbullying detection: A systematic review", Computers in Human Behavior, vol. 93, pp. 333-345, 2019.*
7) *John Hani Mounir, Mohamed Nashaat, Mostafa Ahmed, Zeyad Emad, Eslam Amer and Ammar Mohammed, "Social Media Cyberbullying Detection using Machine Learning", (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 10, pp. 703-707, 2019.*
8) *O. Habimana, Y. Li, R. Li, X. Gu and G. Yu, "Sentiment analysis using deep learning approaches: an overview", Science China Information Sciences, vol. 63, no. 1, pp. 1-36, 2020.*
9) *T. K. Chan, C. M. Cheung and Z. W. Lee, "Cyberbullying on social networking sites: A literature review and future research directions", Information & Management, vol. 58, no. 2, 2021.*
10) *V. Ashok, "Nexus of advanced technology platforms for strengthening cyber-defense capabilities" in Practical applications of advanced technologies for enhancing security and defense capabilities: Perspectives and Challenges for the Western Balkans, IOS Press, pp. 14-31, 2022.*
11) *A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in Twitter data," 2015, doi: 10.1109/EIT.2015.7293405.*
12) *A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," Artif. Intell. Rev., vol. 53, no. 6, 2020, doi: 10.1007/s10462-019-09794-5.*
13) *E. Wulczyn, N. Thain, and L. Dixon, "Ex machina: Personal attacks seen at scale," 2017, doi: 10.1145/3038912.3052591.*
14) *I. H. Ting, W. S. Liou, D. Liberona, S. L. Wang, and G. M. T. Bermudez, "Towards the detection of cyberbullying based on social network mining techniques," in Proceedings of 4th International Conference on Behavioral, Economic, and SocioCultural Computing, BESC 2017, 2017, vol. 2018-January, doi: 10.1109/BESC.2017.8256403.*
15) *J. Yadav, D. Kumar, and D. Chauhan, "Cyberbullying Detection using Pre-Trained BERT Model," 2020, doi: 10.1109/ICESC48915.2020.9155700.*
16) *Z. Reynolds, A. Kontostathis, and L. Edwards, "Using machine learning to detect cyberbullying," 2011, doi: 10.1109/ICMLA.2011.152.*
17) *M. Dadvar and K. Eckert, "Cyberbullying Detection in Social Networks Using Deep Learning Based Models; A Reproducibility Study," arXiv. 2018.*
18) *P. Galán-García, J. G. de la Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, "Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying," 2014, doi: 10.1007/978-3-319-01854-6_43.*
19) *R. Zhao, A. Zhou, and K. Mao, "Automatic detection of cyberbullying on social networks based on bullying features," 2016, doi: 10.1145/2833312.2849567.*
20) *S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," arXiv. 2018.*
21) *T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," 2017.*
22) *T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013.*
23) *V. Banerjee, J. Telavane, P. Gaikwad, and P. Vartak, "Detection of Cyberbullying Using Deep Neural Network," 2019, doi: 10.1109/ICACCS.2019.8728378.*
24) *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.*
25) *Z. Waseem and D. Hovy, "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," 2016, doi: 10.18653/v1/n16-2013.*