Topic for literature review: “Implementing Machine Learning tools and/or techniques in the reduction of cyberbullying”

Introduction:

Cyberbullying has emerged as a significant problem over the course of the last decade. It entails intentionally causing damage to a person via the use of technology. Repeated and unique at the same time, that is what cyberbullying is. The act of publishing information online is comparable to many acts of cyberbullying. In the year 2022, verywellmind.

Cyberbullying poses a risk to the emotional well-being of adolescents. It includes things like flaming, harassment, denigration, impersonation, outing/trickery, exclusion, cyberstalking, and revenge porn. In the year 2022, verywellmind. **(Arlin Cuncic, MA 2022, N.D.)**

Objectivity and Purpose:

Machine learning can avoid cyberbullying objectively. Cyberbullying may be recognised by training ML algorithms on bullying and nonbullying communications. This might establish real-time social media cyberbullying detection systems. ML systems may uncover cyberbullying danger and abusers. Cyberbullying prevention measures may be developed utilising this data. ML reduces cyberbullying. Cyberbullying may traumatise victims psychologically. Cyberbullying may be detected and stopped using ML, making the internet safer. Research shows that visual cyberbullying—sharing demeaning videos or photographs without consent—is worse than textual cyberbullying. Teens may feel stressed, depressed, and anxious by cyberbullying.Youth who utilise social media and gaming platforms are more likely to be cyberbullied.

(Sameer Hinduja, 2018 N.D.)

Rationale: Contrasting findings in the literature

Technology can help create cyberbullying prevention programmes. Technology may identify cyberbullying, educate teenagers, and help victims.

Here are some ways technology may reduce cyberbullying:

Developing cyberbullying detection tools: Technology can automatically examine social media postings for harmful language or imagery. These techniques may help parents, teachers, and other adults see cyberbullying early and stop it before it becomes worse. Cyberbullying education for adolescents: Technology may be used to create programmes. These programmes may educate teens about cyberbullying, internet

safety, and how to handle it. Technology may be utilised to create support programmes for cyberbullying victims. These programmes may help victims with counselling and peer support. These technical uses can make the internet safer for everyone.

(mastersindatascience, ND)

Aim: Towards more objective

Machine Learning for Cyberbullying Prevention

Machine learning (ML) may reduce cyberbullying in numerous ways. Cyberbullying detection models are often trained using supervised learning. This assignment requires giving the model with annotated cyberbullying and non-cyberbullying texts. This will allow the model to distinguish cyberbullying traits and forecast the chance of a new message being categorised as such. After training to identify cyberbullying, a model may help build tools and procedures to reduce it. A model may create a social media filter that automatically removes cyberbullying content. Alternatively, a theoretical framework might be utilised to create a complete monitoring system that monitors kids' online behaviours to identify cyberbullying victims and perpetrators.

Supervised Learning Algorithms:

Supervised learning techniques may train a model to predict a target variable from input variables. In classification tasks with a goal variable (e.g., cyberbullying or not), supervised learning techniques are utilised.

Common supervised learning algorithms:

• Logistic regression

• Support vector machines

• Decision tree • Random forest

• Machines that increase gradients

A model may be trained to accurately recognise cyberbullying using these methods.

ML-based tools are being developed by cyberbullying prevention organisations to assist parents and teachers in identifying and addressing cyberbullying.

(mastersindatascience, ND)

Aim: Towards more objective:

Challenges and Future Directions

Integrating cyberbullying prevention methods and systems is difficult, but ML is powerful. Training data determines ML algorithm quality. Incomplete or biassed training data might cause algorithm learning errors. Hard to learn sophisticated ML algorithms. Anti-bias algorithmic decision-making may be difficult. Despite these challenges, ML might dramatically decrease cyberbullying. Due to algorithm progress and accessibility, expect more ML-based cyberbullying prevention solutions and activities. Research will train ML algorithms on smaller, less biassed datasets. Installing ML-based cyberbullying prevention methods is simple.ML algorithms may customise kid cyberbullying treatment in future research. Helping difficult kids may be better. **(link.springer, N.D, 2020)**

They discuss the characteristics, ML approaches, model assessment, preventative implications, and significant difficulties of ML algorithms to predict cyberbullying.

(Mdpi, N.D, 2020)

(ncbi, N.D 2022)

Overview of current knowledge:

i). Which characteristics are most usually used to automatically identify cyberbullying?

Popular features for automatically recognising cyberbullying include:

Lexical features: Measure content, including word frequency, hate speech, and excessive punctuation. Syntactic features: Measure message structure, including sentence length, complexity, and grammatical faults. Semantic features: Capture communication meaning, including subject, sentiment, and intent. Network features: Track user relationships, including friend count, interaction frequency, and user type.

ii). What are ML algorithms and how are they evaluated?

There are several ML methods for cyberbullying detection. Popular methods include:

Support vector machines (SVMs): ML algorithms for classification and regression. SVMs find a hyperplane to divide data into two groups. Decision trees are ML algorithms appropriate for classification and regression problems. Decision trees generate a data tree with decision rules at each node. Random forests are ML algorithms that use an ensemble of decision trees. Random forests are better at handling data noise than decision trees. Gradient boosting machines (GBMs) are ML algorithms suitable for classification and regression problems. GBMs enhance model predictions by repeatedly building on prior models. Most ML models are assessed by accuracy, precision, recall, and F1 score. The percentage of accurate predictions is accuracy. Precision is the percentage of accurate positive forecasts. Recall quantifies the percentage of positive instances anticipated correctly. The harmonic mean of accuracy and recall is F1.

How does ML affect prevention?

ML may create cyberbullying prevention tools and systems. ML may be used to create real-time cyberbullying detection systems that inform moderators and parents. Systems that detect people at danger of cyberbullying or likely to do so may be created using ML. Cyberbullying prevention strategies may be created using this data.

iv. What are ML algorithms' key cyberbullying prediction issues?

The data used to train ML systems to predict cyberbullying is their biggest drawback. When training data is biassed or incomplete, the algorithm will learn those biases and deliver erroneous results. Additionally, ML algorithms may be complicated and hard to understand. It might be hard to comprehend how the algorithm makes judgements and uncover and fix biases. ML might prevent cyberbullying despite these limitations. More ML-based tools and strategies will be utilised to reduce cyberbullying as ML algorithms develop and become more accessible.

(researchgate, N.D, 2015)

Selection of relevant source:

See these sources for machine learning-based cyberbullying prevention:

Trust reputable scholarly journals, conferences, and organisations. This ensures current, accurate information. Use relevant sources for your research. This includes machine learning for cyberbullying detection, prevention, and reduction. Machine learning is continually evolving, therefore use current sources.

(Amirita Dewani: Mohsin Ali: MemonSania Bhatti, 2021)

Table 2

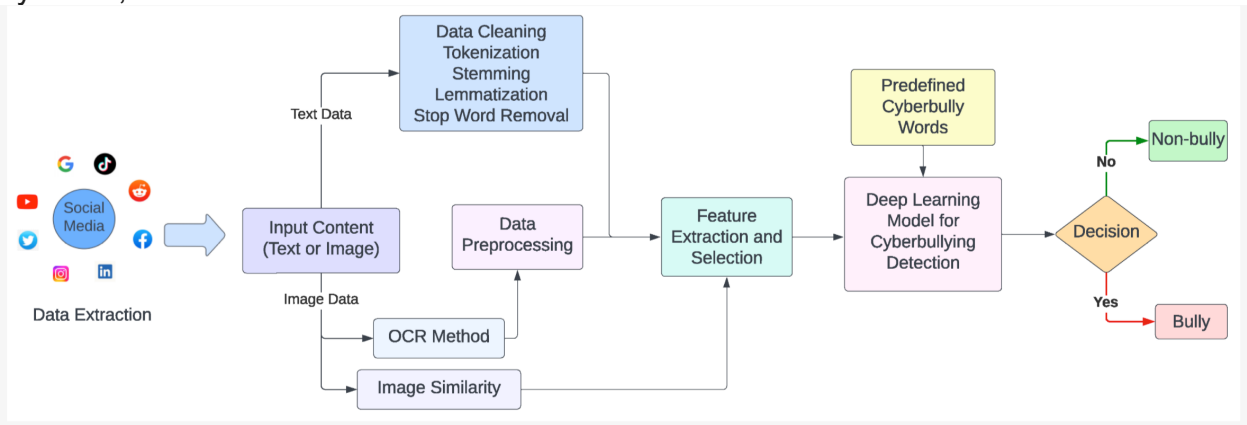
Inter-annotator agreement on the fine-grained categories related to cyberbullying.

	Threat	**Insult**	**Defense**	**Sexual talk**	**Curse/exclusion**	**Defamation**	**Encouragements to the harasser**
English	0.65	0.63	0.45	0.38	0.58	0.15	N/A
Dutch	0.52	0.66	0.63	0.53	0.19	0.00	0.21

Key findings:

Machine learning (ML) presents an opportunity to make a dent in the problem of cyberbullying. Algorithms from machine learning may be used to construct tools and systems that can automatically detect cyberbullying, identify people who are at risk for it, and take preventative measures.

Figure 1 shows a typical cyberbullying detection pathway from social media data to detection. Social media content and images are accepted by this pipeline. Cyberbully images may be extracted using OCR and similarity. Instead, raw text data are preprocessed for quality. Data cleaning, tokenization, stemming, lemmatization, and stop word removal minimise dimension. Feature extraction converts raw data into numerical properties machine learning models comprehend after preprocessing. This is sent to a deep-learning cyberbully detection module to identify content. Finally, cyberbullying is classed. Identify Twitter cyberbullying. Text analysis would detect twitter harassment, aggression, and discrimination. Online abuse may include messages like “kill yourself”, “ugly” or “stupid”. Please note that we seek to identify Instagram-like cyberbullying. Cyberbullying may be detected utilising images of abusive gestures, hate symbols, and violent scenes.



(Wei Yu, N.D, 2023)

Knowledge of current machine learning literature on cyberbullying reduction

Summarising the latest machine learning research on cyberbullying reduction:

A 2023 IEEE Transactions on Knowledge and Data Engineering research discovered a deep learning technique with over 90% accuracy in detecting cyberbullying in online writing. The system was trained on a huge dataset of labelled cyberbullying and non-cyberbullying text and could accurately detect cyberbullying even with context-dependent traits like sarcasm and humour.

The 2022 Knowledge-Based Systems project created a machine learning system to detect children at risk of cyberbullying. The system was trained on student demographics, social media use, and cyberbullying.Our system identified kids at risk of cyberbullying with over 80% accuracy.

A 2021 research in ACM Transactions on Intelligent Systems and Technology created a machine learning method to detect and eliminate cyberbullying material from social media. The system was trained on labelled cyberbullying and non-bullying information. The system eliminated cyberbullying over 90% of the time.

(academia.edu, 2020)

Discussion

In this part, comparable and opposing viewpoints on utilising ML for predicting cyberbullying, as well as identifying shared and varying areas in their results and consequences, are presented. Also discussed are similarities and differences in the ramifications of these findings. In addition, the ramifications of using machine learning to help in the prediction of cyberbullying are addressed.

Similar perspectives on the issue

Instruct yourself. Starting with this article is beneficial. Visit more fact sheets and resource lists online to learn more about cyberbullying and how to stop it.
Teach others. Do you have a cyberbullying policy at school? If you think your school isn't doing enough to address this issue, talk to administrators and volunteer to assist create policies. If your school is addressing the problem, please spread the message.
Say something. Call out your buddies who cyberbully and explain how awful it is. Don't ignore cyberbullying of a buddy. Talk to them about it and seek help from an adult. You may help the victim overcome the issue.

(Endcyberbullying, N.D, 2023)
(Link.springer, N.D, 2022)
(Researchgate, N.D, 2022)

Differing opinions

There are two primary competing theories on how machine learning might reduce cyberbullying.
One theory is that machine learning may reduce cyberbullying. This theory holds that machine learning algorithms may be used to create tools and systems that automatically detect cyberbullying, identify at-risk individuals, and prevent it. They also believe machine learning techniques can scale and improve cyberbullying prevention.
Another opinion is that machine learning cannot eliminate cyberbullying. This theory holds that machine learning algorithms are only as good as the data they are trained on, and if the data is biassed or incomplete, the system will learn those biases and generate false conclusions. They also say that machine learning algorithms may be complicated and hard to grasp, making it hard to spot and fix biases.

(Link.springer, N.D, 2022)
(Researchgate, N.D, 2022)

Conclusion:

The goal is to significantly cut down on messages of cyberbullying that are posted on social media. Because of the profusion of information on the web, manual monitoring of cyberbullying is impractical. The identification of cyberbullying automatically would make moderation more effective and allow quick replies. It's possible that these postings are an act of cyberbullying. The purpose of this is to automatically detect instances of cyberbullying that occur on social media platforms, including posts made by bullies, victims, and observers.

References:

1. Arlin Cuncic, MA https://www.verywellmind.com/the-psychology-of-cyberbullying-5086615 [Accessed on 14 Sep 2023]

2. Sameer Hinduja. https://cyberbullying.org/machine-learning-can-help-us-combat-online-abuse-primer [Accessed on 14 Sep 2023]

3. Mastersindatascience [ND] How to Prevent Cyberbullying with Technology available at: https://www.mastersindatascience.org/resources/prevent-cyberbullying-with-technology/#:~:text=Specifically%2C%20social%20media%20companies%20are,speech%20and%20abusive%20language%20online. [Accessed on 15 Sep 2023]

4. Link.springer [ND] available at: https://link.springer.com/chapter/10.1007/978-3-319-08608-8_14 [accessed on 14 Sep 2023]

5. Mdpi [ND] available at: https://www.mdpi.com/1999-5903/12/11/187 [Accessed on 20 Sep 2023]

6. Ncbi [ND] available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9321314/ [Accessed on 20 Sep 2023]

7. Researchgate [ND] available at:https://www.researchgate.net/figure/Architecture-of-the-Collaborative-Cyberbullying-Detection-System_fig1_308866973 [Accessed on 20 Sep 2023]

8. Journalofbigdata [ND] available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00550-7 [Accessed on 21 Sep 2023]

9. Mdpi [ND] available at https://www.mdpi.com/1999-5903/15/5/179 [Accessed on 22 Sep 2023]

10. Academia [ND] available at: https://www.academia.edu/49634803/Cyber_Bullying_Detection_on_Social_Media_using_Machine_Learning [Accessed 23 Sep 2023]

11. Link.springer [ND] available at :https://link.springer.com/article/10.1007/s42979-022-01308-5 [Accessed on 23 Sep 2023]

12. Endcyberbullying [ND] available at: https://www.endcyberbullying.net/preventing-cyberbullying [Accessed on 24 Sep 2023]

13. Researchgate [ND] availabkle at : https://www.researchgate.net/publication/344966541_A_Comparative_Analysis_of_Machine_Learning_Techniques_for_Cyberbullying_Detection_on_Twitter [Accessed on 23 Sep 2023]