General Information

Instagram has a feature called "Comment Controls" that allows users to block comments that contain specific keywords or phrases that are often associated with bullying or harassment. If a user attempts to post a comment containing those words or phrases, it will be blocked in real-time and will not be visible to other users.

TikTok also uses real-time detection to prevent bullying content from being uploaded to the platform. If a user tries to upload a video that contains bullying behavior or language, the system will flag it for review and may prevent it from being published.

**Problems of the CYBER BULLYING FIELD**

The field of cyberbullying detection faces several challenges and problems, including:

1. Defining and identifying cyberbullying: Cyberbullying can take many forms and can be difficult to define and identify. Different cultures and communities may have different definitions and perceptions of what constitutes cyberbullying, making it challenging to develop a universal approach to detection.
2. Language use: Cyberbullying often involves the use of language that is abusive, harassing, or threatening. However, the language used to perpetrate cyberbullying can be complex and subtle, making it difficult for algorithms to detect.
3. Contextual understanding: Cyberbullying often occurs within a specific context, such as on social media or in online gaming communities. Understanding the nuances of these contexts is important for accurately detecting cyberbullying behavior.
4. False positives and false negatives: Machine learning algorithms used to detect cyberbullying can produce false positives, where content is flagged as cyberbullying when it is not, or false negatives, where cyberbullying behavior is missed. Balancing these errors is a challenge in developing effective cyberbullying detection systems.
5. Ethical considerations: The use of machine learning algorithms to detect cyberbullying raises ethical considerations, such as user privacy and data protection. Additionally, algorithms may unintentionally perpetuate biases or discriminate against certain groups, which must be carefully addressed in the development of cyberbullying detection systems.
6. Rapidly evolving technologies: social media and other online platforms are constantly evolving, with new features and behaviors emerging regularly. Keeping up with these changes and updating detection algorithms accordingly is a significant challenge for the field of cyberbullying detection.
7. User engagement: Users may be hesitant to report instances of cyberbullying, either out of fear of retaliation or because they do not believe the platform will take effective action. Ensuring user engagement and trust in the cyberbullying detection system is therefore critical for its success.

**Underexplored Areas in the CYBERBULLYING DETECTION FIELD**

There are several areas that are currently under-explored or have limited research in the context of cyberbullying detection. Some potential ideas for your dissertation could be:

1. Privacy-preserving cyberbullying detection: Currently, most cyberbullying detection methods require access to user data, which raises privacy concerns. Developing privacy-preserving methods that can detect cyberbullying without compromising user privacy could be a valuable contribution to the field.
2. Multimodal cyberbullying detection: Most existing cyberbullying detection methods focus on text-based content, such as messages or comments. However, cyberbullying can also occur through images, videos, and audio content. Developing multimodal cyberbullying detection methods that can analyze multiple types of content could be a useful addition to the field.
3. Context-aware cyberbullying detection: Cyberbullying can vary depending on the context in which it occurs. For example, the same message could be considered cyberbullying in one context but not in another. Developing context-aware cyberbullying detection methods that can take into account the different social, cultural, and linguistic contexts in which cyberbullying occurs could be an important area of research.
4. Real-time cyberbullying detection: Most cyberbullying detection methods rely on analyzing content after it has been posted or sent. Developing real-time cyberbullying detection methods that can identify and prevent cyberbullying as it is happening could be a valuable contribution to the field.
5. Cyberbullying detection in non-English languages: Most cyberbullying detection methods have been developed and tested on English-language content. Developing cyberbullying detection methods that can analyze content in non-English languages could be an important area of research, especially given the global nature of social media platforms.

Paper 2 – Objective

There is numerous remarkable research conducted to overcome cyberbullying specifically on detection of cyberbullying with the help of machine learning to study the sentiment and contextual features from the conversation or medium. This project will tackle this problem by performing a classification on social media data to detect the act of cyberbullying. The objectives of this project are as follows: First, to classify posting susceptibility to cyberbullying with certain degree of confidence; Second, to investigate the context of discrimination on posting susceptible to cyberbullying; Finally, to study the severity of cyberbullying based on the context of cyberbullying.

1. Resampling on unbalanced dataset. (Python library “iambalancedlearn”)

2. Decision Tree, Random Forest, SVM.

3. Classifying a cyberbullying post as negative (as no cyberbullying) is more costly that classifying a non-cyberbullying post as positive (as cyberbullying).

4. SVM Very good results.

5. TF-IDF used also.

Paper 4

Issues Regarding Automatic Cyberbullying detection

1. Capturing the complexity of the phenomenon requires well-defined criteria to develop appropriate digital tools which integrate automatic detection features.

Frequently Used dataset for cyber bullying detection training: Formspring

There are 22 research projects in this systematic review. The paper includes their datasets.

Synthetic Oversampling / Under sampling to balance the datasets.

One attempt to capture the repetitiveness of aggression by detecting cyberbullying in sessions consisting of streams with several messages.