

**Paper Title:** An Ensemble-Based Multi-Classification Machine Learning Classifiers Approach to Detect Multiple Classes of Cyberbullying

**Paper Link:** <https://www.mdpi.com/2504-4990/6/1/9>

**Summary:**

1. **Motivation:** With the expansion of communication through social media cyberbullying has become one of the most growing issues. With the surge in online activity and data volume, harmful content like offensive tweets and bullying instances has also increased. Cyberbullying can be done in different contexts which makes it difficult to manually categorize it. Thus, in this paper, they combine multi-classification models to detect cyberbullying
2. **Contribution:** This study dives into the potential for constructing a framework to identify multiple categories of cyberbullying. Through their experimentation, the researchers devised an automated system capable of detecting cyberbullying tweets, outperforming conventional methods. They aimed to make a contribution by developing a system using an ensemble of transformer models, proposing a new framework in this field. This investigation adds value by introducing this automated system tailored to recognize diverse forms of cyberbullying specifically on the Twitter platform.
3. **Methodology:** They used the NLTK library for data pre-processing and incorporated the N-gram feature with TF-IDF for feature extraction. Their framework employs an ensemble method to achieve performance results from three multi-classification classifiers combined into two different types of ensemble techniques: voting and stacking. The three multi-classification classifiers they used are Decision Tree, Random Forests, and XGBoost.
4. **Conclusion:** The experimental outcomes revealed that the framework efficiently identifies six distinct types of cyberbullying, boasting an accuracy rate of 0.9071. Surpassing previous studies, this experiment yielded superior results across all evaluation metrics, effectively classifying entire datasets. The results indicate that both voting and stacking ensemble classifier techniques outshine recent experiments targeting cyberbullying detection with the same dataset.

**Limitations:** While the proposed framework demonstrates superior accuracy in cyberbullying detection, it exhibits limitations, particularly in capturing nuanced word meanings. Moreover, the study solely relies on Twitter data, raising doubts about its effectiveness across diverse social platforms. The researchers note their focus on English language data and express intentions to explore other languages in future endeavors. Additionally, the research exclusively addresses text classification, overlooking potential cyberbullying conveyed through emojis or images, posing another constraint.

**Synthesis:** This paper introduces a novel framework utilizing an ensemble of transformer models to detect various forms of cyberbullying on Twitter. By combining three multi-classification classifiers using voting and stacking ensemble techniques, the framework achieves an impressive accuracy of 90.71%. Despite its success, limitations include difficulties in capturing nuanced word meanings and the exclusive focus on English language data from Twitter, neglecting potential cyberbullying conveyed through emojis or images. Future research aims to address these limitations and explore detection across diverse social platforms and languages. Also working on detection of cyberbullying via emojis and images could be a good area of research