



A Comprehensive Study of Hate Speech Detection in Child Abuse-Related YouTube Comments: Insights from Traditional and Deep Learning Approaches

S. Sweshthika* and R. Yasotha

Department of Physical Science, Faculty of Applied Science, University of Vavuniya, Sri Lanka

Abstract: The widespread use of social media has significantly contributed to the proliferation of user comments by posing serious threats to individuals and communities. Due to the emotionally charged and nuanced nature of the language, it presents unique challenges for detecting hate speech in comments. YouTube is a video-sharing and social media platform that promotes interaction all over the world. This research considers YouTube comments relating to child abuse. This research study proposes a comprehensive machine learning-based approach to automatically identifying hate speech in YouTube comments related to child abuse. A dataset of approximately 2,500 comments was collected through web scraping using Selenium and ensured a balanced mix of hate and non-hate speech. The dataset was processed using advanced natural language processing (NLP) techniques, including CountVectorizer, TF-IDF, Word2Vec, and Fast-Text, to extract key textual features. Various machine learning models were tested; among them, the gradient boosting model combined with CountVectorizer achieved the highest accuracy at 78%. Ensemble methods such as soft voting and stacking classifiers also demonstrated strong performance by reaching an accuracy of up to 75%. The research highlighted the importance of precision and recall metrics in evaluating model effectiveness. The Gradient boosting model's superior performance underscored its potential in enhancing hate speech detection systems, offering actionable insights for policymakers and platform administrators. By addressing hate speech in discussions related to sensitive issues like child abuse, this study contributes to the creation of safer, more respectful digital environments.

Keywords: Child abuse, CountVectorizer, Gradient boosting, Hate speech, Machine learning, Selenium, YouTube.