

Detection of Cyberbullying to Reduce Mental Health Problems using Machine Learning Algorithms

M. V. V. Perera^{1*}, K. M. Piyumal²

¹*Department of Computer Systems Engineering, University of Kelaniya, Dalugama, Sri Lanka, vihanm@kln.ac.lk*

²*Department of Computer Systems Engineering, University of Kelaniya, Dalugama, Sri Lanka, madhushkak@kln.ac.lk*

Social networks and other online platform services are where people are more likely to experience issues with cyberbullying, including kids, young and older adults who are addicted to them. Cyberbullying is an activity that takes place on digital platforms. Victims are threatened or bullied individually or in groups by messages or comments online. Various cyberbullying detection techniques are continuously used on social media platforms. However, not all online platform services follow those cyberbullying mechanisms, which may lead to psychological problems that can cause depression and lead to suicide because people are unaware of taking action to prevent it. Many past cyberbullying detection studies used a small data set and omitted to disclose the total number of features used to train the model. To fill this gap, this study explores how the model performance changes with the feature count and what happens to the model performance when the data set size increases. Therefore, two cyberbullying datasets with a combined total of 47,183 and 120,556 were used, which contained suspicious activities on Twitter and Facebook that most commonly belong to the cyberbullying category. To compare the performance metrics of each model, three methods for feature extraction and three classifiers were used, namely, Logistic Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). The highest accuracy for the models created utilizing 47,183 data under the three feature extraction approaches was 94.43%, and the highest accuracy for the 120,556 data was 89.96%.

Keywords: *Cyberbullying detection, Cybercrime, Feature extraction, Machine learning*