

## **ABSTRACT**

Cyberbullying is an act that harass, humiliate, and hassle people by means of electronic devices, including the Internet. Cyberbullying in internet is more dangerous than traditional bullying because it has a big potential to amplify to a much wider audiences. UNICEF and a survey conducted by the Indonesian Ministry of Communication and Information have revealed that 58% of 400 adolescents in Indonesia did not understand about cyberbullying. Some of them might ever be the bullies, but they did not understand about cyberbullying and neither recognize the negative effect of what they have done. The bullies might not recognize the harm of their act because they do not see the immediate response from their victims.

This research aims to detect Indonesian cyberbullying texts and actors. This research used data collection from Twitter and since the data are unlabeled, a web-based labelling tool is developed to classify tweets into cyberbullying and non-cyberbullying tweets. This tool collected 301 cyberbullying tweets, 399 non-cyberbullying tweets, 2,015 negative words, and 129 swearwords. Afterwards, SVM and KNN algorithm was applied to learn and detect cyberbullying texts. The result showed that SVM results in the highest f1-score i.e. 67%. The credibility of Twitter users was measured. As the results there were of 257 Normal Users, 45 Harmful Bullying Actors, 53 Bullying Actors, and 6 Prospective Bullying Actors.

Besides, this research aims to analyze the impact of cyberbullying notification to raise the Indonesian cyberbullying awareness. After the system found the cyberbullying actors, it sent cyberbullying notification to the actors. The results showed that the system was not able to raise cyberbullying actors awareness because there were no replies to the notification nor feedback to the website from the actors.

**Keywords:** cyberbullying detection, cyberbullying actors, SVM, credibility analysis of user