

Accurate Cyberbullying Detection Using Software Based Approach

A Dissertation By
Andrea Perera
w1654556/ 2016432

submitted in partial fulfilment of the requirements for the
BSc (Hons) Computer Science degree

Department of Computer Science

INFORMATICS INSTITUTE OF TECHNOLOGY
In Collaboration with
UNIVERSITY OF WESTMINSTER, UK (UOW)

2020

Abstract

The usage of digital/social media is increasing day by day with the advancement of technology. People in the twenty-first century are being raised in an internet-enabled world with social media. Communication has been just one button click. Even though there are plenty of opportunities with digital media people tend to misuse it. People spread hatred toward a person in social networking. Cyberbullying affects people in different aspects. It doesn't affect only for health, there are more different aspects which will lead a life to threat. Cyberbullying is a worldwide modern phenomenon which humans cannot avoid hundred percent but can be prevented. Most existing solutions have shown techniques/approaches to detect cyberbullying, but they are not freely available for end-users to use. They haven't considered the evolution of language which makes a big impact on cyberbullying text.

The author proposes a system for automatic detection and prevention cyberbullying considering the main characteristics of cyberbullying such as Intention to harm an individual, Repeatedly and over time and Using abusive curl language or hate speech using supervised machine learning. The system relies on the detection of cyberbullying text along with the themes/categories associated with cyberbullying such as race, sexuality, physical and Political, using support vector machines and Logistic regression.

Author of this research presents a new hypothesis for cyberbullying detection that the circumstances and usage of texting and its language have changed by time. Most of the studies have considered calling someone stupid, ugly and idiot as cyberbullying. I do believe that things have changed, such words (stupid, ugly and idiot) may or may not always be a bullying incident. Now a days we use Dumbass for an idiotic person. If a person wants intentionally to harm an individual, they will use extreme words such as Cunt, Bitch, Dickhead etc., not just idiot, stupid. Evaluated the proposed system using Recall, Precision-score curve and confusion matrix along with accuracy of 74%.

Keywords: Cyberbullying; Supervised machine learning; Natural language processing; Social media; Health problems.