

**DETECT HATE AND OFFENSIVE SPEECH
(SINHALA/ ROMANIZATION SINHALA TEXT) IN
SOCIAL MEDIA**

UDITH ISURU SENANAYAKE

A dissertation submitted in partial fulfilment of the requirement for
Master of Sciences (Honours) degree in Big Data Analytics

**Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University Aberdeen, UK**

2022

Abstract

Social media become a merge part of our lives. People more often use these platforms for various reasons. As these platforms create a cyber world for community different minded people join for these for various kind of reasons. As anything have two sides, the same theory has valid for the social media also. Because people use good technologies for do bad and dark things. Some people join these communities for searching for victims. In the past decade numerous people lost their lives because of the incidents took place in social media platforms. As a result of these situations most of social media users have a doubt whether the social media is a safer place or not. Share hateful content throughout these networks, share radical thoughts, cyber bullying, cyber threaten, cyber abusing, cyber harassment, malicious communications, share deception news and threats on sexually humiliation are most common activities in these days.

To prevent those activities and create a safe social media community different social media companies use different technologies over their networks. Also, different countries apply different laws to prevent these situations. Some companies, organizations and research communities try to find a solution with the use of computer vision, artificial intelligent, data mining, image processing and text analysis. When considering text analysis approaches most research conduct either to detected hate speech or cyberbullying. But there is no proper research conduct to identify both at the same time. Also, there are very lesser research conduct for native Sinhala languages. So, this research conduct to narrow down that knowledge gaps in same time.