Informatics Institute of Technology

In collaboration with

University of Westminster, UK

A Machine Learning Approach in the Identification of Sinhala Toxic Language on Social Media

A dissertation by

Ms. Shavindya Kariyawasam

Supervised by

Ms. Aloka Fernando

Submitted in partial fulfillment of the requirements for the

BEng (Hons) in Software Engineering degree

Department of Computing

May 2019

Abstract

With the technological advancement of all means of communication, Social media has become an essential part of our daily routines. Although it paves the way to achieve enhancements in our sociability, diving too deep into social media will always place anyone at the risk of facing online harassment. In popular Social networking platforms, intentional attacks such as fraud identity, releasing private pictures and causing harm verbally through toxic language, can be observed to a great extent.

Toxic language, which can also be defined as rude and disrespectful language, plays a major role in online harassment. With the increasing number of online harassment cases that get reported every day, automatic detection of toxic language has received global attention. Hence, several attempts have been taken in order to identify toxic language in European and Non-European languages such as Arabic and Hindi. Yet, the identification of toxic language in Sinhala, is an area of research that has not been addressed before.

In the present day, Sinhala is spoken by a major percentage of the total Sri Lankan population, and there are around 6 million active social media users in the country. With the rise of severe cases of online toxicity that lead Sri Lankan authorities to go as far as blocking social media platforms in the country several times, it is clear that the automatic detection of Sinhala toxic language needs to be given more attention.

Thus, with online harassment and toxicity being identified as a major issue in Sri Lanka, in this research, an initial attempt has been taken to provide a solution for online harassment. Hence, for the purpose of identifying Sinhala toxic language, a machine learning approach that follows multi-label text classification techniques has been proposed.

Keywords:

Online harassment, Toxic language, Sinhala text classification, Machine learning, Multi-label classification, NLP

Shavindya Kariyawasam