



INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

HATE SPEECH DETECTION FOR LOW RESOURCE LANGUAGE.

A Final Thesis by

Ms. Thushamini Chathushika.

W1789943 / 2019644

Supervised by

Mr. Mithushan Jalangan

This project thesis is submitted in partial fulfillment of the requirements for the
BEng(Hons) Software Engineering degree at the
University of Westminster.

02nd May 2023

Abstract

Despite all the benefits offered by the internet, we occasionally encounter harmful behaviors including cyberbullying, cyberthreats, emotional abuse, malicious communication, and online hate speech. With the recent rise in internet usage, hate speech has a tendency to become more prevalent online. On most platforms, there aren't many hurdles to prevent the aforementioned offenses because Sinhala is a low-resource language. Sinhala comments and texts can also be written in a variety of formats and writing styles, which makes it more difficult to conduct research on them. The foremost focus of this study is to recognize hate speech and detect it using emojis in Sinhala. The chosen strategy is chosen from a variety of opposing points of view in order to arrive at a single, reliable solution. Before detecting hate speech, the author suggests finding the distinctive sinhala text writing styles. In contrast to the majority of other studies on this subject of hate speech, the method used here does not simply rely on the detection of banned terms. In this strategy, the author is experimenting with machine learning and deep learning techniques.

The accuracy of the LSTM Model that the author was able to develop was about 86%. In order to validate and confirm the outcomes from the training and testing process, the author further analyzed the framework in use of the metrics recall, F1, precision, confusion matrix, and graphs.

Keywords - Hate Speech, Natural language Processing , Deep Learning, Machine learning

ACM Subject Descriptions

Computing methodologies → Artificial intelligence

Computing methodologies → Machine learning → Machine learning approaches → Neural networks