



INFORMATICS INSTITUTE OF TECHNOLOGY

**In Collaboration with
UNIVERSITY OF WESTMINSTER**

**Automated Hate Speech Detection for
Transliterated Forms of Low Resource Languages**

Final Project Report (FPR) by

Mr. Mahen Dunusinghe

Supervised by

Ms. Ganesha Thondilege

Submitted in partial fulfilment of the requirements for the BEng in Software
Engineering degree at the University of Westminster.

July 2022

Abstract

Hate speech has been on the rise for quite a while now, big tech moguls such as Facebook, Instagram and Twitter have attempted to reduce hate speech on their platforms, spent millions on algorithms that could reduce hate on their platforms. Even such attempts have not managed to eradicate or produce results that could make a substantial change. The power of social media is quite often underestimated, with enough people grouping together it can bring down a nation even through digital means with just the right provoking.

Low resource languages are now being used in social media platforms, various transliterate forms of such languages have very low hate speech detection capabilities in most platforms. Since most platforms have the lack of knowledge on such languages it is quite hard for them to act swiftly when riots occur through social media platforms. Romanized Sinhala is one such language that will be taken as the primary language throughout this research.

This research involves a suitable method to automatically detect hate speech on transliterate forms of languages in social media platforms. The researcher hopes to deliver findings through extensive research on existing solutions, benchmarking and evaluations on the proposed solution. According to the authors evaluation this system can be vital for the ongoing battle of hate speech detection in social media platforms.

Keywords: Hate Speech, Transliteration, Machine Learning, Natural Language Processing, Deep Learning