LANKAN HATE CHECK - A DEEP LEARNING APPROACH THE DETECTION & CATEGORIZATION OF SINHALA OR SINGLISH HATE SPEECH

S. C. SIRIMANNE

A dissertation submitted in partial fulfilment of the requirement for Bachelor of Science (Honours) degree in Computer Science

School of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2023

ABSTRACT

With the increasing use of the internet and social media, the spread of hate speech has become a major concern. This research aims to address the issue of hate speech detection and categorization in the context of Sri Lanka, specifically for Sinhala and Singlish (a mix of Sinhala and English) languages. While the existing research in the domain of hate speech detection in the Sri Lankan region focuses on Sinhala or Singlish hate speech detection with machine learning, those solutions do not possess the ability to categorize hate speech. This system uses deep learning to detect and categorize both Sinhala and Singlish hate speech, which has not been addressed thus far. A novel approach is taken by the author to train seven LSTM models with binary classification to obtain the overall result of each of the seven models, which are used to produce the final result. Seven datasets were manually created for the purpose of training the models. A back-transliteration function is used to convert the Singlish text into Sinhala and then fed to the model. The idea was to find a complete solution that can automatically detect hate speech and warn the users to in order to reduce Singlish and Sinhala hate speech in Sri Lankan social media. Notably, this research looks into categorizing Sinhala and Singlish hate speech into racism, sexism and xenophobia, addressing the gap in existing research.

Key Words: Sinhala, Singlish, Hate Speech Detection, Hate Speech Categorization, Deep Learning, NLP, LSTM