HATE SPEECH DETECTION AND SENTIMENT ANALYSIS
ON CONTENT WRTTEN IN SINHALA LANGUAGE:
A STUDY ON SOCIAL MEDIA AND E-COMMERCE SITES

Neranjaya Bandara Ehelepola

A dissertation submitted in partial fulfilment of the requirement for

Master of Science degree in Business Analytics

School of Business

Informatics Institute of Technology, Sri Lanka in collaboration with

Robert Gordon University, UK

2022

Abstract

In the early days, social media was used as a platform to connect people. With the advancements in the social media platforms such as Facebook marketplaces and higher volumes of users engaging social media is currently used as a prominent business platform by many small, medium, and large businesses to promote and brand their businesses. In the Sri Lankan context, the rise of social media users has increased gradually over the years and user engagement is higher than ever now. The e-commerce platforms are another platform where sellers and users are met. There are several e-commerce platforms in Sri Lanka and Daraz is one of the main e-commerce platforms used by Sri Lankans.

Writing content using Sinhala letters was not very popular among Sri Lankans before developing virtual keyboards such as Helakuru and now it is used by most Sri Lankans. Sri Lankan people engaging in social media platforms, and e-commerce platforms and expressing their ideas using the Sinhala language has become a common thing. The study is done to analyze the content written in the Sinhala language. The study mainly focuses on businesses on social media platforms and other e-commerce platforms. By analyzing the textual data, the study has been able to detect whether the content written is Hate speech or not which ensures that the written content does not violate community standards. Furthermore, the study has focused on doing a sentiment analysis of the content whether it is positive or negative and which aspect of the business the comment is regarding.

Machine learning models are used in analyzing and doing predictions. Several machine learning models are trained and tested to come up with the best solution for the research problem. The results of the study can be used to detect hate speech written by a user before posting and block or warn the user before posting it will minimize the harm done to the brand of the company and it will make the platform a safe space for the users. The sentiment analysis can help the sellers and the platform to analyze how the users are responding to the service of the sellers helping them to improve their business and expand the customer base.