Identify Hateful Comments in Sinhala Language on Social Media

W.W.E.N. Fernando
189461H

Faculty of Information Technology
University of Moratuwa

July 2021

Identify Hateful Comments in Sinhala Language on Social Media

W.W.E.N. Fernando
189461H

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of Degree of Master of Science in Information Technology.

July 2021

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student
W.W.E.N. Fernando

Signature of Student

Date

Supervised by
Name of Supervisor
Mr. S. C. Premaratne

Signature of Supervisor

Date

Acknowledgements

I am so grateful to my supervisor Mr. Saminda Premaratne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for academic guidance, advice, time, encouragement and patience that has seen me through, making this work success.

Also I would like to give a special thanks to the Prof. Rathnasiri Arangala, Senior Professor in Sinhala, University of Sri Jayewardenepura for his guidance, advice, time and encouragement in carrying out this research. As well as special thanks to Mr. K.K. Premarathne, Former Principle, Bandaragama National College for his precious support as an annotator and his encouragement.

Furthermore, my thanks also go out to my beloved family members for their support, love, advice and encouragement throughout the period of my study. I lastly wish to extend my sincere gratitude toward my colleagues who have tirelessly guided and encouraged me at all times through my course.

Abstract

In present, the spread of hate speech through social media has become a very serious problem, both globally and locally. The route cause for this is the increasing use of social media with the rapid expansion of computer science and information technology. Therefore, it is very important to use same to control this kind of situations. Although there is a mechanism in place on social media to automatically control such hate speech in English language, but it is still not seen in Sinhala Language. The reason for this is the lack of knowledge about the native languages such as Sinhala in the social media service providers. Therefore, the identification of hateful contents in Sinhala language is an urgent and vital task that needs to be addressed.

This research propose lexicon based and machine learning based approaches for the automatic identification of hateful speech in Sinhala on social media. With different pre-processing techniques and machine learning algorithms, machine learning algorithm based approach was conducted with four different approaches. These approaches were begun with 3000 comments which is equally divided into hateful and non-hateful. Using these comments, it was able to identify the most appropriate featured groups and model to identify the hateful speech in Sinhala language on social media.

Table of Contents

List of Figures

List of Tables