HATE CHECKER - HATE SPEECH DETECTION IN SINHALA-ENGLISH CODE-MIXED LANGUAGE

Oshadhi Gayangika Liyanage

A dissertation submitted in partial fulfilment of the requirement for
Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2021

Abstract

With the steady increase internet usage, the propagation of hate speech on the internet also steadily increases. Social media sites, review forums, micro blogging sites encourage users to convey their thoughts with minimum restrictions. This leads to expressing hate towards others who do not believe their beliefs. This study focuses on identifying hate speech texts that are written in Sinhala-English code-mixed language (Singlish) which is mostly used by Sri Lankans on the internet. Due to the unavailability of Sinhala-English code-mixed datasets, dataset was created using comments on YouTube and Facebook. Eight machine learning algorithms and three ensemble approaches were developed to detect hate speech in Singlish and their performance were evaluated in terms of accuracy, f1-score precision and recall. Support Vector Machine (SVM), Multinominal Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB) and Logistic Regression classifiers were used to develop ensemble approaches. They were developed using soft voting, hard voting, and stacking. The stacking approach outperformed other baseline algorithms with 85.71% accuracy and 83.78% f1-score..

Keywords— Hate Speech Detection, Sinhala-English Code-Mixed Language, Ensemble Approach, Stacking, Singlish