Paper No: SC-09 **Smart Computing**

Hate Words Detection Among Sri Lankan Social Media Text Messages

J. A. D. U. Shalinda*
Software Engineering Teaching Unit
Faculty of Science, University of Kelaniya, Sri Lanka
udithshalinda2@gmail.com

Lankeshwara Munasinghe
Software Engineering Teaching Unit
Faculty of Science, University of Kelaniya, Sri Lanka
lankesh@kln.ac.lk

***Abstract* - The number of Sri Lankan social media users have been increased with the rapid growth of 23% between 2020and 2021, reaching 7.9 million in 2021 January. Social media platforms became more popular when they started supporting native languages. The problems with social media also evolved as popularity grows. Social media platforms were banned for SriLankan users in 2019 to prevent the spreading of hate messagesand incorrect information among citizens. The lack of automatically recognizing tools for hate messages in Sinhala andRomanized Sinhala was reported as the reason for the ban. It'salso a waste of time and money to manually identify them. Manystudies have been conducted to identify hate messages in both English and Sinhala separately. Users in Sri Lanka tend to combine Sinhala, Romanized Sinhala, and English phrases while expressing their opinions." Mama job ekakata apply kara," for example. To train, an open-source data set which consists of 2500 comments, was used. And the comments were categorized as either hateful or non-hateful. To pre-process thedata set, an Open-source stop word corpus and stem word corpus in Sinhala were utilized, and two corpus were manually converted into Romanized Sinhala stop word corpus and Romanized Sinhala stem word corpus to identify stop words andstem words in Romanized Sinhala. All English words wererecognized using an open- source English word corpus, and a library was utilized to obtain stop word corpus and stem Englishwords. As a result, doing research to identify hate speech in all of the languages indicated above will be more effective in reaching Sri Lankan users. The bag of words and term frequency-inverse document frequency were compared for feature engineering. Linear Support vector classifier, Random Forest Classification, SGD classifier, Logistic Regression, XGBoost classifier and multinomial Naive Bayes classifier are used as classification algorithms and evaluated. Using the SGD classification using TF-IDF with uni&bi-gram, the highest accuracy was determined to be 74.2%.**

Keywords - English, hate speech detection, NLP, RomanizedSinhala, Sinhala

[1]