

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER (UOW)

MIZARD

Multilingual Sentiment Analysis on Social Media Platforms

By

2014509 - Sashini Piyaratne

Supervised By

Krishnakripa Jayakumar

Submitted in partial fulfillment of the requirements for the

BEng/BEng. (Hons) in Software Engineering

Department of Computing

2018/2019

©The copyright for this project and all its associated products resides with
Informatics
Institute of Technology.

Abstract

In present day society, people have become increasingly active on social media and now include it as a part of their daily life. In Sri Lanka, the case stands the same, where people use multiple languages to communicate on social media. Sinhala, English and Tamil are some of the most commonly used tongues, while combination languages such as Singlish, a mixture of Sinhala and English, are also widely used. Being a double edged sword, social media can have both favorable and unfavorable outcomes.

For example, if considering the negative aspect of social media, extremists of different religions can use it to spread discord by voicing racist opinions and theories that may result in societal conflicts; building mistrust among communities. On the other hand, social media has its positive side as well; it can be used to build understanding among different cultural and religious groups, where people will be exposed to the diverse viewpoints of these groups, encouraging them to be more open-minded and creating a sense of unity.

To solve this issue, many researches have been conducted using several approaches. Yet a major gap in these research areas lie in the fact they are targeted to detect hate speech in a single language, without taking mixed languages into consideration.

Through an in-depth literature review and domain study, MIZARD will demonstrate how to identify racism in a mixed language. Only Singlish (Sinhala words written in English) will be considered here. Using Neural networks, MIZARD's system has been developed as a chrome plug-in where, once it is enabled, Singlish racist posts will be detected and covered. The user may uncover the post by hovering the mouse over it.

System accuracy was tested on tweets that were written in Singlish. The usage of Neural networks make MIZARD produce acceptable results with better performance.

Key words: Machine learning, Natural language processing, Sentiment analysis, neural networks