INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER, UK

Detection of Mental Stress Level via code mixed Social media posts using Sentiment Analysis

A Dissertation by

Miss Sandani Gunasekara

W1810476 | 20200208

Supervised by

Mr. Guhanathan Poravi

Submitted in partial fulfilment of the requirements for the MSc in

Advanced Software Engineering degree at the

University of Westminster.

September 2023

ABSTRACT

Detecting mental stress level using code-mixed social media posts lies in the unique linguistic and contextual challenges posed by the combination of multiple languages within a single post. Code-mixing, the phenomenon of blending languages, is prevalent in online communication, especially on social media platforms. This mixing of languages introduces complexities in natural language processing tasks, making it difficult to accurately detect and classify mental stress levels. The inclusion of multiple languages, variations in grammar and syntax, slang, and cultural references further complicate the analysis process. As a result, existing stress detection methods may not adequately handle code-mixed data, leading to reduced accuracy and effectiveness in identifying and understanding mental stress levels within this specific linguistic context.

This problem requires a resolution that utilizes machine learning and natural language processing methods to autonomously handle and assess social media data, offering valuable insights into the emotional stress levels of people. By addressing this problem, the stress detection system aims to contribute to early intervention, support, and resources for individuals experiencing mental stress.

The classification report underscores the precision, recall, F1-score, and the general accuracy of the model. By using more than 23000 data, SVM algorithm showed the best accuracy of 88% while random forest algorithm showed 84% and gradient boosting had 85%. After performing cross validation as well, SVM showed the best accuracy with least time to process the model. Therefore, SVM algorithm is used to build the model.

Subject Descriptors:
- Computing methodologies → Machine learning → Learning paradigms → Supervised learning
- Computing methodologies → Machine learning → Machine learning algorithms → Ensemble methods