SUICIDAL TENDENCY PREDICTION USING TEXT ANALYTICS:
A STUDY ON SRI LANKAN FACEBOOK COMMUNITY

Nipuni Karunaratne

A dissertation submitted in partial fulfilment of the requirement for Master of Science degree in Business Analytics

School of Business

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University, UK

2022

Abstract

Suicidal tendencies occur mostly due to excessive stress, anxiety, and prolonged depression conditions. Suicide has become one of the leading causes of death in the world according to WHO statistics. Sri Lanka has a higher suicide rate than the world average. This has a huge impact on the country's economy as well as the social well-being of the victims' loved ones. With the growing popularity of social media platforms, people have gotten accustomed to sharing anything and everything on social media platforms, even posts with suicidal ideations. Therefore, such platforms can be utilized positively to diagnose individuals with suicidal ideations and severe depression conditions and direct them for medical assistance.

This study has been conducted to predict suicidal tendencies in individuals using Facebook posts from users in Sri Lanka. Machine Learning and NLP algorithms such as KNN, Random Forest, Logistic Regression, SVM, Gradient Boost, Multinomial Naïve Bayes and BERT have been used to build the models to predict suicidal tendencies using text analytics. Different word embedding techniques such as Word2vec and TF-IDF vectorization have also been compared in this study.

Moreover, it was evident from the preliminary analysis of the dataset used in the study that, people from Sri Lanka uses a different key set of words when depicting their suicidal ideations than the words used by people in Western countries. Therefore, models trained on data from other countries with different cultural and social backgrounds cannot be used to successfully predict suicidal tendencies in individuals.

According to the study's results, the machine learning algorithms have succeeded in achieving this study's objective. Among all the algorithms compared during the study, the TF-IDF-Base KNN model and BERT model have given the best outcomes in model evaluation. These developed models could be utilized to identify individuals with suicidal tendencies and direct them to proper medical care.

Keywords- Suicidal Tendencies, Facebook, Machine Learning, Natural Language Processing