

**PREDICTION OF SOCIAL MEDIA TEXT-BASED
COMMENTS TOXICITY DURING SRI LANKAN CRISIS
WITH BOOSTING ALGORITHMS**

Nethmie Kulanga Pasqual

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

**Department of Computing Science and Digital Media
Informatics Institute of Technology, Sri Lanka
In collaboration with
Robert Gordon University, Aberdeen, UK**

2023

ABSTRACT

Nowadays, most of the times, communication happens through social media platforms like Twitter, Facebook and Youtube. Therefore, loads of information produced on a daily basis through social media platforms and online forums. Comments are a way of expressing one's opinion about some content on those platforms. But in some cases, people tend to use these platforms to spread disrespectful, harmful thoughts about others by posting abusive content. It negatively impacts the content creators and other viewers, closing off the entire comment section for all users. These types of comments are called toxic, and they should be avoided in every possible way. There are many studies conducted for identifying the toxicity and filtering out those comments using various methodologies. The machine learning techniques are most popularly used in this context, but there is a lack of research conducted to evaluate those methodologies comparatively. This research analyzes such text-based comments made during the Sri Lankan Crisis which is still ongoing.

This study discusses what classification algorithms that can be used in this problem domain and focuses only on text-based comments while analyzing three separate data sets retrieved by manually by web scraping on Facebook. Those data sets contained somewhat of a large number of comments and each and every dataset was trained and tested on six classification models considering the distribution of the dataset as well. As the feature extraction techniques, Bag-of-words (Bow) and Tf-Idf were used. Boosting Algorithms like Gradient Boosting, Catboost, XGBoost, Adaboost, Stacked Models and Hybrid Models are used as the classification algorithms to train the datasets. This study is based on a supervised machine learning approach that uses binary classification as the classification technique.

Key Words – Toxicity Prediction, Classification Models, Boosting Algorithms, Sri Lankan Crisis, GoHomeGota