

**DETECTING HUMAN EMOTION FROM SOCIAL
MEDIA COMMENTS USING SENTIMENT ANALYSIS
TECHNIQUES**

Loku Bogahawaththege Amila Sankha Dayarathna

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

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

Robert Gordon University, UK

2021

Abstract

Contemplating the fact that business enterprises and individuals, globally, benefit from social media to gain profits, this dissertation discusses the importance and impact of analyzing mechanisms of consumer responses on social media to maximize the gained profit and take befitting future decisions while reducing the risk of cyber-crimes and terrorism. The main objective is applying sentiment analysis techniques on the collected responses to analyze them in order to identify the emotions behind. This study focuses on finding out the best model for emotion categorization of the collected responses using the sentiment analysis techniques.

The collected responses are preprocessed and undergo several techniques to achieve data normalization which enable fast and effortless querying. Cleansed data is analyzed in order to place the comments under six main emotion categories related to sentiment analysis. This analysis mainly consists of feature extraction and training the models. The two built-in methods available in scikit learn toolkit, namely, *TfidfVectorizer()* and *CountVectorizer()* are used for the feature extraction using bi-gram features and those extracted features are used to train the selected four (04) classification models in order to find out the best performing model with the two (02) feature types extracted. Subsequently, this study will investigate and compare different features for the different classifiers when categorizing emotions on social media. Then record the accuracy of each model with each feature type and select the best performing model with the accuracy of 80% with a 0.1909 train-test accuracy difference.

Keywords: Social Media, Emotion Categorization, scikit-learn, Sentiment Analysis, Feature Extraction