

Text-based emotion prediction system using machine learning Approach

1 Ahmad Fakhri Ab. Nasir, 1 Eng Seok Nee, 1 Chun Sern Choong, 1 Ahmad Shahrizan
Abdul Ghani, 1 Anwar P. P. Abdul Majeed, 1 Asrul Adam and 2 Mhd. Furqan

1 Faculty of Manufacturing and Mechatronic Engineering Technology,
Universiti Malaysia Pahang, Pekan, Malaysia

2 Department of Computer Science, Universitas Islam Negeri Sumatera Utara Medan

Abstract:

Text-based input becomes a common channel for humans in sharing their opinions/emotions to the product or service through online social media, shopping platform etc. Humans are easy to make errors in interpreting emotions, especially the emotion that derived from text based. The main aim of this study is to develop text-based emotion recognition and prediction system. Several market challenges facing in the advancement of emotion analysis with accuracy being the main issue. Therefore, four supervised machine learning classification algorithms such as Multinomial Naïve Bayes, Support Vector Machine, Decision Trees, and kNearest Neighbors were investigated. The model was developed based on Ekman's six basic emotions which are anger, fear, disgust, joy, guilt and sadness. Data pre-processing techniques such as stemming, stop-words, digits and punctuation marks removal, spelling correction, and tokenization were implemented. A benchmark of ISEAR (International Survey on Emotion Antecedents and Reactions) dataset was used to test all models. Multinomial Naïve Bayes classifier resulted the best performance with an average accuracy of 64.08%. Finally, the best model was integrated to graphical user interface using Python Tkinter library to complete the whole system development. Besides, the detailed performance of the best model such as tf-idf and count vectorizer, confusion matrix, precision-recall rate, as well as ROC (Receiver Operating Characteristic) score were also discussed. Text-based emotion prediction system to interpret and understand human emotions was successfully developed.

Keywords: Multinomial Naïve Bayes; Support Vector Machine; Decision Trees; Receiver Operating Characteristic (ROC)

Acknowledgement

The authors acknowledge the internal grant from Universiti Malaysia Pahang (RDU180383).