

ABSTRACT

This study aims to enhance sentiment classification performance in the spiritual dimension of multidimensional sentiment analysis on COVID-19 vaccination-related tweets in Indonesia. Three keyword expansion methods were evaluated: Mass Diffusion of Co-occurrence Approach (MDCA), Word2Vec, and TF-IDF. MDCA was tested with Top-10, Top-20, Top-30, and Top-50 configurations. All expanded keywords underwent human evaluation by three spiritual experts. Validated keywords were used to collect new tweets, which were then preprocessed, sentiment-labeled using a fine-tuned IndoBERT model, and merged with the original dataset. Each resulting dataset was tested using Naïve, SVM, IndoBERT Base, and IndoBERT Large P2. The results showed that MDCA Top-30 performed best in traditional classifiers, with accuracy improvements of up to +2.6% and F1-score gains of +8.0% in SVM, while MDCA Top-50 achieved the highest performance in deep learning models, yielding up to +7.8% accuracy and +10.0% F1-score improvement in IndoBERT Base. Misclassified tweet analysis revealed frequent errors related to geographic and educational terms, suggesting the need for additional dimensions. Overall, the findings demonstrate that MDCA, particularly Top-30 and Top-50, is more effective than Word2Vec and TF-IDF in expanding spiritual-related keywords and improving sentiment classification in multidimensional analysis of COVID-19 vaccine discourse in Indonesia.

Keywords: keyword expansion, multidimensional sentiment analysis, spiritual dimension, COVID-19 vaccination tweets, mass diffusion of co-occurrence approach (MDCA)