ABSTRACT

The 2024 East Java Gubernatorial Election has become a widely discussed topic on Twitter, reflecting public opinion on the candidate pairs. This study utilizes Support Vector Machine (SVM) to automatically classify tweet sentiments. The data was collected using Tweet-Harvest and then processed with pre-processing techniques such as case folding, tokenization, special character removal, and word normalization. Next, the data was represented using the Term Frequency-Inverse Document Frequency (TF-IDF) method to construct features for classification. The model was tested in several scenarios, including experiments with the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance and the impact of stopword removal on model performance. Evaluation was conducted using a confusion matrix with Accuracy, Precision, Recall, and F1-Score metrics. The results showed that the best approach was without stopword removal and without SMOTE, achieving an accuracy of 0.67. The presence of stopwords helped the model distinguish sentiment, especially between negative and neutral classes. The main challenges in this study were class overlapping, which made sentiment classification difficult, and the sparse matrix in TF-IDF representation, which hindered the model from capturing optimal patterns. Future research can explore other methods and text representations based on Word2Vec or BERT to improve classification accuracy.

Keywords: Sentiment Analysis, 2024 Election, SMOTE, Support Vector Machine, Stopword Removal, TF-IDF