ABSTRACT

The 2024 Central Java Regional Election (Pilkada) was a major political focus in Indonesia, as the province contributed the largest number of votes and had no incumbent running, creating opportunities for new candidates to gain public support. Social media platform X played a crucial role in political communication among the public and political figures, making it a valuable source for analyzing public sentiment toward competing candidates. This study aimed to compare the accuracy of various feature extraction methods in public sentiment analysis regarding the 2025–2030 Central Java gubernatorial candidates using SVM and to examine the correlation between sentiment analysis of tweets about the two candidates and the official results of the 2024 Central Java Regional Election. Feature extraction methods were used to convert text into numerical representations that could be processed by the SVM model, which then classified public sentiment into positive, neutral, and negative categories. The results showed that the use of TF-IDF feature extraction with the RBF kernel achieved the highest accuracy of 76.34%. Meanwhile, the Bag of Words method obtained an accuracy of 66.87% with the RBF kernel as the optimal choice. Word embedding-based feature extraction techniques, such as Word2Vec and FastText, yielded lower accuracy rates of 62.75% and 56.18%, respectively. The sentiment analysis results did not align with the official election results. Although candidate Andika Perkasa received the highest positive sentiment at 65.79%, the election results showed that Ahmad Luthfi won with 59.14% of the votes.

Keywords : Feature Extraction, TF-IDF, Bag of Words, Word2Vec, FastText, Sentiment Analysis, Gubernatorial Candidates, Regional Election, Central Java, SVM