

ABSTRACT

The role of social media in shaping public opinion is increasingly significant, especially in the context of local politics such as the 2024 East Java Regional Head Election (Pilkada). This study aims to build a sentiment classification model for governor and deputy governor candidates based on uploads on social media X. The method used is the Random Forest algorithm, which is applied through the stages of text preprocessing, sentiment labeling, feature weighting using TF-IDF, and feature selection with Chi-Square and Mutual Information. To overcome the class imbalance, sampling techniques such as Random Undersampling, SMOTE, Borderline-SMOTE, and a combination of SMOTE + Tomek Links were applied, tested at data ratios of 90:10, 80:20, and 70:30, respectively. Performance evaluation was carried out using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The best results were obtained in the combination of TF-IDF with Borderline SMOTE at a ratio of 80:20, with an accuracy of 82.40% and an F1-score of 82.44%. These findings suggest that proper balancing strategies play an important role in improving the effectiveness of the model. In addition, this study also implements an interactive dashboard based on Streamlit that displays visualization of sentiment distribution and input sentiment prediction directly, making it easier for users to explore data.

Keywords: *East Java Regional Elections, Sentiment Analysis, Social Media X, Random Forest, Feature Selection, Balancing Techniques.*