

ABSTRAK

Analisis sentimen masyarakat terhadap sengketa pemilihan Presiden Indonesia 2024 melalui sosial media X menggunakan algoritma *Random Forest*. X menjadi *platform* yang penting bagi masyarakat untuk menyampaikan opini dalam bentuk sentimen positif dan sentimen negatif. Algoritma *Random Forest* dipilih karena kemampuannya dalam mengelola data teks serta meningkatkan akurasi klasifikasi melalui teknik pohon keputusan. Tujuan dari penelitian ini adalah untuk menganalisis sentimen masyarakat terhadap sengketa pemilihan Presiden Indonesia 2024 dan mengukur kinerja model *Random Forest* dalam mengklasifikasikan sentimen positif dan negatif. Penelitian ini menggunakan sampel sebanyak 1.671 *tweet* berbahasa Indonesia. Perancangan sistem dilakukan melalui proses pengumpulan data, *preprocessing* data (*cleansing*, *case folding*, *tokenization*, *filtering/stopword removal* dan *stemming*) dan klasifikasi menggunakan algoritma *Random Forest*. Hasil klasifikasi ini kemudian diintegrasikan ke dalam sebuah aplikasi web menggunakan Streamlit, yang menyediakan *user interface* yang mudah diakses untuk melihat hasil analisis sentimen. Hasil pengujian menunjukkan bahwa model memiliki akurasi sekitar 76%. Evaluasi kinerja menggunakan *k-fold cross validation* menunjukkan akurasi rata-rata sebesar 76%, dengan *precision* 75%, *recall* sebesar 73%, dan *f1-score* sebesar 74% dengan variasi performa disetiap *fold* antara 70% hingga 82%.

Kata Kunci : Analisis sentimen, *Random Forest*, *Preprocessing*, *Streamlit*, *Confusion Matrix*, *K-fold Cross Validation*.



ABSTRACT

Sentiment analysis of public opinion on the 2024 Indonesian presidential election dispute through social media platform X using the Random Forest algorithm. X has become an important platform for the public to express opinions in the form of positive and negative sentiments. The Random Forest algorithm was chosen for its ability to handle textual data and improve classification accuracy through decision tree techniques. The aim of this research is to analyze public sentiment on the 2024 Indonesian presidential election dispute and to measure the performance of the Random Forest model in classifying positive and negative sentiments. This research uses a sample of 1,671 Indonesian-language tweets. The system design involves data collection, data preprocessing (cleansing, case folding, tokenization, filtering/stopword removal, and stemming), and classification using the Random Forest algorithm. The classification results are then integrated into a web application using Streamlit, providing a user-friendly interface for viewing sentiment analysis results. Testing results show that the model achieved an accuracy of approximately 76%. Performance evaluation using k-fold cross-validation showed an average accuracy of 76%, with a precision of 75%, recall of 73%, and an f1-score of 74%, with performance variation across folds ranging from 70% to 82%.

Keywords: Sentiment Analysis, Random Forest, Preprocessing, Streamlit, Confusion Matrix, K-Fold Cross Validation.

