

ABSTRAK

“UJI PELABELAN MENGGUNAKAN INDONESIA ROBERTA BASE SENTIMENT CLASSIFIER BERDASARKAN UJI ATTRIBUT HASIL LEMBAGA SURVEI DAN KLASIFIKASI SENTIMEN PEMILIHAN PRESIDEN MENGGUNAKAN ALGORITMA SVM”

Glorious Aldo Putra Jatmiko

Universitas Sanata Dharma

2024

Penelitian ini bertujuan untuk mengeksplorasi korelasi antara sentimen publik di media sosial *Instagram* dengan hasil survei untuk pemilihan presiden Indonesia tahun 2024. Analisis korelasi antara data sentimen hasil pelabelan dan hasil survei menunjukkan variasi yang signifikan antar periode yang berbeda. Beberapa periode menunjukkan korelasi positif yang kuat, sementara yang lain menunjukkan korelasi negatif yang lemah atau tidak ada korelasi sama sekali. Penelitian lebih lanjut disarankan untuk meningkatkan keragaman data dan pendekatan metodologis, termasuk penggunaan algoritma lain dan metode validasi eksternal. Berdasarkan nilai *p_value* menunjukkan bahwa *p_value* yang lebih besar dari 0,05 mengindikasikan bahwa korelasi yang diamati tidak signifikan secara statistik, sehingga tidak ada cukup bukti untuk menyatakan bahwa ada hubungan linear yang nyata antara "Persentase positif" dan "Rata-Rata Survei". Analisis sentimen dilakukan menggunakan algoritma *Support Vector Machine* (SVM), dengan menggunakan *Indonesian RoBERTa Base Sentiment Classifier* untuk pelabelan. Langkah-langkah klasifikasi meliputi pengumpulan data, pelabelan, *preprocessing teks*, dan pemberian bobot kata. Model SVM yang dioptimalkan menggunakan *k-fold cross-validation* (5) dengan parameter terbaik ($C=1$, $\Gamma=1$ untuk kernel *RBF*) mencapai akurasi sebesar 84,50%, presisi 83,46%, *recall* 82,34%, dan *F1-score* 82,84%.

Kata kunci : Analisis Sentimen, *Support Vector Machine*, Indonesian RoBERTa, Pemilihan Presiden 2024, Korelasi Survei

ABSTRACT

**“CORRELATION LABELING TEST USING THE INDONESIAN
ROBERTA BASE ON ATTRIBUTE TESTS OF SURVEY INSTITUTIONS
RESULTS AND PRESIDENTIAL ELECTION SENTIMENT
CLASSIFICATION USING THE SVM ALGORITHM”**

Glorious Aldo Putra Jatmiko

Universitas Sanata Dharma

2024

This research aims to explore the correlation between public sentiment on Instagram social media and survey results for the 2024 Indonesian presidential election. The correlation analysis between labeled sentiment data and survey results shows significant variations across different periods. Some periods exhibit strong positive correlations, while others show weak negative correlations or no correlation at all. Further research is recommended to increase data diversity and methodological approaches, including the use of other algorithms and external validation methods. Based on the p-value, it indicates that a p-value greater than 0.05 suggests that the observed correlation is not statistically significant, thus there is not enough evidence to claim a real linear relationship between "Positive Percentage" and "Survey Average." Sentiment analysis was conducted using the Support Vector Machine (SVM) algorithm, with Indonesian RoBERTa Base Sentiment Classifier for labeling. The classification steps included data collection, labeling, text preprocessing, and word weighting. The SVM model optimized using k-fold cross-validation (5) with the best parameters (C=1, Gamma=1 for RBF kernel) achieved an accuracy of 84.50%, precision of 83.46%, recall of 82.34%, and F1-score of 82.84%.

Keywords : Sentiment Analysis, Support Vector Machine, Indonesian RoBERTa, 2024 Presidential Election, Survey Correlation