

ABSTRAK

Pemindahan Ibu Kota Negara ke Nusantara (IKN) yang menjadi topik hangat oleh masyarakat di media sosial TikTok, Instagram, dan YouTube. Dalam penelitian ini, dilakukan analisis sentimen mengenai relokasi IKN dengan menggunakan algoritma *Random Forest Classifier*. Untuk meningkatkan akurasi model, penelitian ini menerapkan koreksi ejaan dengan metode *Damerau-Levenshtein Distance* (DLD) serta tokenisasi *N-Gram* dengan $n=2$ (*Bigram*). Dataset penelitian terdiri dari 1634 komentar diperoleh melalui proses *scraping*. Pelabelan data menunjukkan 966 data berlabel positif dan 668 data berlabel negatif, yang didapatkan dari proses *majority voting* dan validasi pakar bahasa. Penelitian ini lebih sistematis dengan digunakannya kerangka kerja CRISP-DM. Model *Random Forest Classifier* diterapkan untuk mengklasifikasikan sentimen, dengan seleksi fitur menggunakan *Information Gain* untuk mengatasi *overfitting*. Temuan pada penelitian ini menunjukkan kombinasi metode yang digunakan menghasilkan akurasi tertinggi sebesar 90,21%, sedangkan tanpa penerapan koreksi ejaan dan *N-Gram*, akurasi menurun menjadi 89,69% pada tiap subset *balancing data* yang paling optimal yakni *Random OverSampling* dibandingkan dengan SMOTE dan *Random UnderSampling*. Akan tetapi, DLD terdeteksi mengalami *margin of error* sebesar 8,62% dan berhasil dengan benar sebesar 91,38%. Dengan demikian, penelitian ini menunjukkan bahwa kombinasi metode koreksi ejaan, ekstraksi fitur, serta seleksi fitur yang digunakan mampu memperoleh performa optimal dalam analisis sentimen terkait relokasi IKN.

Kata Kunci: Analisis Sentimen, Ibu Kota Nusantara (IKN), *Damerau-Levenshtein Distance*, *N-Gram*, *Random Forest*

ABSTRACT

The relocation of Indonesia's capital city to Nusantara (IKN) has become a hot topic among the public on social media platforms such as TikTok, Instagram, and YouTube. This study conducts a sentiment analysis regarding the IKN relocation using the Random Forest Classifier algorithm. To enhance the model's accuracy, the research applies spelling correction with the Damerau-Levenshtein Distance (DLD) method and N-Gram tokenization with $n=2$ (Bigram). The research dataset consists of 1,634 comments obtained through web scraping. Data labeling indicates 966 positive and 668 negative samples, derived from a majority voting process and validation by language experts. The study adopts a more systematic approach using the CRISP-DM framework. The Random Forest Classifier model is applied to classify sentiments, with feature selection using Information Gain to address overfitting. The findings of this study reveal that the combination of methods used achieved the highest accuracy of 90.21%, while without the implementation of spelling correction and N-Gram, the accuracy dropped to 89.69% on the most optimal data balancing subset, which is Random OverSampling compared to SMOTE and Random UnderSampling. However, DLD exhibited a margin of error of 8.62% and a success rate of 91.38%. Therefore, this study demonstrates that the combination of spelling correction, feature extraction, and feature selection methods employed can achieve optimal performance in sentiment analysis related to the IKN relocation.

Keywords: *Sentiment Analysis, Capital City Nusantara (IKN), Damerau-Levenshtein Distance, N-Gram, Random Forest*