

ABSTRAK

Media sosial X menjadi ruang publik bagi masyarakat dalam menyampaikan opini terhadap RUU Perampasan Aset. Penelitian ini menganalisis sentimen publik terhadap RUU tersebut menggunakan model *Bidirectional Long Short-Term Memory* (BiLSTM) berbasis *Word2Vec*, serta mengidentifikasi topik dominan pada tiap kategori sentimen menggunakan *Latent Dirichlet Allocation* (LDA). Metodologi mengacu pada CRISP-DM, meliputi *crawling* 9.651 cuitan dengan rentang waktu Januari 2025 s.d. Oktober 2025 yang menghasilkan 9612 data final setelah tahapan seleksi, *preprocessing*, pelabelan otomatis ke dalam dua kelas sentimen (positif dan negatif) berbasis leksikon *SentiWordNet*, ekstraksi fitur *Word2Vec* (CBOW dan *Skip-gram*), penyeimbangan kelas dengan SMOTE, serta pembagian data 80:20. Delapan skenario diuji untuk membandingkan pengaruh *stemming* dan *stopword removal* serta arsitektur *Word2Vec*. Hasil menunjukkan skenario terbaik adalah *Skip-gram* dengan *stemming* tanpa *stopword removal*, mencapai akurasi 85,91%, presisi 89,82%, *recall* 92,50%, dan *F1-score* 91,14%. *Stemming* meningkatkan kinerja, sedangkan *stopword removal* tidak selalu menunjukkan hasil yang optimal. *Skip-gram* unggul dibanding CBOW. LDA mengidentifikasi tiga topik dominan pada sentimen positif dan sembilan topik pada sentimen negatif. Model terbaik diimplementasikan ke aplikasi *Streamlit* untuk analisis sentimen dan topik secara langsung.

Kata kunci: Analisis Sentimen, RUU Perampasan Aset, *SentiWordNet*, *Word2Vec*, BiLSTM, LDA, X (Twitter).

ABSTRACT

The social media platform X serves as a public space for citizens to express their opinions regarding the Asset Confiscation Bill. This study analyzes public sentiment toward the bill using a Word2Vec-based Bidirectional Long Short-Term Memory (BiLSTM) model, and identifies dominant topics within each sentiment category using Latent Dirichlet Allocation (LDA). The methodology is based on the CRISP-DM framework, involving the crawling of 9,651 tweets from January to October 2025, which yielded 9,612 final data points after the selection phase. Subsequent steps included preprocessing, automatic labeling into two sentiment classes (positive and negative) based on the SentiWordNet lexicon, Word2Vec feature extraction (CBOW and Skip-gram), class balancing using SMOTE, and an 80:20 data split. Eight scenarios were tested to compare the effects of stemming and stopword removal, as well as the Word2Vec architectures. The results demonstrated that the best scenario was Skip-gram with stemming but without stopword removal, achieving an accuracy of 85.91%, a precision of 89.82%, a recall of 92.50%, and an F1-score of 91.14%. Stemming improved model performance, whereas stopword removal did not consistently yield optimal results. Furthermore, Skip-gram outperformed CBOW. The LDA identified three dominant topics within the positive sentiment and nine topics within the negative sentiment. The best-performing model was implemented into a Streamlit application for real-time sentiment and topic analysis.

Keywords: Sentiment Analysis, Asset Confiscation Bill, SentiWordNet, Word2Vec, BiLSTM, LDA, X (Twitter).