

ABSTRAK

Opini publik di media sosial berkembang pesat dan dapat membentuk persepsi masyarakat terhadap kebijakan pemerintah. Pembentukan Badan Pengelola Investasi Daya Anagata Nusantara (Danantara) memunculkan respons pro dan kontra di platform X, sehingga diperlukan analisis sentimen untuk memetakan kecenderungan dukungan maupun penolakan secara objektif. Penelitian ini menggunakan model IndoBERT (indobenchmark/indobert-base-p1) untuk mengklasifikasikan sentimen publik berbahasa Indonesia secara biner (positif dan negatif). Dataset terdiri dari 3.796 cuitan periode Mei–Agustus 2025 yang dilabeli secara manual dan dibagi menggunakan *stratified split* 80% data latih, 10% validasi, dan 10% uji. Proses *fine-tuning* dilakukan pada 10 skenario yang mencakup kondisi baseline serta variasi rasio distribusi kelas (50:50, 70:30, 80:20, 90:10), baik tanpa maupun dengan penerapan SMOTE sebagai teknik penyeimbangan data. Hasil penelitian menunjukkan bahwa kemampuan representasional IndoBERT menjadi faktor utama keberhasilan klasifikasi. Pada kondisi baseline tanpa SMOTE, model mencapai F1-Score = 0,9571, *Accuracy* = 0.9526, *Precision* = 0.9757 dan *Recall* = 0.9392. Hal ini, menunjukkan bahwa IndoBERT sudah sangat kuat bahkan tanpa penyeimbangan data tambahan. Penerapan SMOTE memberikan pengaruh yang bervariasi, tidak berdampak pada data seimbang (50:50), meningkatkan performa pada ketidakseimbangan sedang (70:30 dan 80:20), namun menurunkan performa pada ketidakseimbangan ekstrem (90:10).

Kata kunci: Analisis Sentimen, IndoBERT, Danantara, SMOTE, *Text Mining*, X

ABSTRACT

Public opinion on social media evolves rapidly and can shape societal perceptions of government policies. The establishment of the Daya Anagata Nusantara Investment Management Agency (Danantara) has generated both support and opposition on the X platform, making sentiment analysis necessary to objectively map tendencies of approval and rejection. This study employs the IndoBERT model (indobenchmark/indobert-base-p1) to classify Indonesian-language public sentiment into binary categories (positive and negative). The dataset consists of 3,796 tweets collected from May to August 2025, which were manually labeled and split using a stratified approach into 80% training data, 10% validation data, and 10% testing data. The fine-tuning process was conducted across 10 scenarios, including a baseline condition and variations in class distribution ratios (50:50, 70:30, 80:20, 90:10), both without and with the application of SMOTE as a data balancing technique. The results indicate that the representational capability of IndoBERT is the primary factor contributing to classification performance. Under the baseline condition without SMOTE, the model achieved an F1-Score = 0,9571, Accuracy = 0.9526, Precision = 0.9757 and Recall= 0.9392. This demonstrates that IndoBERT is already highly effective even without additional data balancing. The application of SMOTE shows varying effects: it has no impact on balanced data (50:50), improves performance under moderate imbalance (70:30 and 80:20), but decreases performance under extreme imbalance (90:10).

Keywords: *Sentiment Analysis, IndoBERT, Danantara, SMOTE, Text Mining, X*