

ABSTRAK

ANALISIS PERFORMA *TRANSFORMER-BASED MODEL* PADA DETEKSI *BLACK CAMPAIGN* (Studi Kasus: Pemilihan Presiden Indonesia 2024)

Oleh

JESSEN RAMADEKSA ALLEN

Meningkatnya penggunaan media sosial pada Pemilihan Presiden Indonesia 2024 memperluas ruang komunikasi politik sekaligus memunculkan praktik *black campaign* yang berpotensi merusak reputasi kandidat dan menyesatkan opini publik. Pendekatan klasifikasi teks yang digunakan pada studi awal masih memiliki keterbatasan dalam menangkap hubungan semantik yang kompleks pada teks politik, sehingga diperlukan metode yang mampu memahami representasi bahasa secara lebih mendalam. Penelitian ini bertujuan untuk menganalisis performa model berbasis *Transformer* dalam mendeteksi konten *black campaign* pada data Twitter terkait Pemilihan Presiden Indonesia 2024. Dataset yang digunakan berjumlah 7.025 data dengan dua kelas, yaitu *Black Campaign* dan *Non-Black Campaign*.

Tahapan penelitian meliputi *preprocessing* teks, penyeimbangan data menggunakan *oversampling* dan *downsampling*, *splitting*, proses *hyperparameter tuning* dan *fine-tuning*, serta evaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Model yang diuji meliputi *IndoBERT*, *IndoBERTweet*, dan *XLM-RoBERTa*. Hasil penelitian menunjukkan bahwa seluruh model *Transformer* mampu melampaui performa model *benchmark*, dengan performa terbaik diperoleh oleh *IndoBERT* menggunakan pendekatan *oversampling* dan normalisasi teks yang mencapai akurasi 95,45%. Pengujian pada dataset eksternal menunjukkan kemampuan model dalam menghadapi variasi distribusi data dengan akurasi sekitar 50%. Secara keseluruhan, arsitektur *Transformer* menunjukkan kemampuan yang lebih baik dalam memahami representasi semantik teks politik dibandingkan model sekuensial.

Kata Kunci: *Black Campaign*, *Transformer-based Model*, Klasifikasi Teks, *Natural Language Processing*, Media Sosial, Pemilihan Presiden 2024, *IndoBERT*.

ABSTRACT

PERFORMANCE ANALYSIS OF TRANSFORMER-BASED MODELS IN BLACK CAMPAIGN DETECTION (Case Study: The 2024 Indonesian Presidential Election)

By

JESSEN RAMADEKSA ALLEN

The increasing use of social media during the 2024 Indonesian Presidential Election has expanded the space for political communication while also enabling the spread of black campaign content that may damage candidates reputations and mislead public opinion. Early text classification approaches still face limitations in capturing complex semantic relationships within political texts, highlighting the need for methods capable of learning deeper language representations. This study aims to analyze the performance of Transformer-based models in detecting black campaign content on Twitter data related to the 2024 Indonesian Presidential Election. The dataset used consists of 7,025 labeled instances with two classes: Black Campaign and Non-Black Campaign.

The research stages include text preprocessing, data balancing using oversampling and downsampling, data splitting, hyperparameter tuning, and fine-tuning processes, followed by evaluation using Accuracy, Precision, Recall, and F1-score metrics. The models evaluated in this study include IndoBERT, IndoBERTweet, and XLM-RoBERTa. The results show that all Transformer models outperform the benchmark model, with the best performance achieved by IndoBERT using oversampling and text normalization, reaching an accuracy of 95.45%. Evaluation on an external dataset further provides insight into the model's ability to handle variations in data distribution, achieving an accuracy of approximately 50%. Overall, the Transformer architecture demonstrates a stronger capability in capturing the semantic representation of political texts compared to sequential models.

Keywords: Black Campaign, Transformer-based Model, Text Classification, Natural Language Processing, Social Media, 2024 Presidential Election, IndoBERT.