

## ABSTRAK

Penyebaran berita palsu pada *platform* berita *online* telah menjadi tantangan serius yang mengancam integritas informasi dan stabilitas sosial di era digital. Penelitian ini bertujuan untuk mengembangkan model deteksi berita palsu yang akurat menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT) dengan pendekatan *fine-tuning* pada *dataset* berita berbahasa Indonesia. *Dataset* penelitian terdiri dari 248 artikel berita yang dikurasi dari berbagai *platform* media *online*, kemudian diseimbangkan menjadi 196 artikel dengan distribusi 50:50 antara berita asli dan berita palsu. Metodologi penelitian mencakup *text preprocessing*, tokenisasi menggunakan IndoBERT *tokenizer* dengan panjang maksimal 128 token, dan pembagian data dengan *stratified train-test split* 80:20 menghasilkan 156 sampel *training* dan 40 sampel *testing*. Model IndoBERT di-*fine-tune* selama 3 *epoch* dengan konfigurasi *batch size* 4, *learning rate* 2e-05, *gradient accumulation steps* 2, dan *optimizer* AdamW. Hasil penelitian menunjukkan performa yang baik dengan *accuracy* sebesar 87,50%, *Macro F1-Score* 87,43%, *Sensitivity* 95,00%, dan *Specificity* 80,00%. Model menunjukkan *Precision* 0,9412 dan *Recall* 0,8000 untuk deteksi berita palsu (*FAKE*), serta *Precision* 0,8261 dan *Recall* 0,9500 untuk berita asli (*REAL*). Validasi menggunakan *confusion matrix* menunjukkan kemampuan model dalam mengidentifikasi indikator linguistik berita palsu dengan karakteristik model yang cenderung konservatif dalam melabeli berita sebagai palsu sehingga meminimalkan sensor berlebihan terhadap berita *legitimate*. Penelitian ini membuktikan bahwa metode BERT efektif untuk deteksi berita palsu berbahasa Indonesia dan dapat diimplementasikan sebagai sistem moderasi konten otomatis pada *platform* berita *online*.

**Kata Kunci:** BERT, deteksi berita palsu, *natural language processing*, klasifikasi teks, *deep learning*.

## **ABSTRACT**

*The spread of fake news on online news platforms has become a serious challenge threatening information integrity and social stability in the digital era. This study aims to develop an accurate fake news detection model using the Bidirectional Encoder Representations from Transformers (BERT) method with a fine-tuning approach on an Indonesian-language news dataset. The research dataset consists of 248 news articles curated from various online media platforms, subsequently balanced to 196 articles with a 50:50 distribution between real and fake news. The research methodology includes text preprocessing, tokenization using the IndoBERT tokenizer with a maximum length of 128 tokens, and data splitting using stratified train-test split 80:20, resulting in 156 training samples and 40 testing samples. The IndoBERT model was fine-tuned for 3 epochs with a configuration of batch size 4, learning rate  $2e-05$ , gradient accumulation steps 2, and AdamW optimizer. The results demonstrate strong model performance with an accuracy of 87.50%, Macro F1-Score of 87.43%, Sensitivity of 95.00%, and Specificity of 80.00%. The model achieved a Precision of 0.9412 and Recall of 0.8000 for fake news detection (FAKE), and a Precision of 0.8261 and Recall of 0.9500 for real news (REAL). Validation using a confusion matrix demonstrates the model's capability in identifying linguistic indicators of fake news, with the model exhibiting conservative characteristics in labeling news as fake, thereby minimizing over-censorship of legitimate news. This study proves that the BERT method is effective for fake news detection in the Indonesian language and can be implemented as an automated content moderation system on online news platforms.*

**Keywords:** *BERT, fake news detection, natural language processing, text classification, deep learning.*