ABSTRAK

Muchtar, Nabillah Ufairoh. 2025. Analisis Kinerja Transformer Untuk *Named Entity Recognition* (NER) Menggunakan Indobert Pada Transkrip Video Politik Berbahasa Indonesia. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember

Pembimbing: Deni Arifianto, M,Kom.; Dr. Reni Umilasari, S.Pd., M.Si.

Penelitian ini menganalisis kinerja *Named Entity Recognition* (NER) berbasis IndoBERT dalam mengidentifikasi entitas pada transkrip pidato politik Presiden Republik Indonesia. Proses penelitian mencakup transkripsi video menjadi teks, *preprocessing*, pelabelan entitas dengan format BIO, serta pelatihan model melalui *fine-tuning*. Evaluasi kinerja dilakukan menggunakan beberapa skema *k-fold cross validation* yaitu 2-*Fold*, 3-*Fold*, 4-*Fold*, dan 6-*Fold* untuk memastikan reliabilitas model. Indikator evaluasi yang digunakan meliputi *precision*, *recall*, *F1-score*, serta *confusion matrix*. Hasil penelitian menunjukkan bahwa IndoBERT mampu memberikan performa yang konsisten dengan nilai *F1-score* yang stabil di seluruh skema validasi, meskipun terdapat variasi kecil pada masing-masing *fold*. Hal ini menegaskan bahwa IndoBERT efektif dalam mengidentifikasi entitas penting (PER, ORG, LOC, DATE, CRD) pada teks politik berbahasa Indonesia. Penelitian ini memberikan kontribusi pada pengembangan teknologi NLP untuk mendukung analisis wacana politik serta potensi implementasi dalam sistem pemantauan konten digital.

Kata kunci: *Named Entity Recognition*, IndoBERT, Pidato Politik, *Transformer*, *Cross Validation*

ABSTRACT

Muchtar, Nabillah Ufairoh. 2025. *Transformer Performance Analysis for Named Entity Recognition (NER) Using Indobert on Indonesian Political Video Transcripts. Final Project. Undergraduate Program, Informatics Engineering Study Program, Muhammadiyah University of Jember.*
Advisors: Deni Arifianto, M.Kom.; Dr. Reni Umilasari, S.Pd., M.Si.

This research analyzes the performance of the IndoBERT-based Named Entity Recognition (NER) model in identifying entities within political speech transcripts of the President of the Republic of Indonesia. The process involved converting videos into text, preprocessing, entity labeling using the BIO format, and fine-tuning the model. Performance evaluation was conducted using multiple k-fold cross validation schemes, namely 2-Fold, 3-Fold, 4-Fold, and 6-Fold, to ensure the model's reliability. Evaluation metrics included precision, recall, F1-score, and confusion matrix. The results indicate that IndoBERT achieved consistent and stable performance across all validation schemes, with slight variations between folds. These findings highlight IndoBERT's effectiveness in recognizing political entities (PER, ORG, LOC, DATE, CRD) in Indonesian texts. This study contributes to the advancement of NLP technologies for political discourse analysis and their potential application in digital content monitoring systems.

Keywords: *Named Entity Recognition, IndoBERT, Political Speech, Transformer, Cross Validation*