

ABSTRAK

PEMODELAN PENILAIAN ESAI OTOMATIS BAHASA INDONESIA MENGUNAKAN INDOBERT DAN ARSITEKTUR BILSTM- ATTENTION

LUTHFI ARIE ZULFIKRI

H1D022061

Proses penilaian esai secara manual kerap membutuhkan waktu yang lama dan memiliki kerentanan terhadap bias subjektivitas penilai. Di sisi lain, upaya pengembangan *Automated Essay Scoring* (AES) untuk bahasa Indonesia masih terkendala oleh minimnya korpus berlabel serta tingginya kompleksitas linguistik yang khas. Untuk mengatasi permasalahan tersebut, penelitian ini bertujuan merancang sekaligus mengevaluasi keandalan model AES berbahasa Indonesia. Pendekatan yang diusulkan adalah arsitektur hibrida yang mengintegrasikan IndoBERT, *Bidirectional Long Short-Term Memory* (BiLSTM), dan mekanisme *Self-Attention*. Proses pelatihan dan pengujian model memanfaatkan *dataset* PERSUADE 2.0 yang telah dialihbahasakan secara presisi ke dalam bahasa Indonesia menggunakan arsitektur *machine translation* NLLB-200. Pada tahap pelatihan, penelitian ini mengimplementasikan fungsi *Weighted Smooth L1 Loss* guna memitigasi masalah ketimpangan kelas. Langkah tersebut kemudian dipadukan dengan teknik optimasi ambang batas berbasis Nelder-Mead untuk memetakan prediksi regresi yang bersifat kontinu ke dalam ruang ordinal. Hasil evaluasi menunjukkan bahwa integrasi arsitektur IndoBERT-BiLSTM-Attention mampu menghasilkan performa yang optimal. Kinerja ini dibuktikan dengan perolehan metrik *Quadratic Weighted Kappa* (QWK) sebesar 0,8508, *Exact Accuracy* sebesar 63,27%, dan *Mean Absolute Error* (MAE) sebesar 0,4277. Selain itu, metode kalibrasi kelas yang diaplikasikan terbukti krusial dalam meningkatkan presisi pada kelompok skor minoritas ekstrem hingga mencapai 18,37%. Peningkatan tersebut didukung oleh tingkat stabilitas kinerja yang telah divalidasi menggunakan *Bootstrap Confidence Intervals* 95% pada rentang nilai [0,8426; 0,8586]. Temuan dari penelitian ini memberikan kontribusi metodologis yang signifikan terhadap kemajuan sistem AES berbahasa Indonesia. Model yang dikembangkan memiliki potensi untuk diimplementasikan sebagai instrumen penunjang evaluasi pendidikan yang lebih efisien, terukur, dan objektif.

Kata Kunci: *Automated Essay Scoring*, Bahasa Indonesia, BiLSTM, IndoBERT, Mekanisme *Attention*, *Weighted Huber Loss*.

ABSTRACT

A HYBRID INDOBERT AND BILSTM-ATTENTION ARCHITECTURE FOR INDONESIAN AUTOMATED ESSAY SCORING

LUTHFI ARIE ZULFIKRI

H1D022061

Manual essay grading is inherently time-consuming and highly susceptible to subjective evaluator bias. The development of Automated Essay Scoring (AES) systems for the Indonesian language, however, remains constrained by a scarcity of labeled corpora and unique linguistic complexities. To address these limitations, this study designs and evaluates a robust Indonesian AES model employing a hybrid architecture that integrates IndoBERT, Bidirectional Long Short-Term Memory (BiLSTM), and a Self-Attention mechanism. The model was trained and evaluated using the PERSUADE 2.0 dataset, which was precisely translated into Indonesian utilizing the NLLB-200 machine translation architecture. To mitigate class imbalance during the training phase, a Weighted Smooth L1 Loss function was implemented. This approach was subsequently combined with a Nelder-Mead threshold optimization technique to effectively map continuous regression predictions into an ordinal space. Empirical evaluations demonstrate that the integrated IndoBERT-BiLSTM-Attention architecture achieves optimal performance, yielding a Quadratic Weighted Kappa (QWK) of 0.8508, an Exact Accuracy of 63.27%, and a Mean Absolute Error (MAE) of 0.4277. Furthermore, the applied class calibration method proved critical in enhancing precision within extreme minority score groups by up to 18.37%. This improvement is corroborated by a high degree of performance stability, validated through 95% Bootstrap Confidence Intervals within the [0.8426, 0.8586] range. These findings offer a significant methodological contribution to the advancement of Indonesian AES systems, positioning the developed model as a highly efficient, scalable, and objective instrument for educational assessment.

Keywords: *Attention Mechanism; Automated Essay Scoring; BiLSTM; IndoBERT; Indonesian Language; Weighted Huber Loss*