

ABSTRACT

Manual essay grading in academic settings often faces issues of subjectivity, inconsistency, and high time demands on teachers. The objective of this study is to develop a novel automated essay grading model by combining Cosine Similarity and Model (LLM) to evaluate student responses. The cosine similarity is employed to assess the lexical similarity between the student responses and the answer keys, while the LLM is utilized for a more profound semantic analysis, encompassing argument structure, cohesion, and the extent of analysis. The research process involves a series of steps, including text preprocessing (tokenization, capitalization, stop word removal, and lemmatization) and Term Frequency (TF) weighting on an Indonesian essay dataset. The evaluation was conducted in three distinct scenarios: The models considered include the cosine similarity model, the LLM model, and a hybrid model. The evaluation results indicate that Cosine Similarity exhibits a Mean Absolute Percentage Error (MAPE) of 19.08%, while LLM demonstrates superior performance with a MAPE of 7.85%. The hybrid model, which integrates Cosine Similarity (40%) and LLM (60%), achieved the most optimal accuracy with a MAPE of 5.49%. This hybrid approach was found to improve accuracy by 71.2% compared to the initial method and demonstrated a consistent decrease in MAPE across all models, reflecting significant improvements not only in accuracy but also in evaluation consistency. These findings indicate that the integration of both methods significantly enhances the effectiveness of the automatic essay scoring system, thereby rendering it a more objective, efficient, and reliable alternative to manual scoring in the context of academic evaluation.

Keywords: Automated Essay Scoring, Cosine Similarity, Large Language Model (LLM), Mean Absolute Percentage Error (MAPE), Text Preprocessing

ABSTRAK

Penilaian esai manual di lingkungan akademik seringkali menghadapi masalah subjektivitas, inkonsistensi, dan beban waktu yang tinggi bagi pengajar. Penelitian ini bertujuan mengembangkan model penilaian esai otomatis inovatif dengan mengombinasikan *Cosine Similarity* dan *Large Language Model* (LLM) untuk mengevaluasi jawaban mahasiswa. *Cosine Similarity* digunakan untuk mengukur kesamaan leksikal antara jawaban mahasiswa dan kunci jawaban, sedangkan LLM dimanfaatkan untuk analisis semantik yang lebih mendalam, mencakup struktur argumen, kohesi, dan kedalaman analisis. Proses penelitian meliputi *text preprocessing* (*tokenization, lower casing, stopword removal, lemmatization*) dan pembobotan *Term Frequency* (TF) pada *dataset* esai berbahasa Indonesia. Penilaian dilakukan dalam tiga skenario: *Cosine Similarity* saja, LLM saja, dan model hibrida. Hasil evaluasi menunjukkan *Cosine Similarity* memiliki *Mean Absolute Percentage Error* (MAPE) 19,08%, sementara LLM menunjukkan performa lebih baik dengan MAPE 7,85%. Model hibrida, yang mengombinasikan *Cosine Similarity* (40%) dan LLM (60%), mencapai akurasi paling optimal yaitu dengan MAPE 5,49%. Pendekatan hibrid ini terbukti meningkatkan akurasi sebesar 71,2% dibandingkan metode awal dan menunjukkan penurunan MAPE yang konsisten dari setiap model, mencerminkan perbaikan signifikan tidak hanya dalam akurasi, tetapi juga konsistensi penilaian. Temuan ini menegaskan bahwa integrasi kedua metode secara signifikan meningkatkan efektivitas sistem penilaian esai otomatis, menjadikannya alternatif yang lebih objektif, efisien, dan andal dibandingkan penilaian manual dalam konteks evaluasi akademik.

Kata Kunci: Penilaian Esai Otomatis, *Cosine Similarity*, *Large Language Model* (LLM), *Text Preprocessing*, *Mean Absolute Percentage Error* (MAPE)