

ABSTRAK

Evaluasi jawaban uraian singkat secara manual rentan terhadap subjektivitas dan memerlukan waktu yang relatif lama, sehingga mendorong pengembangan sistem *Automatic Short Answer Grading* (ASAG) yang mampu melakukan penilaian secara otomatis dan konsisten. Meskipun berbagai pendekatan telah dikembangkan, masih terbatas penelitian yang membandingkan secara langsung pendekatan *Semantic Embeddings* berbasis Sentence-BERT (SBERT) dengan pendekatan *Generative Large Language Models* (LLMs) pada tipe soal yang berbeda sekaligus mempertimbangkan efisiensi operasional. Penelitian ini membandingkan enam metode ASAG, yaitu SBERT *zero-shot*, *Fine-tuned* SBERT, GPT-5.4-nano *zero-shot*, GPT-5.4-nano *few-shot*, GPT-5.4 *zero-shot*, dan GPT-5.4 *few-shot* menggunakan *dataset* Mohler ASAG yang dipisahkan menjadi *subset Open-Ended* (OE) dan *Close-Ended* (CE). Pada pendekatan SBERT, model all-distilroberta-v1 di-*fine-tune* menggunakan CosineSimilarityLoss pada *training set* yang telah diseimbangkan melalui augmentasi berbasis LLM, sedangkan pendekatan *Generative LLMs* memanfaatkan pemanggilan API GPT dengan strategi *zero-shot* dan *few-shot prompting*. Evaluasi dilakukan menggunakan *F1-Score*, *Accuracy*, *Precision*, dan *Recall*, serta waktu inferensi dan estimasi biaya API sebagai ukuran efisiensi operasional. Hasil penelitian menunjukkan bahwa *Fine-tuned* SBERT memperoleh *F1-Score* tertinggi pada *subset* OE sebesar 0,9241, sedangkan GPT-5.4 *few-shot* mencapai performa terbaik pada *subset* CE dengan *F1-Score* sebesar 0,9750. Selain itu, *fine-tuning* secara konsisten meningkatkan performa SBERT, sedangkan efektivitas *few-shot prompting* bergantung pada kapasitas model yang digunakan. Penelitian ini menyimpulkan bahwa tidak terdapat satu metode yang unggul secara universal untuk seluruh tipe soal ASAG. *Fine-tuned* SBERT lebih sesuai untuk soal *Open-Ended*, sedangkan GPT-5.4 *few-shot* lebih efektif pada soal *Close-Ended*, sehingga pemilihan metode sebaiknya disesuaikan dengan karakteristik soal dan kebutuhan efisiensi operasional.

Kata kunci : *Automated Short Answer Grading*, Sentence-BERT, *Generative Pre-trained Transformer*, klasifikasi biner, *fine-tuning*, *prompt engineering*