

ABSTRACT

Learning evaluation was a critical aspect of education to measure student understanding and enhance the quality of educational processes. One of the main challenges was providing useful feedback on subjective explanations of student answer deficiencies, which required high educator subjectivity. Appropriate and constructive feedback could improve students' grasp of concepts, learning materials, and critical thinking skills. To address this, Natural Language Processing (NLP) technology with text generation approaches, such as the BART-Large architecture, was adopted to automate evaluations. However, large architectures like BART-Large required specialized training on large-scale datasets, making them less efficient for small-scale research. Therefore, this study proposed developing an Automatic Feedback model for Indonesian short-answer essay questions using the DistilBART architecture, a lightweight, distilled version of BART that retained the original model's performance. Experiments were conducted to train the feedback model with various configurations of similarity thresholds, batch sizes, and learning rates to identify the optimal setup for generating accurate and relevant feedback. When comparing DistilBART with BART-Large, results demonstrated that DistilBART was more efficient in inference and delivered higher feedback accuracy. Evaluations showed that DistilBART outperformed BART-Large in BLEU (0.1650) and METEOR (0.4166) metrics, while BART-Large achieved higher performance in ROUGE-L (0.4430) on the Sistem Cerdas dataset. Similarly, on the MPI dataset, these findings confirmed that DistilBART was more accurate and efficient than BART-Large in generating automated feedback for Indonesian short-answer essay responses.

Keywords : Short Answer Questions, Automatic Feedback, Architecture DistilBART, Architecture BART-Large, BLEU, ROUGE, METEOR