ABSTRACT

The short answer question is considered an appropriate type of exam question for evaluating cognitive, understanding, and knowledge. However, the answers generated are very varied, making manual grading time-consuming and lead to bias or inconsistency. Automated Short Answer Grading (ASAG) is the assessment of short answers using computational techniques that output a score or cognitive category. ASAG input in the form of questions and answers helps to recognize varied correct answers based on the relevance of context or semantic attributes of two text representations. The current popular text representation method in ASAG, Siamese-BERT sentence embedding, is able to produce fixed-size semantic vectors. The representation of questions and answers from this method can be assessed for their relevance with the matching matrix method and can be understood for their context with BiLSTM. Therefore, ASAG research was conducted using matching matrix and BiLSTM on the question and answer representation produced by Siamese-BERT sentence embedding. IndoSBERT, an Indonesian Siamese-BERT model, was fine-tuned and used to represent questions and answers from four Indonesian short answer datasets. Four ASAG model architectures were built to determine the effect of matching matrix and BiLSTM. The model without matching matrix with BiLSTM is the best model for closed-ended questions with analytical answers (Siscer dataset) with SMAPE 13.2262%. The model with matching matrix and BiLSTM is the best model for open-ended questions with analytical answers (MPI and Ukara B dataset) with SMAPE 14.2271% and 30.7918%. The model with matching matrix without BiLSTM is the best model for open-ended questions with non-analytical answers (Ukara A dataset) with SMAPE 19.7149%. Matching matrix was observed to have a positive effect on open-end questions, while BiLSTM was observed to have a positive effect on analytical answers.

Keywords : *Automated Short Answer Grading*, BiLSTM, *Matching Matrix*, SMAPE