ABSTRACT

Recommender systems play an important role in helping users discover relevant information amid the large number of options available on digital platforms. In the context of tourism, recommender systems are utilized to assist users in finding tourist destinations that match their preferences. A key challenge in content-based recommender systems is how to represent tourist destination information so that similarity between destinations can be measured accurately. Therefore, this study conducts a comparative analysis of text representation methods based on TF-IDF and IndoBERT, including both pre-trained IndoBERT and fine-tuned IndoBERT, in the application of Content-Based Filtering for tourist destination recommendation. TF-IDF is employed as a simple and computationally efficient frequency-based approach, while IndoBERT is an Indonesian Transformer-based language model capable of generating contextual representations. Text representations produced by both methods are used to construct feature vectors for each tourist destination, which are then compared using Cosine Similarity to obtain similarity scores. The performance of the recommender system is evaluated using the Normalized Discounted Cumulative Gain (NDCG) metric with graded relevance and binary relevance schemes. The experimental results indicate that fine-tuned IndoBERT with a learning rate of 5×10^{-5} achieves the best performance, obtaining Graded NDCG values of 0.727768 at a recommendation cutoff of five and 0.739136 at a cutoff of ten, as well as binary NDCG values of 0.626778 at a cutoff of five and 0.625513 at a cutoff of ten. In contrast, TF-IDF demonstrates the highest computational efficiency, achieving the fastest maximum execution time of 0.001468 seconds using the unigram configuration. Overall, this study shows that contextual embedding-based text representations such as IndoBERT are more effective in improving the quality of tourist destination recommendations, while TF-IDF remains a competitive and efficient baseline method.

Keywords: Recommendation System, Content-Based Filtering, Normalized Discounted Cumulative Gain, Cosine Similarity, TF-IDF, IndoBERT.