

ABSTRAK

Pencarian ayat Al-Qur'an yang relevan dengan kondisi emosional pengguna sering kali menjadi tantangan karena keterbatasan sistem pencarian berbasis *Semantic Matching* yang mampu memahami makna semantik yang lebih dalam. Penelitian ini bertujuan untuk membangun sistem pencocokan ayat Al-Qur'an berdasarkan ungkapan perasaan pengguna dengan pendekatan *Semantic Text Similarity* menggunakan model IndoBERT. Penelitian ini menggunakan dua dataset yang berbeda jumlahnya. Tahap pertama, model dilakukan *fine-tuning* menggunakan dataset sebanyak 3.813 data. Lalu kedua, dilakukan *fine-tuning* lagi menggunakan dataset sebanyak 4.137 data. Proses pelatihan dilakukan pada tiga skenario kombinasi *epoch* dan *batch size*, lalu dievaluasi menggunakan metrik Recall@1, @3, @5, Mean Reciprocal Rank (MRR), dan *cosine similarity*. Hasil penelitian menunjukkan bahwa semakin besar dan berkualitas dataset, maka semakin tinggi nilai akurasi. Skenario paling bagus ditunjukkan dari skenario ke-5 (*epoch* 45, *batch size* 32) memberikan performa terbaik dengan Recall@5 sebesar 0.8889 dan MRR sebesar 0.8564. dari penelitian yang telah dilakukan pengaruh kuantitas dan kualitas dataset, semakin ingginya *epoch* serta semakin kecilnya *batch size* mempengaruhi hasil kinerja dari suatu model. Hasil ini membuktikan bahwa pendekatan IndoBERT mampu menangkap keterkaitan semantik antara ekspresi emosional dan ayat Al-Qur'an dengan akurasi yang cukup baik. Sistem ini berpotensi menjadi alat bantu dakwah dan konsultasi Islami berbasis teknologi.

Kata Kunci: IndoBERT, Recall@K, MRR, *Cosine Similarity*, *Semantic Text Similarity*



ABSTRACT

Finding Qur'anic verses that are relevant to a user's emotional condition is often challenging due to the limitations of Semantic Matching systems in understanding deeper semantic meanings. This study aims to develop a verse-matching system based on users' emotional expressions using a Semantic Text Similarity approach with the IndoBERT model. Two datasets of different sizes were used in this research. In the first phase, the model was fine-tuned using a dataset of 3,813 pairs. In the second phase, additional fine-tuning was performed using 4,137 pairs. The training process involved three different combinations of epochs and batch sizes, and the model was evaluated using Recall@1, @3, @5, Mean Reciprocal Rank (MRR), and cosine similarity metrics. The results showed that larger and higher-quality datasets contribute to improved accuracy. The best performance was achieved in scenario 5 (45 epochs, batch size 32), with a Recall@5 of 0.8889 and an MRR of 0.8564. The findings indicate that both the quantity and quality of the dataset, as well as increased epochs and smaller batch sizes, significantly influence model performance. These results demonstrate that the IndoBERT-based approach can effectively capture the semantic relationship between emotional expressions and Qur'anic verses. This system holds potential as a technological tool for Islamic preaching and counseling.

Keywords: IndoBERT, Recall@K, MRR, *Cosine Similarity*, *Semantic Text Similarity*.

