

ABSTRAK

Penelitian ini mengembangkan perangkat bantu deteksi kemiripan naskah dengan mengintegrasikan algoritma *Bidirectional Encoder Representations from Transformers* (BERT) dan *cosine similarity*. Sistem ini dirancang untuk meningkatkan akurasi dalam mendeteksi kemiripan teks, termasuk pada teks parafrase dan terjemahan lintas bahasa. Hasil penelitian menunjukkan bahwa kombinasi BERT dan *cosine similarity* memberikan performa yang baik dengan F1 score sebesar 0,83 dan *Mean Absolute Error* (MAE) sebesar 0,139. Sistem ini mampu mendeteksi kemiripan semantik dengan akurasi yang cukup tinggi, mencapai skor similarity hingga 1,0 pada teks identik dan 0,97 pada teks parafrase. Selain itu, sistem ini menunjukkan fleksibilitas dalam mendeteksi kemiripan konten terjemahan dalam berbagai bahasa. Penelitian ini membuktikan bahwa integrasi BERT dan *cosine similarity* efektif dalam mendeteksi kemiripan teks, terutama dalam menangkap makna semantik dan teks terjemahan atau lintas bahasa.

Kata Kunci : BERT, Cosine Similarity, Deteksi Kemiripan



ABSTRACT

This research develops a tool for detecting text similarity by integrating the Bidirectional Encoder Representations from Transformers (BERT) algorithm and cosine similarity. The system is designed to enhance accuracy in detecting text similarity, including paraphrased texts and translations across languages. The results indicate that the combination of BERT and cosine similarity performs well, achieving an F1 score of 0.83 and a Mean Absolute Error (MAE) of 0.139. The system can detect semantic similarities with high accuracy, reaching similarity scores of up to 1.0 for identical texts and 0.97 for paraphrased texts. Additionally, the system demonstrates flexibility in detecting similarity in translated content across various languages. This research proves that the integration of BERT and cosine similarity is effective in detecting text similarity, particularly in capturing semantic meanings and translated or cross-language texts.

Keywords: BERT, Cosine Similarity, Similarity Detection

