

ABSTRAK

Fenomena *information overload* pada portal berita daring berbahasa Indonesia sering kali memicu tingginya tingkat redundansi informasi lintas artikel, sehingga menyulitkan pembaca dalam memperoleh intisari peristiwa secara efisien. Penelitian ini mengusulkan metode peringkasan multi-dokumen ekstraktif yang mengintegrasikan *Agglomerative Clustering* sebagai algoritma pemilah topik, representasi semantik *Sentence-BERT* (SBERT) menggunakan varian *paraphrase-multilingual-MiniLM-L12-v2*, serta algoritma *Maximal Marginal Relevance* (MMR) sebagai mekanisme penyeleksi kalimat anti-redundansi. Pendekatan komputasi berbasis *dense embedding* ini diimplementasikan guna mengatasi keterbatasan representasi statistik konvensional dalam mendeteksi makna parafrasa. Pengujian arsitektur dilakukan terhadap *benchmark dataset* dari portal berita Liputan6 yang terdistribusi ke dalam enam kategori topik utama. Kualitas ringkasan dievaluasi menggunakan metrik ROUGE terhadap *human ground truth* yang telah tervalidasi reliabilitasnya dengan nilai kesepakatan aktual (*Percentage Agreement*) sebesar 63,21%. Hasil eksperimen komputasional pada skenario tingkat kompresi 20%, 30%, 40%, dan 50% memvalidasi bahwa tingkat kompresi 30% merupakan titik keseimbangan (*sweet spot*) paling optimal antara kepadatan informasi utama dan keberagaman frasa. Pada ambang batas kompresi tersebut, metode usulan secara keseluruhan meraih nilai rata-rata performa kompetitif dengan *F1-Score* ROUGE-1 sebesar 0,4750, ROUGE-2 sebesar 0,3034, dan ROUGE-L sebesar 0,3391. Penelitian ini membuktikan secara empiris bahwa integrasi pemahaman semantik mendalam dan metrik diversitas komputasional efektif untuk mereduksi redundansi leksikal sekaligus mempertahankan akurasi faktual teks berita bagi pembaca digital.

Kata Kunci : Peringkasan Multi-Dokumen Ekstraktif, *Sentence-BERT*, *Maximal Marginal Relevance*, ROUGE.