

ANALISIS PENGARUH *CONVEX COMBINATION* TERHADAP PERFORMA *HYBRID* BM25 DAN INDOSBERT PADA BERITA BERBAHASA INDONESIA

Irawati Kusumadewi¹, Deni Arifianto², Nanda Kurnia Wardati³

¹²³Program Studi Sistem Informasi, Fakultas Teknik,
Universitas Muhammadiyah Jember

Email: irrawkd12@gmail.com¹, deniarifianto@unmuhjember.ac.id²,
nandakurniawardati@unmuhjember.ac.id³

ABSTRAK

Metode *Information Retrieval* berbasis leksikal seperti BM25 efektif dalam mencocokkan kata kunci, namun memiliki keterbatasan dalam memahami hubungan semantik antar kata. Sebaliknya, IndoSBERT mampu merepresentasikan makna dokumen secara kontekstual, tetapi kurang memperhatikan kecocokan istilah secara eksplisit. Penelitian ini bertujuan menganalisis efektivitas pemeringkatan BM25, IndoSBERT, dan *Hybrid Retrieval*, serta pengaruh variasi pembobotan *Convex Combination* dalam menggabungkan kedua metode tersebut pada pencarian berita berbahasa Indonesia. Dataset penelitian diperoleh melalui proses *web scraping* berita dari detik.com, kemudian dilakukan tahapan *preprocessing*, pembentukan indeks BM25, *chunking*, pembentukan *embedding* IndoSBERT, serta proses *retrieval*. Evaluasi dilakukan terhadap *candidate pool* yang dibentuk menggunakan teknik *Depth-10 Pooling* dan divalidasi melalui *expert judgment*, dengan menggunakan metrik *Precision@5*, *Precision@10*, *NDCG@5*, *NDCG@10*, dan *latency*. Hasil penelitian menunjukkan bahwa variasi pembobotan pada metode *Convex Combination* memengaruhi hasil pemeringkatan dokumen. Berdasarkan hasil evaluasi, konfigurasi $\alpha = 0.8$ menghasilkan nilai *NDCG@5* sebesar 1 dan *NDCG@10* sebesar 0.9989, serta memberikan kualitas pemeringkatan yang paling mendekati *ground truth* pada skenario evaluasi penelitian ini. Temuan tersebut menunjukkan bahwa pendekatan *Hybrid Retrieval* berpotensi meningkatkan kualitas pemeringkatan dibandingkan BM25 maupun IndoSBERT pada skenario evaluasi yang digunakan.

Kata kunci: *Information Retrieval*, BM25, IndoSBERT, *Hybrid Retrieval*, *Convex Combination*.