

ABSTRAK

Perkembangan media sosial seperti platform X menghasilkan volume besar opini publik terkait rehabilitasi ekosistem mangrove di Indonesia, sehingga memerlukan pendekatan analisis teks yang efektif sekaligus transparan. Meskipun pemodelan topik menggunakan BERTopic mampu menangani teks pendek dengan baik, sifatnya yang *black-box* menyulitkan pengguna untuk menginterpretasi alasan di balik hasil tersebut. Oleh karena itu, penelitian ini bertujuan untuk mengoptimasi fungsi objektif guna menyeimbangkan koherensi dan diversitas topik, serta mengimplementasikan *Explainable Artificial Intelligence* (XAI) untuk meningkatkan transparansi model. Dalam proses implementasinya, sebanyak 59.460 tweet berbahasa Indonesia dimodelkan menggunakan representasi *embedding* IndoBERTweet dengan mengeksplorasi tiga skema evaluasi multiobjektif: *weighted sum*, *geometric mean*, dan *weighted sum* dengan penalti. Untuk proses interpretasinya, digunakan *surrogate model* berbasis Logistic Regression yang dianalisis secara global menggunakan SHAP dan secara lokal menggunakan LIME. Hasil tersebut kemudian divalidasi lebih lanjut melalui pelabelan topik oleh *Large Language Model* (LLM) serta evaluasi berbasis nalar manusia. Hasil penelitian menyimpulkan bahwa skenario penalti ($\alpha = 0,7$) memberikan kinerja algoritmik terbaik dengan akurasi *surrogate* sebesar 0,85 dan proporsi *outlier* 7,34%, serta mampu menjaga wacana ekologis dari *noise* secara efektif. Namun, evaluasi berbasis penilaian manusia justru menunjukkan preferensi terhadap skenario *geometric mean*, karena dianggap menghasilkan klaster yang lebih tajam dan mudah dipahami secara logis. Temuan ini secara langsung menegaskan bahwa terdapat perbedaan yang jelas antara performa matematis algoritma dan interpretabilitas manusia, di mana metode XAI terbukti berperan penting untuk menjembatani keseimbangan tersebut dalam analisis wacana publik.

Kata kunci : Pemodelan Topik, BERTopic, Evaluasi Multiobjektif, Explainable AI, Mangrove