

ABSTRAK

Kualitas udara yang baik merupakan kebutuhan dasar manusia. Jakarta sebagai pusat urbanisasi mengalami tingkat polusi udara yang tinggi dan berdampak pada kesehatan masyarakat. Opini publik terkait isu ini banyak disuarakan melalui media sosial dan media massa. Twitter (sekarang X) memungkinkan masyarakat mengekspresikan pendapat secara spontan, sementara media massa menyampaikan informasi secara lebih terstruktur. Penelitian ini menggunakan kerangka kerja *Knowledge Discovery in Databases* (KDD) untuk menganalisis persepsi publik terhadap kualitas udara Jakarta yang disuarakan melalui Twitter dan portal berita daring. Data yang digunakan terdiri atas 2.084 *tweet* dan 535 judul artikel berita daring yang diambil sepanjang tahun 2024, dikumpulkan dengan pencarian berbasis kata kunci. Analisis sentimen dilakukan menggunakan model IndoBERT dengan strategi *focal loss* dan augmentasi untuk data Twitter, serta *class weighting* untuk data berita. Hasil evaluasi menunjukkan akurasi sebesar 87,8% (Twitter) dan 88.2% (berita). Analisis *net sentiment* menunjukkan nada negatif pada Twitter (-76,6%) dan lebih netral pada berita (+20,75%). Uji *Chi-Square* menghasilkan nilai $p = 0,3454$, menunjukkan tidak terdapat hubungan signifikan antara distribusi sentimen antarplatform, dengan nilai *Cramer's V* sebesar 0,14 yang mengindikasikan kekuatan asosiasi yang sangat lemah. Untuk memperkaya analisis, dilakukan pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA), yang menunjukkan bahwa karakteristik Twitter cenderung menekankan keluhan emosional dan pengalaman pribadi, sedangkan berita memuat narasi institusional dan kebijakan pemerintah.

Kata kunci: Analisis sentimen, IndoBERT, Korelasi, Kualitas udara, Pemodelan topik.