

ABSTRAK

Kualitas udara di Indonesia terus menjadi perhatian akibat tingginya polusi yang salah satunya disebabkan oleh emisi Pembangkit Listrik Tenaga Uap (PLTU) berbahan bakar batu bara. Permasalahan ini memicu kekhawatiran publik, yang dapat diamati melalui media sosial dan media berita. Oleh karena itu, penelitian ini penting dilakukan untuk memahami bagaimana persepsi masyarakat terbentuk dan berkembang terkait isu PLTU di Indonesia. Penelitian ini bertujuan mengelompokkan wacana publik dan mengidentifikasi tema dominan yang berkembang. Metode yang digunakan adalah *Knowledge Discovery in Database* (KDD) yang tahapannya terdiri dari *data selection*, *data processing*, *data transformation*, *data mining*, dan *evaluation/interpretation*. Algoritma K-Means diterapkan untuk proses pengelompokan dan dilanjutkan dengan *topic modeling* untuk menemukan topik utama pada setiap kluster. Data penelitian berasal dari Twitter dan media berita daring nasional selama periode 2023–2024. Hasil penelitian menunjukkan bahwa algoritma K-Means berhasil mengelompokkan data menjadi tiga kluster utama dengan nilai *silhouette score* sebesar 0,6380. Kesimpulan dari penelitian ini menunjukkan bahwa opini publik terbagi menjadi tiga tema utama, yaitu kekhawatiran terhadap dampak lingkungan PLTU, perhatian pada aspek teknis dan inovasi operasional, serta pembahasan mengenai sumber energi alternatif pembangkit listrik. Temuan ini membuktikan bahwa *machine learning* dapat memetakan persepsi masyarakat terkait PLTU secara sistematis.

Kata Kunci: K-Means, *machine learning*, media sosial, PLTU, *topic modeling*