

ABSTRAK

SYAUQIYAH MUJAHIDAH YAHYA, 2026. Implementasi Algoritma *CatBoost* Berbasis *Text Features* pada Klasifikasi Sentimen Publik Terkait Konflik Ekologi dan Ekonomi Pertambangan (Dibimbing oleh Muhyiddin A. M. Hayat, S.Kom., M.T., dan Ir. Desi Anggreani, S.Kom.,M.T.).

Isu pertambangan kerap memicu dilema antara eksploitasi ekonomi dan pelestarian ekologi, yang memunculkan polarisasi opini publik secara masif di media sosial. Kumpulan opini publik yang berupa data tidak terstruktur ini berpotensi menjadi acuan mitigasi konflik jika dianalisis dengan tepat, namun metode analisis manual terbukti tidak efisien dan rentan terhadap bias. Penelitian ini bertujuan untuk mengimplementasikan algoritma *CatBoost* berbasis *Text Features* guna mengklasifikasikan sentimen publik terhadap isu pertambangan ke dalam kelas positif, negatif, dan netral. Data penelitian bersumber dari komentar TikTok dan YouTube yang diekstraksi menggunakan metode web scraping, menghasilkan 2.049 data bersih. Pemrosesan dilakukan dengan memanfaatkan fitur *Native Text Handling* pada arsitektur *CatBoost*. Melalui fitur ini, teks diproses secara internal menggunakan tokenisasi dan kalkulasi kamus frekuensi (*Bag-of-Words* dan *BM25*) tanpa memerlukan ekstraksi fitur matriks sparse eksternal seperti *TF-IDF*, sehingga mampu mereduksi beban komputasi dan memori secara signifikan. Data dibagi menggunakan teknik *stratified random sampling* dengan rasio 80:20, yang terdiri dari 1.639 data latih dan 410 data uji. Hasil evaluasi model menggunakan *5-Fold Cross Validation*, *Confusion Matrix*, dan optimasi hyperparameter *Optuna* (dengan penyesuaian kedalaman pohon/depth = 5) menunjukkan tingkat Akurasi Keseluruhan sebesar 79% (0,7878) dan nilai rata-rata *F1-Score Macro* sebesar 0,79. Penurunan parameter depth dari 12 menjadi 5 serta penghapusan matriks *TF-IDF* secara signifikan mempercepat waktu komputasi pelatihan model (*training time*). Secara spesifik, kelas Netral mencapai tingkat kepekaan (*Recall*) tertinggi sebesar 85%, sedangkan kelas Negatif mencatatkan tingkat kehati-hatian (*Precision*) tertinggi sebesar 86%. Model klasifikasi berbasis *CatBoost* terbukti tangguh, efisien, dan seimbang dalam memetakan polaritas opini publik secara otomatis sebagai landasan deteksi dini mitigasi konflik sosial-lingkungan.

Kata Kunci: Analisis Sentimen, *CatBoost*, Konflik Pertambangan, *Native Text Features*, TikTok dan YouTube.