

ABSTRAK

Aduan masyarakat merupakan salah satu indikator penting dalam memahami permasalahan publik, khususnya di bidang transportasi dan lalu lintas. Banyaknya aduan yang masuk melalui media sosial, seperti platform X (Twitter), menuntut adanya sistem klasifikasi otomatis yang mampu mengelola informasi secara cepat tepat dan efisien. Penelitian ini bertujuan untuk membangun model klasifikasi teks aduan masyarakat pada akun media sosial X Suara Surabaya (@e100ss) dengan memanfaatkan model *Extreme Gradient Boosting* (XGBoost) dan berbagai metode fitur ekstraksi, yaitu *Term Frequency — Inverse Document Frequency* (TF-IDF), *Count Vectorizer*, dan *Word2vect*. Pengembangan sistem mengikuti tahapan *Cross-Industry Standard Process for Data Mining* (CRISP-DM), dimulai dari pengumpulan data menggunakan teknik *web scraping*, *preprocessing* teks, hingga evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, *f1-score*, *confusion matrix*, dan *ROC Curve*. Hasil penelitian menunjukkan bahwa kombinasi XGBoost dan *CountVectorizer* memberikan performa terbaik dengan *accuracy* 93%, *precision* 90%, *recall* 86%, dan *f1-score* 88%. Model ini dinilai paling seimbang dalam mendeteksi teks aduan tanpa mengorbankan ketepatan klasifikasi (*precision*). Model diimplementasikan dalam bentuk aplikasi website berbasis *Streamlit* yang bersifat interaktif dan dilengkapi dengan fitur manajemen frasa, guna menambahkan frasa baru yang belum dikenali oleh model. Selain itu, penerapan ambang batas prediksi sebesar 0,4 digunakan untuk meningkatkan kemampuan sistem dalam menangkap aduan yang jarang muncul. Dengan pendekatan ini, sistem yang dibangun diharapkan dapat mendukung peningkatan kualitas pelayanan publik melalui pemanfaatan teknologi *machine learning* dalam analisis teks secara otomatis.

Kata Kunci: *klasifikasi teks, xgboost, aduan masyarakat, countvectorizer, pelayanan publik.*