

ABSTRAK

Kehadiran layanan transportasi publik bus Trans Jatim merupakan salah satu upaya pemerintah dalam meningkatkan mobilitas masyarakat di wilayah Surabaya–Bangkalan. Seiring meningkatnya penggunaan layanan tersebut, masyarakat memberikan berbagai ulasan di media sosial yang mencerminkan tingkat kepuasan maupun keluhan terhadap kualitas layanan. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap layanan bus Trans Jatim Koridor V dengan memanfaatkan data komentar dari platform TikTok dan Instagram. Data dikumpulkan melalui proses crawling, kemudian dilakukan serangkaian *preprocessing* teks berupa pembersihan data, *case folding*, *tokenisasi*, *normalisasi*, penghapusan *stopword*, dan *stemming*. Pelabelan sentimen dilakukan menggunakan metode *lexicon-based*. Model klasifikasi dibangun menggunakan algoritma *Random Forest* dengan representasi fitur TF-IDF, serta penanganan ketidakseimbangan kelas menggunakan SMOTE. Pengujian dilakukan dengan empat skenario pembagian data latih dan data uji (80:20 dan 90:10, dengan dan tanpa SMOTE). Hasil evaluasi menunjukkan bahwa performa terbaik diperoleh pada Skenario 4, yaitu pembagian data 90:10 dengan SMOTE, menggunakan parameter $n_estimators = 100$, $max_depth = None$, $min_samples_split = 2$, dan $min_samples_leaf = 1$, dengan tingkat akurasi 90,08%. Temuan ini menunjukkan bahwa proporsi data latih yang lebih besar dan penyeimbangan kelas berpengaruh signifikan terhadap peningkatan akurasi model. Penelitian ini menghasilkan gambaran yang lebih jelas mengenai persepsi masyarakat terhadap layanan bus Trans Jatim serta membuktikan bahwa *Random Forest* efektif digunakan dalam analisis sentimen berbasis teks.

Kata kunci: Layanan Bus, Analisis Sentimen, *Random Forest*, TF-IDF, SMOTE, Trans Jatim Koridor V

ABSTRACT

The presence of the Trans Jatim public transportation service is one of the government's efforts to improve community mobility in the Surabaya–Bangkalan region. As the use of this service increases, the public has shared various opinions on social media, reflecting their satisfaction as well as complaints regarding the service quality. This study aims to analyze public sentiment toward the Trans Jatim Corridor V bus service by utilizing user comments collected from TikTok and Instagram. The data were collected using a crawling technique, followed by a series of text preprocessing steps including data cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Sentiment labeling was conducted using a lexicon-based approach. The classification model was built using the Random Forest algorithm with TF-IDF feature representation, and data imbalance was addressed using the SMOTE technique. The evaluation was performed using four scenarios of training–testing data split (80:20 and 90:10, with and without SMOTE). The results show that the best performance was achieved in Scenario 4, which applied a 90:10 training–testing split combined with SMOTE, using the parameters $n_estimators = 100$, $max_depth = None$, $min_samples_split = 2$, and $min_samples_leaf = 1$, resulting in an accuracy of 90.08%. These findings indicate that a larger portion of training data and balanced class distribution significantly improve model accuracy. The study provides a clearer understanding of public perception toward the Trans Jatim service and demonstrates that Random Forest is effective for text-based sentiment analysis.

Keywords: *Bus Service, Sentiment Analysis, Random Forest, TF-IDF, SMOTE, Trans Jatim Corridor V.*