

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen dan mengidentifikasi topik yang terdapat pada ulasan pengguna layanan transportasi publik TransJatim yang diperoleh dari Google Play Store, media sosial X (Twitter), dan TikTok. Penelitian ini menggunakan kombinasi metode Logistic Regression untuk klasifikasi sentimen dan Non-Negative Matrix Factorization (NMF) untuk pemodelan topik, sehingga hasil analisis dapat memberikan gambaran yang lebih menyeluruh. Dataset yang digunakan terdiri dari 2.656 ulasan yang telah melalui tahap preprocessing, ekstraksi fitur menggunakan TF-IDF, serta penyeimbangan data dengan metode SMOTE untuk mengatasi ketidakseimbangan kelas. Hasil pengujian menunjukkan bahwa model Logistic Regression memiliki kinerja yang sangat baik dengan nilai akurasi, presisi, recall, dan F1-score sebesar 0,9524. Model ini juga sangat baik dalam mendeteksi sentimen negatif dengan nilai recall mencapai 1,00, namun masih memiliki keterbatasan dalam mengenali sentimen positif dengan recall sebesar 0,73. Pada tahap pemodelan topik menggunakan NMF, diperoleh empat topik utama yang menggambarkan fokus ulasan pengguna, yaitu terkait rute perjalanan, operasional layanan, informasi layanan, serta pengalaman pengguna dalam menggunakan transportasi. Evaluasi menggunakan coherence score menghasilkan nilai 0,5564 yang menunjukkan bahwa topik yang dihasilkan cukup konsisten dan sesuai dengan makna data. Secara keseluruhan, kombinasi Logistic Regression dan NMF terbukti efektif dalam menganalisis sentimen sekaligus mengungkap struktur topik dari ulasan pengguna. Hasil penelitian ini diharapkan dapat menjadi dasar dalam pengambilan keputusan untuk meningkatkan kualitas layanan transportasi publik berbasis data.

Kata Kunci : Analisis Sentimen, Logistic Regression, Non-Negative Matrix Factorization, TransJatim

ABSTRACT

This study aims to analyze sentiment and identify topics contained in user reviews of TransJatim public transportation services obtained from the Google Play Store, social media X (Twitter), and TikTok. This study uses a combination of the Logistic Regression method for sentiment classification and Non-Negative Matrix Factorization (NMF) for topic modeling, so that the analysis results can provide a more comprehensive picture. The dataset used consists of 2,656 reviews that have gone through the preprocessing stage, feature extraction using TF-IDF, and data balancing using the SMOTE method to address class imbalance. The test results show that the Logistic Regression model has excellent performance with accuracy, precision, recall, and F1-score values of 0.9524. This model is also very good at detecting negative sentiment with a recall value reaching 1.00, but still has limitations in recognizing positive sentiment with a recall of 0.73. In the topic modeling stage using NMF, four main topics were obtained that describe the focus of user reviews, namely related to travel routes, service operations, service information, and user experience in using transportation. Evaluation using the coherence score yielded a value of 0.5564, indicating that the resulting topics were quite consistent and aligned with the meaning of the data. Overall, the combination of Logistic Regression and NMF proved effective in analyzing sentiment and uncovering the topic structure of user reviews. The results of this study are expected to inform decision-making to improve the quality of data-driven public transportation services.

Keywords: Sentiment Analysis, Logistic Regression, Non-Negative Matrix Factorization, TransJatim