

ABSTRAK

Kemacetan lalu lintas di Jakarta mendorong pemerintah untuk meningkatkan efisiensi transportasi umum, salah satunya melalui pengoperasian MRT Jakarta. Sebagai layanan pendukung, aplikasi MyMRTJ dikembangkan untuk menyediakan informasi mengenai jadwal, rute, dan pemesanan tiket MRT. Namun, ulasan pengguna di *Google Play Store* masih belum tersusun dengan baik dan sulit untuk dipahami. Oleh karena itu, diperlukan analisis sentimen untuk memahami kecenderungan opini pengguna terhadap aplikasi MyMRTJ. Penelitian ini bertujuan untuk mengkaji lebih lanjut sentimen dalam ulasan pengguna aplikasi MyMRTJ dengan jumlah data yang dianalisis sebanyak 1210 ulasan, lalu dikategorikan ke dalam sentimen positif, negatif, dan netral. Selanjutnya, data diproses melalui tahapan *pre-processing* yang mencakup *cleaning*, *case folding*, *tokenization*, *filtering*, *normalization*, dan *stemming*. Metode yang digunakan pada penelitian ini yaitu *K-Nearest Neighbor* (K-NN) dan pendekatan *Lexicon-Based*. Selain itu, karena terdapat ketidakseimbangan pada dataset, peneliti menerapkan *Synthetic Minority Oversampling Technique* (SMOTE) dalam mengatasi masalah tersebut. Evaluasi model dilakukan menggunakan *confusion matrix* dengan mengukur metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Dari hasil pengujian diperoleh hasil sebagai berikut, metode *K-Nearest Neighbor* (K-NN) dan pendekatan *Lexicon-Based* tanpa penerapan SMOTE menghasilkan *accuracy* 71,52%, *precision* 88,15%, *recall* 44,49%, dan *f1-score* 47,60%. Sedangkan hasil setelah menerapkan teknik SMOTE, akurasi meningkat menjadi 80,99%, dengan *precision* 71,14%, *recall* 53,43%, dan *F1-score* 57,95% pada nilai $K=3$. Hasil ini menunjukkan bahwa penggunaan SMOTE dapat meningkatkan performa model dengan memperbaiki prediksi pada kelas minoritas.

Kata kunci: Analisis Sentimen, *K-Nearest Neighbor*, *Lexicon-Based*, *Synthetic Minority Oversampling Technique* (SMOTE), MyMRTJ, *Google Play Store*.

ABSTRACT

Traffic congestion in Jakarta has encouraged the government to improve the efficiency of public transportation, one of which is through the operation of the Jakarta MRT. As a supporting service, the MyMRTJ application was developed to provide information on MRT schedules, routes, and ticket reservations. However, user reviews on the Google Play Store are still not well organized and difficult to understand. Therefore, sentiment analysis is needed to understand the tendency of user opinions towards the MyMRTJ application. This study aims to further examine the sentiment in user reviews of the MyMRTJ application with the amount of data analyzed as many as 1210 reviews, then grouped into positive, negative, and neutral sentiments. Furthermore, the data is processed through pre-processing stages including cleaning, casefolding, tokenization, filtering, normalization, and stemming. The methods used in this study are K-Nearest Neighbor (K-NN) and the Lexicon-Based approach. In addition, because there is an imbalance in the dataset, researchers apply the Synthetic Minority Oversampling Technique (SMOTE) to overcome this problem. Model evaluation is carried out using a confusion matrix by measuring the metrics of accuracy, precision, recall, and F1-score. From the test results, the following results were obtained, the K-Nearest Neighbor (K-NN) method and the Lexicon-Based approach without applying SMOTE produced an accuracy of 71.52%, precision of 88.15%, recall of 44.49%, and f1-score of 47.60%. While the results after applying the SMOTE technique, the accuracy increased to 80.99%, with precision of 71.14%, recall of 53.43%, and F1-score of 57.95% at a value of $K = 3$. These results indicate that the use of SMOTE can improve model performance by improving predictions in the minority class.

Keywords: *Sentiment Analysis, K-Nearest Neighbor, Lexicon-Based, Synthetic Minority Oversampling Technique (SMOTE), MyMRTJ, Google Play Store.*