

ABSTRAK

Telaga Sarangan merupakan salah satu destinasi wisata populer di Indonesia yang banyak menerima ulasan dari pengunjung melalui platform digital seperti *Google Maps*. Ulasan-ulasan tersebut mencerminkan opini dan tingkat kepuasan pengunjung, namun banyaknya ulasan menyulitkan pihak pengelola maupun wisatawan dalam memperoleh informasi yang terstruktur. Penelitian ini bertujuan untuk melakukan klasifikasi sentimen *multi-class* (positif, netral, dan negatif) terhadap ulasan Telaga Sarangan menggunakan metode *Robustly Optimized BERT Pretraining Approach* (RoBERTa). Data yang digunakan sebanyak 3.875 ulasan hasil *scraping* dari *Google Maps*. Untuk mengatasi ketidakseimbangan distribusi kelas sentimen, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). Hasil pelatihan model menunjukkan bahwa kombinasi SMOTE dan RoBERTa secara signifikan meningkatkan performa klasifikasi. Tanpa SMOTE dengan kombinasi *batch size* 8 dan *epoch* 8, akurasi tertinggi hanya mencapai 82.29%, sedangkan setelah penerapan SMOTE akurasi meningkat hingga 93.50%. Selain itu, terjadi peningkatan menyeluruh pada nilai *precision*, *recall*, dan *F1-score*. Penelitian ini membuktikan bahwa kombinasi SMOTE dan RoBERTa efektif dalam menghasilkan model klasifikasi sentimen yang akurat dan seimbang, serta berpotensi mendukung analisis opini publik secara lebih informatif di sektor pariwisata.

Kata Kunci: Analisis Sentimen, *Multi-Class*, RoBERTa, *SMOTE*

ABSTRACT

Telaga Sarangan is one of the popular tourist destinations in Indonesia that receives many reviews from visitors through digital platforms such as Google Maps. These reviews reflect the opinions and satisfaction levels of visitors, but the large number of reviews makes it difficult for managers and tourists to obtain structured information. This study aims to perform multi-class sentiment classification (positive, neutral, and negative) on Telaga Sarangan reviews using the Robustly Optimized BERT Pretraining Approach (RoBERTa) method. The data used were 3875 reviews scraped from Google Maps. To overcome the imbalance of sentiment class distribution, Synthetic Minority Over-sampling Technique (SMOTE) is used. The model training results show that the combination of SMOTE and RoBERTa significantly improves the classification performance. Without SMOTE with the combination of batch size 8 and epoch 8, the highest accuracy only reached 82.29%, while after the application of SMOTE the accuracy increased to 93.50%. In addition, there was an overall improvement in the precision, recall, and F1-score values. This research proves that the combination of SMOTE and RoBERTa is effective in producing accurate and balanced sentiment classification models, and has the potential to support more informative analysis of public opinion in the tourism sector.

Keywords: *Sentiment Analysis, Multi-Class, RoBERTa, SMOTE*