

ABSTRAK

Peran ulasan wisatawan pada platform digital seperti Google Maps sangat penting dalam membentuk citra destinasi pariwisata, termasuk wisata pantai di Kabupaten Sampang. Namun, data ulasan yang tidak terstruktur, mengandung kesalahan penulisan, serta memiliki distribusi kelas yang tidak seimbang menyebabkan proses analisis secara manual menjadi kurang efektif dan memakan waktu. Penelitian ini menggunakan dataset hasil *web scraping* Google Maps sebanyak ± 2.879 ulasan wisata pantai di Kabupaten Sampang yang terdiri atas 1.917 sentimen positif, 724 sentimen negatif, dan 238 sentimen netral. Penelitian ini bertujuan untuk membangun model analisis sentimen yang mampu mengklasifikasikan ulasan wisatawan ke dalam tiga kategori, yaitu positif, negatif, dan netral, menggunakan metode *Long Short-Term Memory* (LSTM) yang mampu memahami konteks data sekuensial. Untuk meningkatkan performa model, penelitian ini mengintegrasikan teknik *Damerau-Levenshtein Distance* (DLD) untuk koreksi kesalahan penulisan, Text Augmentation untuk mengatasi ketidakseimbangan data, serta fitur N-gram untuk memperkaya representasi teks. Hasil penelitian menunjukkan bahwa penerapan DLD saja belum mampu meningkatkan performa secara optimal, terutama dalam mengenali kelas minoritas. Namun, ketika dikombinasikan dengan *Text Augmentation*, performa model meningkat secara signifikan. Model terbaik diperoleh pada skenario penggunaan fitur unigram yang dikombinasikan dengan DLD dan Text Augmentation, dengan nilai akurasi sebesar 93,05%, precision 93%, recall 93%, dan F1-score 93%. Hasil tersebut menunjukkan bahwa kualitas data dan keseimbangan distribusi kelas memiliki peran penting dalam meningkatkan kinerja model. Dengan demikian, penelitian ini memberikan kontribusi dalam pengembangan analisis sentimen berbasis deep learning serta dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam pengelolaan dan pengembangan pariwisata berbasis data di Kabupaten Sampang.

Kata Kunci : Analisis Sentimen, *Long Short-Term Memory*, Wisata Pantai, *Text Augmentation*, *Damerau-Levenshtein Distance*, N-Gram, Kabupaten Sampang.

ABSTRACT

The role of tourist reviews on digital platforms such as Google Maps is highly important in shaping the image of tourism destinations, including beach tourism in Sampang Regency. However, unstructured review data, spelling errors, and imbalanced class distributions make manual analysis less effective and time-consuming. This study utilized a dataset obtained through web scraping from Google Maps, consisting of approximately 2,879 beach tourism reviews in Sampang Regency, including 1,917 positive sentiments, 724 negative sentiments, and 238 neutral sentiments. This research aims to develop a sentiment analysis model capable of classifying tourist reviews into three categories: positive, negative, and neutral, using the Long Short-Term Memory (LSTM) method, which is able to capture the contextual information of sequential data. To improve model performance, this study integrates the Damerau-Levenshtein Distance (DLD) technique for spelling correction, Text Augmentation to address data imbalance, and N-gram features to enrich text representation. The results indicate that the application of DLD alone was not sufficient to significantly improve model performance, particularly in recognizing minority classes. However, when combined with Text Augmentation, the model's performance improved substantially. The best-performing model was achieved using the unigram feature combined with DLD and Text Augmentation, resulting in an accuracy of 93.05%, precision of 93%, recall of 93%, and an F1-score of 93%. These findings demonstrate that data quality and balanced class distribution play a crucial role in enhancing model performance. Therefore, this study contributes to the development of deep learning-based sentiment analysis and can serve as a foundation for data-driven decision-making in tourism management and development in Sampang Regency.

Keywords: Sentiment Analysis, Long Short-Term Memory, Beach Tourism, Text Augmentation, Damerau-Levenshtein Distance, N-Gram, Sampang Regency.