

ABSTRAK

KLASIFIKASI TEMPAT WISATA DI INDONESIA BERDASARKAN SENTIMEN ULASAN PENGUNJUNG MENGGUNAKAN ALGORITMA *BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS*

Oleh:

Rd. Imam Saepul Millah

1217050118

Perkembangan sektor pariwisata di Indonesia menghasilkan volume ulasan pengunjung pada platform digital yang sangat besar, sehingga sering kali menimbulkan *information overload*. Kondisi ini menyulitkan wisatawan dalam menyaring informasi yang relevan dan pengelola destinasi untuk memahami persepsi publik secara efisien. Penelitian ini bertujuan untuk membangun dan mengevaluasi model klasifikasi tempat wisata di Indonesia berdasarkan sentimen pada ulasan pengunjung menggunakan algoritma *Bidirectional Encoder Representations from Transformers* (BERT). Metodologi yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang mencakup tahap pengumpulan data ulasan dari Google Maps dan Tripadvisor, persiapan data, pemodelan, hingga evaluasi. Ulasan diklasifikasikan ke dalam dua kategori, yaitu sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa model IndoBERT yang dikembangkan mampu mencapai kinerja sangat tinggi dengan tingkat akurasi terbaik sebesar 96,64%. Meskipun demikian, analisis lebih lanjut menunjukkan adanya pengaruh kuat dari ketidakseimbangan data, di mana model sangat unggul dalam mengenali kelas mayoritas (positif) namun belum optimal pada kelas minoritas (negatif). Penelitian ini membuktikan bahwa algoritma BERT sangat efektif untuk analisis sentimen berbahasa Indonesia dalam konteks pariwisata dan dapat menjadi landasan untuk pengembangan sistem informasi destinasi yang lebih cerdas dan berbasis data.

Kata Kunci: Klasifikasi Sentimen, BERT, Pariwisata, Ulasan Pengunjung, CRISP-DM, *Natural Language Processing*.

ABSTRACT

CLASSIFICATION OF TOURIST ATTRACTIONS IN INDONESIA BASED ON VISITOR REVIEW SENTIMENT USING THE BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS ALGORITHM

Oleh:

Rd. Imam Saepul Millah

1217050118

The development of the tourism sector in Indonesia has resulted in a huge volume of visitor reviews on digital platforms, often leading to information overload. This condition makes it difficult for tourists to filter relevant information and destination managers to understand public perception efficiently. This research aims to build and evaluate a classification model of tourist attractions in Indonesia based on sentiment in visitor reviews using the Bidirectional Encoder Representations from Transformers (BERT) algorithm. The methodology used is the Cross-Industry Standard Process for Data Mining (CRISP-DM), which includes the stages of collecting review data from Google Maps and Tripadvisor, data preparation, modeling, and evaluation. Reviews are classified into two categories, namely positive and negative sentiments. The results showed that the developed IndoBERT model was able to achieve very high performance with the best accuracy rate of 96.64%. However, further analysis shows the strong influence of data imbalance, where the model excels in recognizing the majority (positive) class but is not optimal for the minority (negative) class. This research proves that the BERT algorithm is highly effective for Indonesian sentiment analysis in the context of tourism and can serve as a foundation for the development of smarter, data-driven destination information systems.

Keywords: Sentiment Classification, BERT, Tourism, Visitor Reviews, CRISP-DM, Natural Language Processing.