

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan ulasan pengunjung wisata Kali Gajah Wong berdasarkan tema dominan yang terkandung dalam Google Reviews. Data ulasan tersebut dianalisis menggunakan pendekatan pemodelan topik berbasis teks untuk memberikan gambaran yang terstruktur mengenai persepsi pengunjung terhadap destinasi wisata tersebut. Proses pra-pemrosesan data meliputi lowercasing, penghapusan angka dan tanda baca, normalisasi kata, penghapusan kata berhenti (stopword removal), serta stemming untuk meningkatkan kualitas dan konsistensi data teks. Metode utama yang digunakan adalah BERTopic, sebuah teknik pemodelan topik terkini yang memanfaatkan representasi teks berbasis embedding, reduksi dimensi menggunakan Uniform Manifold Approximation and Projection (UMAP), serta clustering dengan Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Pendekatan ini memungkinkan identifikasi topik utama secara otomatis tanpa memerlukan penentuan jumlah cluster secara eksplisit.

Hasil pemodelan menunjukkan adanya enam topik dominan yang mewakili berbagai aspek pengalaman pengunjung, yaitu kebersihan, aksesibilitas, fasilitas, suasana lingkungan, dan tema wisata. Visualisasi distribusi ulasan pada masing-masing topik mengindikasikan dominasi persepsi positif terhadap pengelolaan kawasan wisata.

Penelitian ini memberikan kontribusi signifikan dengan penerapan BERTopic pada dataset ulasan berbahasa Indonesia dalam sektor pariwisata, yang masih tergolong terbatas pada kajian akademik. Dengan memadukan teknik text mining modern dan metode pembelajaran tanpa pengawasan (unsupervised learning) berbasis transformer, penelitian ini menunjukkan efektivitas BERTopic dalam memahami opini publik terhadap destinasi wisata lokal. Temuan ini dapat menjadi referensi bagi pengelola destinasi wisata dalam meningkatkan kualitas pelayanan dan pengalaman pengunjung.

Kata kunci: Pemodelan topik, BERTopic, Google Reviews, pariwisata, text mining, unsupervised learning, UMAP, HDBSCAN, ulasan wisata, analisis opini.

ABSTRACT

This study aims to cluster visitor reviews of Kali Gajah Wong tourism site based on dominant themes extracted from Google Reviews. The reviews were analyzed using a text-based topic modeling approach to provide a structured overview of visitors' perceptions towards the destination. The data preprocessing stage involved lowercasing, removal of numbers and punctuation, word normalization, stopword removal, and stemming to improve the quality and consistency of the textual data. The primary method employed was BERTopic, a state-of-the-art topic modeling technique that leverages text representation through embeddings, dimensionality reduction via Uniform Manifold Approximation and Projection (UMAP), and clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). This approach enables automatic identification of main topics without the need to specify the number of clusters in advance.

The modeling results revealed six dominant topics representing various aspects of visitor experience, including cleanliness, accessibility, facilities, environmental atmosphere, and tourism themes. Visualization of review distributions within each topic indicated a predominance of positive perceptions regarding the management of the tourism area.

This study contributes significantly by applying BERTopic to an Indonesian-language review dataset in the tourism sector, which remains limited in academic research. By integrating modern text mining techniques and transformer-based unsupervised learning methods, the study demonstrates the effectiveness of BERTopic in capturing public opinion about local tourism destinations. The findings are expected to serve as a reference for tourism managers to enhance service quality and visitor experience.

Keywords: Topic modeling, BERTopic, Google Reviews, tourism, text mining, unsupervised learning, UMAP, HDBSCAN, visitor reviews, opinion analysis.