

Abstrak

Sektor pariwisata memiliki peran penting dalam mendorong pertumbuhan ekonomi regional. Namun, keberagaman destinasi wisata di Kabupaten Tuban menyebabkan kesulitan bagi wisatawan maupun pemerintah dalam mengidentifikasi karakteristik masing-masing objek wisata secara sistematis. Kurangnya pengelompokan berbasis data mengakibatkan strategi promosi dan pengembangan destinasi wisata menjadi kurang optimal dan tidak tepat sasaran. Penelitian ini bertujuan untuk mengelompokkan objek wisata di Kabupaten Tuban menggunakan algoritma K-Means Clustering berdasarkan beberapa karakteristik, yaitu harga tiket, rating, jumlah fasilitas, jarak, dan durasi operasional. Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining melalui algoritma K-Means Clustering. Dataset yang digunakan terdiri dari 200 destinasi wisata yang diperoleh dari Dinas Pariwisata, dengan atribut harga tiket, rating, fasilitas, jarak, dan lama operasional. Tahapan penelitian meliputi: Data preprocessing (cleaning dan transformasi data), sehingga diperoleh 187 data bersih. Penentuan jumlah cluster optimal menggunakan metode Elbow dan Silhouette Score. Proses clustering menggunakan algoritma K-Means. Hasil evaluasi menunjukkan bahwa jumlah cluster optimal adalah 2 cluster dengan nilai silhouette score sebesar 0,7118 yang termasuk kategori kuat (strong). Pengelompokan anggota cluster didapat Cluster 0 sebanyak 28 destinasi wisata. Cluster 1 sebanyak 159 destinasi wisata. Penelitian ini memberikan kontribusi berupa segmentasi destinasi wisata berbasis karakteristik yang dapat dimanfaatkan oleh berbagai pihak, yaitu Bagi pemerintah daerah: membantu dalam merumuskan strategi pengembangan dan promosi wisata yang lebih tepat sasaran, serta mengoptimalkan alokasi sumber daya. Bagi pengelola wisata: memberikan informasi terkait posisi objek wisata dalam kelompok tertentu sebagai bahan evaluasi untuk meningkatkan kualitas layanan, fasilitas, dan daya tarik wisata.

Kata kunci: Data Mining, Clustering, K-Means, Pariwisata, Silhouette Score.

Abstract

The tourism sector plays a vital role in driving regional economic growth. However, the diversity of tourist destinations in Tuban Regency makes it difficult for both tourists and the government to systematically identify the characteristics of each tourist attraction. The lack of data-driven clustering results in tourism promotion and destination development strategies that are suboptimal and miss the mark. This study aims to cluster tourist attractions in Tuban Regency using the K-Means Clustering algorithm based on several characteristics, namely ticket price, rating, number of facilities, distance, and operational duration. This study employs a quantitative approach using data mining techniques through the K-Means Clustering algorithm. The dataset used consists of 200 tourist destinations obtained from the Tourism Office, with attributes including ticket price, rating, facilities, distance, and operating hours. The research stages include: Data preprocessing (cleaning and data transformation), resulting in 187 clean data points. Determination of the optimal number of clusters using the Elbow and Silhouette Score methods. The clustering process uses the K-Means algorithm. Evaluation results indicate that the optimal number of clusters is 2, with a silhouette score of 0.7118, which falls into the “strong” category. Cluster membership is as follows: Cluster 0 contains 28 tourist destinations. Cluster 1 consists of 159 tourist destinations. This study contributes to the segmentation of tourist destinations based on characteristics that can be utilized by various parties, namely: For local governments: to assist in formulating more targeted strategies for tourism development and promotion, as well as optimizing resource allocation. For tourism managers: providing information regarding the position of tourist attractions within specific groups as evaluation material to improve the quality of services, facilities, and tourist appeal

Keyword: Keywords: Data Mining, Clustering, K-Means, Tourism, Silhouette Score.