

ABSTRACT

Tourism is a vital sector that attracts both domestic and international visitors through various facilities and promotions, including the implementation of e-tourism supported by quality internet data services. With e-tourism, lesser-known tourist destinations can be more easily introduced, thereby increasing the number of visitors. This study aims to analyze the tourism visitation patterns in North Kalimantan using Principal Component Analysis (PCA) for data dimensionality reduction, and K-Means Clustering to group tourist visits based on the number of visits. The data used in this study were obtained from the Central Statistics Agency of North Kalimantan, the Ministry of Tourism and Creative Economy, and the National Library, covering the observation period of 2020–2022. The results of the analysis show that tourist visits can be grouped into three clusters: destinations categorized as good, fairly good, and less good. The PCA successfully reduced the dataset dimensions into two principal components, explaining 87.84% of the total variance, while K-Means produced a Silhouette Score of 0.736, indicating very good clustering quality. The combination of PCA and K-Means provides effective results in understanding the tourism visitation patterns in North Kalimantan.

Keyword: Tourism, Visiting Places, Tourists, PCA/Principal Component Analysis, K- Means

ABSTRAK

Pariwisata merupakan sektor penting yang menarik kunjungan wisatawan domestik maupun mancanegara melalui berbagai fasilitas dan promosi, termasuk penerapan *e-tourism* yang didukung oleh layanan data internet berkualitas. Dengan *e-tourism*, destinasi wisata yang kurang dikenal dapat lebih mudah diperkenalkan, sehingga meningkatkan jumlah kunjungan wisatawan. Penelitian ini bertujuan untuk menganalisis pola kunjungan wisatawan di Kalimantan Utara menggunakan metode Principal Component Analysis (PCA) untuk reduksi dimensi data, serta metode K-Means Clustering untuk pengelompokan kunjungan wisatawan berdasarkan jumlah kunjungan. Data yang digunakan diperoleh dari Badan Pusat Statistik Kalimantan Utara, Kementerian Pariwisata dan Ekonomi Kreatif, serta Perpustakaan Nasional, dengan periode pengamatan 2020–2022. Hasil analisis menunjukkan bahwa kunjungan wisatawan dapat dikelompokkan menjadi tiga klaster: tempat kunjungan dengan kategori baik, cukup baik, dan kurang baik. Penggunaan PCA berhasil mereduksi dimensi dataset menjadi dua komponen utama yang menjelaskan 87.84% total varians, sedangkan K-Means menghasilkan nilai Silhouette Score sebesar 0.736, menunjukkan kualitas clustering yang sangat baik. Kombinasi PCA dan K-Means memberikan hasil yang efektif dalam memahami pola kunjungan wisatawan di Kalimantan Utara.

Kata Kunci: Pariwisata, Tempat Kunjungan, Wisatawan, PCA/Principal Component Analysis, K-Means