

Pariwisata merupakan salah satu sektor penting dalam peningkatan ekonomi suatu negara, penciptaan lapangan kerja, peningkatan pendapatan, dan pertumbuhan ekonomi secara keseluruhan. Oleh karena itu, analisis statistik pariwisata diperlukan dalam pengambilan keputusan bagi pemangku kebijakan, pelaku bisnis pariwisata dan UMKM. Namun metode yang digunakan dalam pengumpulan data statistik ini masih bersifat konvensional dengan menggunakan survei yang membutuhkan tenaga kerja dan biaya yang mahal. Ini menyebabkan adanya proses yang lama untuk bisa mengakses informasi statistik pariwisata tersebut.

Untuk mengatasi keterbatasan tersebut, terdapat pendekatan yang lebih modern yang memanfaatkan *mobile positioning data* (MPD). Dimana data ini dihasilkan dari aktivitas telepon seluler. Data ini menyediakan informasi spasial dan temporal yang dapat memberikan wawasan tentang kebiasaan wisatawan. Untuk melakukan hal tersebut, perlu dilakukan proses identifikasi wisatawan. Terdapat beberapa metode yang bisa digunakan untuk mengidentifikasi wisatawan, seperti metode *rule-based*, probabilitas, *machine learning*, ataupun *deep learning*.

Penelitian ini berfokus pada identifikasi wisatawan menggunakan metode *machine learning* pada data MPD aktif. Penelitian ini menggunakan beberapa model *machine learning* seperti *Decision Tree*, *Random Forest Classification*, dan XGBoost untuk menentukan model terbaik. Model dengan performa terbaik di antara ketiganya adalah *Random Forest Classification* dengan akurasi sebesar 99%. Hasil klasifikasi tersebut kemudian digunakan untuk menghasilkan statistik pariwisata yang menjelaskan tentang jumlah wisatawan yang berkunjung ke lokasi wisata tertentu, desa/kelurahan asal wisatawan, dan jarak yang ditempuh oleh wisatawan untuk mencapai lokasi wisata. Hasil penelitian ini menunjukkan bahwa data MPD aktif mampu memberikan solusi alternatif terhadap pembuatan statistik pariwisata.

Kata kunci: *Active MPD*, statistik pariwisata, mobilitas wisatawan, klasifikasi, *machine learning*

## ABSTRACT

*Tourism plays a crucial role in the economic growth of a country, contributing significantly to job creation, income generation, and overall economic development. Therefore, tourism statistical analysis is necessary for decision making among policymakers, tourism industry, and UMKM. However, the current method used in tourism statistical data collection is still rely on conventional approach, resulting time-consuming process that delays the public access to the information.*

*To address this limitation, a modern approach utilizing Mobile Positioning Data (MPD) has emerged, which harnesses the data generated by mobile phone activities. This data provide valuable spatial and temporal information that offers insight of tourists behaviours. Various method including Rule-based method, probabilistic method, machine learning method and deep learning method can be used to identify tourists based on MPD.*

*In this study, our focus is utilizing machine learning method to identify tourists based on Active MPD. We employed Decision Tree, Random Forest, and XGBoost machine learning models, comparing their performance to determine the best model. The best performance model between the three is Random Forest with the 99% accuracy. The predicted result were then used to generate a tourism statistic which explain about the number of tourists that visiting specific destinations, their duration of stay, and the distance they traveled to reach the destination. The experimental result indicate that Active MPD presents an alternative way for creating tourism statistic.*

**Keywords :** *Active MPD, Tourism Statistic, Tourist Mobility, Classification, Machine Learning*