

ABSTRACT

Overtourism is a growing global challenge, particularly in destinations where tourism volumes exceed sustainable limits, leading to overcrowding and environmental strain. Bali, a leading global tourist destination, faces significant pressures in its southern regions. This study investigates the relationship between infrastructure density and perceived overcrowding in Bali by integrating binary text classification and geospatial analysis. A fine-tuned BERT-based model categorizes tourist reviews as “crowded” or “not crowded” to extract perceptions of congestion, while Kernel Density Estimation (KDE) maps infrastructure concentration and identifies overcrowding hotspots. The findings reveal a moderate positive correlation between infrastructure density and perceived overcrowding, emphasizing the impact of tourism infrastructure on visitor dissatisfaction. This study advocates for decentralizing infrastructure to underdeveloped regions and improving visitor flow management during peak seasons to address overtourism. The results provide actionable insights for policymakers, offering a framework to guide sustainable tourism strategies and preserve Bali’s cultural and environmental integrity.

Keywords: Overtourism, Binary Classification, Geospatial, Analysis, Sustainable Tourism, Bali, Pearson Correlation