

Detecting Early Signs of Overtourism in Bali: Sentiment and Multiclass Analysis Using BERT and LSTM

Dini Khairina
School of Economics and Business
Telkom University
Bandung, Indonesia
dinikhairina@student.telkomuniversity.ac.id

Herry Irawan
School of Economics and Business
Telkom University
Bandung, Indonesia
herryir@telkomuniversity.ac.id

Abstract—The rapid growth of tourism in Bali has made it one of the destinations facing overtourism, according to the World Travel & Tourism Council (WTTC). This study analyzes tourist perceptions of overtourism in Bali using TripAdvisor reviews from 2018 to 2024. By applying BERT and LSTM models, the study identifies tourist perceptions through sentiment analysis of destinations and multiclass issue classification. The results show that LSTM achieved slightly higher accuracy than BERT, with 77% for sentiment analysis and 92% for overtourism issue classification. Destination sentiment analysis showed an average of 70% positive sentiment, with some destinations recording below-average positive sentiment, reflecting less satisfactory tourist experiences in those areas. The Physical Environment issue was the most frequently discussed topic by tourists, while the highest negative sentiment was linked to the Tourist Numbers issue, indicating discomfort with overcrowding. The strong correlation between Tourist Numbers and Physical Environment shows that environmental pressure increases as visitor numbers grow. These findings highlight the importance of sustainable tourism management to address overtourism challenges and ensure the long-term sustainability of Bali's tourism sector.

Keywords—bali, BERT, LSTM, multiclass classification, overtourism, sentiment analysis.

I. INTRODUCTION

Bali has long been one of the most popular international tourist destinations. In 2023, Bali received the Travelers' Choice Award for Best of the Best Destinations from TripAdvisor, ranking second in the Top Destinations category, ahead of London and Paris [1]. Before the COVID-19 pandemic, tourist visits to Bali reached about 20 million in both 2018 and 2019 [2]. Although the number of visitors dropped significantly to 2.59 million in 2021, it nearly recovered by 2023, reaching 19.53 million tourists [2].

This rapid growth has positioned Bali as one of the destinations experiencing overtourism, according to the World Travel & Tourism Council (WTTC) from January to November 2023 [3]. The overtourism phenomenon in Bali has led to significant challenges, including traffic congestion and environmental damage [4]. Tourist visits are heavily concentrated in the southern part of Bali, including popular destinations like Seminyak, Kuta, Jimbaran, Nusa Dua, Uluwatu, Canggu, Sanur, and Ubud [5], [6], [7], [8], [9], while other areas, such as North, West, and East Bali, see less activity [10]. According to the Indonesian Minister of Tourism and Creative Economy, Sandiaga Uno, South Bali is already nearing overtourism, and a 10% increase in tourist numbers could worsen the situation to the point where recovery becomes difficult, as seen in Barcelona [11].

Although overtourism is an increasingly urgent issue in Bali, research on this phenomenon is still relatively limited. Previous studies, such as those on residents' perceptions of the impacts of overtourism in Ubud [12] and tourism-driven gentrification in Bali [9], have mostly used conventional methods like questionnaires and interviews, which are costly, time-consuming, and prone to human error [13].

Digital data from review platforms like TripAdvisor, which hosted over 1 billion global reviews in 2023 [14], offers a more efficient way to analyze perceptions of overtourism in Bali. Previous studies utilizing digital data in Bali tourism have primarily focused on general aspects, such as destination image analysis through TripAdvisor reviews [15] and tourist movement patterns identified via geotagged photos [16]. However, no research has specifically addressed the overtourism phenomenon in Bali using a comprehensive digital data-driven approach.

This study addresses this gap by analyzing tourist perceptions of overtourism in South Bali, a high-density visitor area. Using TripAdvisor reviews, sentiment analysis and multiclass text classification with BERT and LSTM models are applied to explore tourist sentiment and identify key overtourism issues. The findings aim to enrich the literature on overtourism and provide insights for stakeholders in developing sustainable tourism management strategies.