Advancing tourism demand forecasting in Sri Lanka: evaluating the performance of machine learning models and the impact of social media data integration

Isuru Udayangani Hewapathirana

Abstract

Purpose – *This study explores the pioneering approach of utilising machine learning (ML) models and integrating social media data for predicting tourist arrivals in Sri Lanka.*

Design/methodology/approach – *Two sets of experiments are performed in this research. First, the predictive accuracy of three ML models, support vector regression (SVR), random forest (RF) and artificial neural network (ANN), is compared against the seasonal autoregressive integrated moving average (SARIMA) model using historical tourist arrivals as features. Subsequently, the impact of incorporating social media data from TripAdvisor and Google Trends as additional features is investigated.*

Findings – *The findings reveal that the ML models generally outperform the SARIMA model, particularly from 2019 to 2021, when several unexpected events occurred in Sri Lanka. When integrating social media data, the RF model performs significantly better during most years, whereas the SVR model does not exhibit significant improvement. Although adding social media data to the ANN model does not yield superior forecasts, it exhibits proficiency in capturing data trends.*

Practical implications – *The findings offer substantial implications for the industry's growth and resilience, allowing stakeholders to make accurate data-driven decisions to navigate the unpredictable dynamics of Sri Lanka's tourism sector.*

Originality/value – *This study presents the first exploration of ML models and the integration of social media data for forecasting Sri Lankan tourist arrivals, contributing to the advancement of research in this domain.*

Keywords *Tourism demand forecasting, Social media analytics, Machine learning, Support vector regression, Random forest, Artificial neural network, Sri Lanka*

Paper type *Research paper*

1. Introduction

Accurate forecasting of tourism demand is pivotal in maintaining a sustainable tourism industry. Inaccurate predictions can lead to issues such as over- or under-supply of essential services like food, accommodations and infrastructure within the destination, disrupting the delicate balance of the tourism ecosystem (Zhang *et al*., 2021). Tourism demand in Sri Lanka is inherently volatile due to external factors such as geopolitical events and economic fluctuations in key source markets. As illustrated in Figure 1, unexpected events like the 2019 Easter Sunday explosions and the economic impacts of the COVID-19 pandemic have exerted adverse effects on tourist arrivals to the country. These sudden disruptions emphasise the need for sophisticated forecasting models robust to unexpected changes.

Isuru Udayangani Hewapathirana is based at the Faculty of Science, University of Kelaniya, Kelaniya, Sri Lanka.

Received 26 June 2023
Revised 24 August 2023
13 October 2023
Accepted 11 November 2023

The author extends her sincere gratitude to the reviewers for their valuable feedback and acknowledges the significant contributions made in improving the manuscript. The author also acknowledges the financial support received for this study by the University Research Grant RP/03/02/09/01/2022 from the University of Kelaniya. The support provided by this grant has been instrumental in conducting and completing the research.