Current Issues in Tourism

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rcit20

Tourism and uncertainty: a machine learning approach

Athanasia Dimitriadou, Periklis Gogas & Theophilos Papadimitriou

CURRENT ISSUES IN TOURISM
https://doi.org/10.1080/13683500.2024.2370380

RESEARCH ARTICLE

Tourism and uncertainty: a machine learning approach

Athanasia Dimitriadou[a], Periklis Gogas[b] and Theophilos Papadimitriou[b]

[a]College of Business, Law and Social Sciences, University of Derby, Derby, UK; [b]Department of Economics, Democritus University of Thrace, Komotini, Greece

ABSTRACT

In this paper, we attempt to create a unique forecasting model to forecast out-of-sample the tourism demand in 24 European Union countries. The initial dataset included 34 relevant variables of annual frequency that span the period from 2010 to 2020 for 40 countries. A data prefiltering process resulted in a final set of 17 relevant variables for 24 countries. Additionally, in the effort to investigate the impact of uncertainty on international tourism, apart from the traditional factors that affect tourism, we also include variables that measure various forms of uncertainty: we use the World Pandemic Uncertainty (WPU) Index, the Global CBOE Volatility Index, the Political Globalisation Index, the Economic Globalisation Index, and the Political Stability Index. In the empirical part of our research, we employ and compare in terms of their forecasting accuracy a set of six state-of-the-art machine learning algorithms, the Support Vector Regression with both a linear and an RBF kernel, the Random Forests, the Decision Trees, the KNN, and gradient-boosting trees. The results show that the Gradient-Boosting Trees algorithm outperforms the other five models providing the most accurate forecasts with a MAPE of 0.10% and 1.36% in the training and the out-of-sample tests, respectively.

ARTICLE HISTORY
Received 6 January 2024
Accepted 13 June 2024

KEYWORDS
Tourist arrivals; tourism demand; uncertainty; machine learning; prediction; forecasting

1. Introduction

In the complex world of global tourism, uncertainty has steadily emerged as a focal point of academic and industry research. As tourism weaves an intricate tapestry of economic, cultural, and social dynamics, the disruptions introduced by unpredictable factors become central to comprehending fluctuations in tourist demand. Accurately decoding the interplay between uncertainty and tourist demand opens doors to robust strategies that safeguard the sector against unforeseen disruptions. The tangible benefits of such an understanding extend far and wide. For travellers, it promises experiences that are insulated from geopolitical, economic, or environmental volatilities. Meanwhile, for the industry, it offers an analytical lens, acting as a Business Intelligence compass to navigate the intricate seas of global tourism. A comprehensive analysis of uncertainty's impact can empower stakeholders to forecast travel patterns, make informed infrastructure investments, and shape policies that cultivate a resilient and thriving tourism ecosystem.

Numerous studies have underlined the impact of political and economic turmoil on tourism. For instance, Neumayer (2004) emphasised that political unrest and economic instability could deter potential tourists, even if these disturbances occur in regions geographically removed from their chosen destination. A real-world illustration of this can be drawn from the 2016 coup attempt in Turkey. Moreover, a recent scholarly investigation conducted by Sharma and Khanna (2023), delve into the relationship between global economic policy uncertainty (GEPU) and tourist arrivals

CONTACT Athanasia Dimitriadou nancy.dimitriadou@gmail.com

across a panel of 19 countries. Their findings suggest that changes in trade, monetary and fiscal policy in important countries induce short-term uncertainty. Tourism exhibits a unique resilience against policy uncertainty, unlike the trade and investment sectors.

However, the broader implications of tourism extend beyond the industry. Governments, recognising the substantial economic contributions of tourism, are deeply invested in understanding factors that influence tourist flows and the demand responsiveness to these dynamics. The World Tourism Organization (WTO) underscores the scope of global tourism with statistics indicating an international tourist as someone venturing on temporary visits across international frontiers, residing from anywhere between 24 h to less than a year. While tourism's robust development in various destinations signals its economic vitality, it' is equally important to acknowledge the significant costs some countries have borne in its wake. The intertwined relationship between political-economic uncertainty and tourist demand, thus, accentuates the need for accurate demand predictions and the sculpting of sagacious tourism policies.

The dynamic and varied industry of tourism is intricately entwined with a wide range of world events and phenomena. Over the time period of 2009–2020, this sector faced a whirlwind of uncertainties shaped by economic downturns, political disruptions, health emergencies, and ecological challenges. In light of this, the International Monetary Fund (IMF) introduced the World Uncertainty Index (WUI), a pivotal tool that captures global uncertainties over time. Notably, the WUI highlights significant spikes during pivotal events such as the 9/11 attacks, the SARS outbreak, the Lehman Brothers collapse, the European debt crisis, El Niño, the Brexit referendum, and the evolving US–China trade tensions. A graphical representation of the WUI can be seen in Figure 1. As observed by Ahir et al. (2022), these uncertainties often exhibit synchronous patterns, particularly among advanced economies and euro-area nations.

These events, while diverse in nature, demonstrated the unpredictability of the patterns of global tourism demand. Forecasting tourism demand during such unstable times is not just a theoretical exercise; it is an essential tool for decision-makers in government, commerce, and other sectors to navigate the volatility. Machine learning (ML) is quickly becoming a significant tool in the forecasting

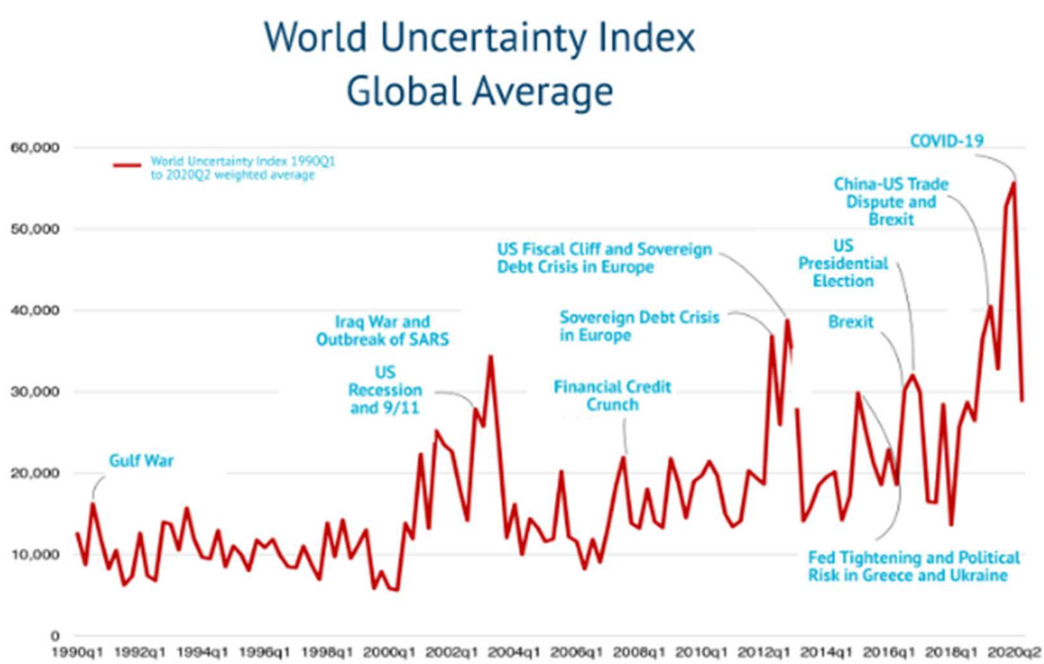

Figure 1. Word Uncertainty Index. Source: Ahir et al. (2022).

arsenal in the era of digitisation. ML has the potential to offer insight into tourism demand as it can extract complex patterns and relationships from large datasets, especially in uncertain situations. However, the efficacy of machine learning models can vary, necessitating a comprehensive comparative analysis (Sulong et al., 2023).

We contribute to the literature on tourism demand in several significant ways. First, we forecast the tourist demand amidst global uncertainties between 2009 and 2020, specifically focusing on domestic tourist arrivals in 24 countries. Through conducting this empirical analysis, we aim to reveal the impact of the global uncertainty events on tourist arrivals in 24 EU countries, such as the World Pandemic Uncertainty Index (WPU) and the Political Stability Index. This investigation allows us to understand the complicated relationship between the uncertainty events that shape these indices and the subsequent combinations in aggregate tourism demand.

Secondly, whereas many previous studies relied solely on either traditional econometric methods (Wu et al., 2021) or various machine learning techniques (Pereira & Cerqueira, 2022; Sulong et al., 2023), our approach is distinct. In this study, we adopt a novel mixed-methodology approach by integrating a variety of machine learning models with conventional econometric regression techniques, as developed by Sofianos et al. (2024). Furthermore, we leverage advanced techniques such as the coarse-to-fine grid-search cross-validation and the variable importance measure to not only identify the best tuning parameters but also determine the most important variables in predicting tourism demand amidst diverse global uncertainties.

Third, the high forecasting accuracy of the models in our study can be valuable to the tourism industry such as hotels, travel agencies, and airlines, allowing these entities to allocate resources more efficiently and, thus, protect their profitability, devise efficient flight schedules, arrange appropriate logistics, and improve customers' needs. Finally, while most of the previous research has used annual or aggregated data individually at a national level, our study utilises a rich and detailed dataset, covering domestic tourist arrivals across 24 EU countries from 2009 to 2020.

From our empirical analysis, the Gradient-Boosting Trees (GBT) model outperformed the other models in all three forecasting accuracy metrics in both the training and the out-of-sample tests. The GBT produced a MAPE of 0.10% and 1.36% in the training and the out-of-sample tests, respectively. This finding not only complements the broader predictive analytics literature, which highlights the effectiveness and competitiveness of ensemble methods like Gradient-Boosting Trees for predictive tasks, employed for both regression and classification (Friedman, 2001), but also presents tangible implications for the tourism sector. Especially, during times of high global uncertainty, tourism stakeholders – from governments to businesses – can use the Gradient-Boosting Tree model's predictive power to improve policy design and implementation in the tourism sector. Given the demonstrated accuracy of the machine learning models in our empirical analysis, we present a strong case for its wider industrial adoption. Finally, the ability to identify tourists' arrival trends one year earlier allows (a) for prompt and effective strategic decision-making from governments and other stakeholders, and (b) for optimal allocation of resources. This has major policy implications, allowing policymakers and industry stakeholders to develop better long-term business plans and effectively respond to expected changes in demand.

The rest of the paper is organised as follows. Section 2 provides a review of the relevant literature. Section 3 describes the data. Section 4 outlines the methodology that we use in our study. Section 5 presents our empirical results. Finally, Section 6 presents conclusions, policy implications, and recommendations for future research.

2. Literature review

The global tourist industry's inherent dynamism is frequently a result of its sensitivity to external factors. Understanding the uncertainties in this area is not just a matter of academic study; it also helps stakeholders create plans that will increase resilience and ensure adaptation in the face of changing conditions (Chase et al., 2023). Building on this context of global events, there arises a natural query about the symbiotic relationship between tourism and broader economic parameters. Numerous studies have explored this link, especially between tourist demand and economic growth across both developed and developing countries (Adnan Hye & Ali Khan, 2013; Cannonier & Burke, 2019; Kreishan, 2010; Scarlett, 2021; Skerritt & Huybers, 2005). Most of this research has found a positive association between tourism and economic growth, which is equally important across large and small countries (Sequeira & Maçãs Nunes, 2008).

With the established significance of tourism with respect to economic growth, the industry's next challenge is forecasting. Over the past decades, there has been a proliferation of methodologies for forecasting tourism demand. Techniques such as support vector regression (Chen et al., 2015), artificial neural networks (Pattie & Snyder, 1996), and Bayesian networks have been applied by researchers. Time series models, including those advocated by Kulendran and King (1997) and Lim and McAleer (2002), also remain central to tourism forecasting. The Gravity Model, which is frequently promoted in academic circles, places an emphasis on how useful distance and population density are for forecasting tourist flows. Guo (2007) computes the gravity model to evaluate the influx of tourists to China, while Khadaroo and Seetanah (2008), leverage it to assess the impact of transportation infrastructure on tourism movements.

In a comprehensive meta-analysis, Peng et al. (2014), explored the relationship between the accuracy of various forecasting models, associated data attributes and study characteristics. Through assessing 65 publications from 1980 to 2011, their meta-regression analysis revealed that factors such as tourist origin, destination type, time duration, modelling technique, data frequency, variable counts, their measurements, and sample size significantly determine the forecasting model's precision. This pioneering work offers insights into selecting suitable forecasting methods tailored to specific forecasting contexts in tourism. Moreover, building on this foundational research, Peng et al. (2015) conducted a recent meta-analysis investigating the influences of several factors on the estimated international tourism demand elasticities. Reviewing 195 studies published between 1961 and 2011, they found that a number of factors had a substantial impact on the estimated demand elasticities, including some of the origin and destination of tourists, the time period analysed, the modelling techniques utilised, the frequency of data and more.

Chen et al. (2024) forecast tourism demand during the Covid-19 pandemic. Their research introduced a unique COVID-19 impact indicator to quantify the pandemic's influence on tourism. In addition, they developed a forecast aggregation algorithm designed to optimise predictions with minimal post-pandemic data. The results of this empirical analysis confirmed the effectiveness of these strategies, showing marked improvements in forecast accuracy and consistency.

Li et al. (2024), presented a new tourism demand forecasting system grounded in an advanced decomposition algorithm. This system first break down the main data into various sub-series, and then employs forecasting models tailored to the characteristics of each sub-series. Using monthly tourist arrival data from Hong Kong, sourced from six countries, they gauged the efficiency of their framework. Their method consistently outperformed conventional models, highlighting the potential of their decomposition approach in forecasting scenarios. Another recent study predicting the international arrivals to Hong Kong has been conducted by Hu et al. (2022). They obtain tourist-generated online review data related to attractions, accommodations, and shopping venues in their demand forecasting system for seven English-speaking countries. The findings suggest that mixed-data sampling (MIDAS) models showed greater efficiency than traditional time series models, when combined with high-frequency electronic review data.

Pattie and Snyder (1996) undertook research that compared traditional time-series forecasting techniques against the more innovative neural network model. Their analysis, which underscored the importance of data integrity, accuracy, and the application of suitable error metrics, utilised a dataset from the US National Park Service. Their investigation was primarily focused on comprehending the operational complexities of forecasting neural networks. Their findings revealed that both the Census II decomposition and the neural network technique emerged as the most precise methods for 12-month-ahead predictions. In a similar vein, but with a different geographical focus, Burger et al. (2001), investigated projecting tourism demand, specifically US demand for travel to Durban, South Africa. Their study spanned various forecasting techniques, from traditional ones like naïve, moving average, decomposition, single exponential smoothing, and ARIMA to more unconventional methods like genetic regression and neural networks. Their results suggested that the neural network method was, once again, recognised as the superior forecasting approach.

Reflecting on the broader sphere of forecasting, Vapnik (1995), created a statistical learning algorithm – Support Vector Machines (SVR) – in the mid-1990s that follows the principle of structural risk minimisation, by attempting to minimise the upper bound of the generalisation error rather than the training error. Augmenting this lineage of advanced methodologies, Chen et al. (2015), predicted the holiday tourist demand flow by applying an approach which hybridises the SVR model with adaptive genetic algorithm (AGA) and the seasonal index adjustment (S), namely AGA-SSVR. The study used daily tourist flow data from 2008 to 2012 for Mountain Huangshan in China. The findings display that the AGA-SSVR model is an effective more accurate approach with a MAPE of 0.1182 than the other alternative models including AGA-SVR and back-propagation neural network (BPNN), which have a MAPE of 0.1479 and 0.2319, respectively.

Expanding upon these forecasting methodologies, research studies have shown the efficacy of machine learning approaches in predicting tourism demand across various domains. Pereira and Cerqueira (2022) conducted a comprehensive investigation, employing 22 different methods, comparing machine learning methods with more traditional forecasting techniques, in an effort to predict hotel demand at the very short. Emphasising real-time, time series data on the daily demand for four-star hotels in southern Europe, their findings show that machine learning methods outperformed traditional forecasting techniques. In particular, compared to traditional techniques, machine learning models were found to reduce the root mean squared error by up to 54% for the 1-day ahead forecast horizon, and up to 45% for the 14-day forecast horizon. Similarly, Sulong et al. (2023), proposed a machine learning approach to predict Halal Tourism Demand (HTD) and Halal Tourism (HT) firms' financial performance. Utilising internet data, specifically, Twitter and Google Trends data, the authors employed two models with 14 machine learning algorithms. Their findings suggested the efficiency of the bagged Classification and Regression Trees (CART) model, achieving an R^2 of 93.71% for HTD forecasting. Similarly, the bagged CART emerged as the optimal model for HT firms' financial performance forecasting, achieving an R^2 of 80.12%. With the increasing intricacies of the tourism industry, various uncertainty indices, such as the World Pandemic Uncertainty (WPU) Index, Economic Policy Uncertainty Index (EPUI) and Global CBOE Volatility Index, have been developed and incorporated in research to gauge the influence of global uncertainties on tourism demand. Ongan and Gozgor (2018), focused on the EPUI as a critical predictor. Analysing data from the first quarter of 1999 to the first quarter of 2015, they discovered that a one standard deviation increase in this index corresponds to a 4.7% drop in the number of Japanese tourists visiting the USA in the long term. Similarly, the intricate relationship between political instability, terrorism, and tourism has been a subject of interest for researchers.

Saha and Yap (2014), embarked on a panel analysis, scrutinising data from 139 countries between 1999 and 2009. Their findings revealed a counterintuitive observation: terrorist attacks seemed to boost tourism demand in nations with low to moderate political risks. However, countries plagued with substantial political threats often suffer significant blows to their tourism industries.

The complex mechanics of tourism demand do not work in isolation. While economic policy uncertainties and political instabilities influence tourist decisions significantly, global economic crises and subsequent policy responses have consistently remained at the forefront of affecting tourism. Boukas and Ziakas (2013), provided light on this by performing a qualitative study on the effects of the Global Economic Crisis on the tourism sector in Cyprus. The study took an immersive approach, including semi-structured interviews with key stakeholders, including tourism officials and suppliers. Their analysis found certain significant characteristics that impacted Cyprus's tourist landscape. These included a notable lack of competition, a significant decline in visitation and associated income, issues of poor quality, and, most importantly, rising prices, which appeared to be a source of discomfort for many potential tourists.

Beyond the conventional confines of financial downturns and their effects on tourism between 2007 and 2010, Hall (2010) brought to light additional unanticipated disasters, such as natural disasters. The 2010 Icelandic volcanic plume, for example, created considerable obstacles to global tourism. Along with the effect of pandemics in disrupting the routine flow of tourism, the overhanging shadow of potential future disasters was discussed. While the literature seems to be predominantly skewed towards examining economic and financial crises in the context of tourism, it would be a disservice to overlook other significant determinants (Kumar & Sanjeev, 2020; Sönmez, 1998).

Moreover, governments can play a significant role in becoming more efficient in resource allocation geared towards intensifying and diversifying the sector. Geopolitical risks, encompassing international conflicts, political instability, and terrorism, can significantly impact tourist demand (Wujie, 2023). Tourists are often deterred from visiting destinations perceived as high-risk, causing sharp declines in visitor numbers (Blake & Sinclair, 2003). Delving deeper into this, Tiwari et al. (2019), uncovered a compelling insight. Their discoveries suggest that the tourism industry is more sensitive to geopolitical risks than to economic policy uncertainty. Furthermore, the lasting impact of these risks is quite disparate. Geopolitical risks, they argued, have long-run implications for the sector, casting prolonged shadows over tourist arrivals. On the other side, economic policy uncertainties, though impactful, predominantly exert their influence in the short run. This difference in the temporal spread of consequences is pivotal for tourism policymakers and stakeholders.

3. The data

Most of the data for this study were obtained from TheGlobalEconomy.com. The data are of annual frequency and span the period from 2010 to 2020. We collected 34 relevant variables over the sample period for 40 countries. However, we had to eliminate any variable with missing values as the methodologies we use require no missing values in order to work properly and efficiently. Next, we tried to overcome the problem of multicollinearity. This problem makes it difficult to determine the true underlying relationship between the variables. Thus, we used two techniques: We first performed a correlation analysis on the independent variables in order to identify the ones exhibiting high correlation. While this technique helped identify and eliminate the variables with the highest correlation, it did not fully address the issue of multicollinearity in our model. Following that, we applied the Variance Inflation Factor (VIF),[1] a more robust technique (Thompson et al., 2017). Using the VIF, we identified and removed the variables with VIF values greater than 10, reducing multicollinearity and increasing the reliability of our results. This prefiltering procedure left us with 16 independent variables and 264 observations for the 24 European Union (EU) countries (Luxembourg, Sweden, Finland, Malta, Denmark, Portugal, the Netherlands, Hungary, Germany, Croatia, Latvia, Italy, Spain, the United Kingdom, Romanian, Bulgaria, France, Serbia, Moldova, Bosnia and Herzegovina, Russia, Belarus, Turkey, and Ukraine).

The number of annual tourist arrivals per country is the dependent variable that we attempt to forecast. The list of possible predictors used includes a variety of 17 variables, categorised by three main groups: Macroeconomic Variables, Political and Environmental Variables, and Financial Market Variables. The macroeconomic factors reflect the overall economic health and stability of the respective country, which can have an important impact on tourism demand (Martins et al., 2017). Moreover, the political stability and the globalisation metrics are crucial for promoting tourism, as they may cause a decrease/ increase in tourism arrivals due to safety concerns or global economic activity and society (Valentinas et al., 2022; Xu & Lv, 2023).

The full list of all the variables, their description and their descriptive statistics are presented in Table 1. The majority of the data, including the number of international inbound tourist arrivals (dependent variable), economic growth, inflation, government sector metrics, and globalisation indices, were obtained from TheGlobalEconomy.com. Furthermore, we included exchange rates

Table 1. Data set and descriptive statistics.

No	Variable	Description	Std. dev.	Mean	Skewness	Kurtosis	Source
1.	Tourist arrivals (thousands)	Number of international inbound tourists	43,956	30,075	2.63	7.27	Global Economy.com
2.	Economic growth	Annual percentage growth rate of GDP at market prices	3.45	1.57	−1.19	4.35	Global Economy.com
3.	Inflation	Percent change in the Consumer Price Index	6.53	3.59	5.57	39.87	Global Economy.com
4.	Unemployment rate	The share of the labour force that is out of work but available and looking for work.	5.33	9.30	1.70	2.63	Global Economy.com
5.	Government spending, % of GDP	General government final consumption	3.51	19.55	0.29	−0.65	Global Economy.com
6.	Fiscal balance, % of GDP	Government Revenue minus Government Expenditure	3.03	−2.68	−0.62	0.24	Global Economy.com
7.	Government debt, % of GDP	The total stock of direct government fixed-term contractual commitments to others that are outstanding as of a specific date.	38.62	57.69	3.97	34.37	Global Economy.com
8.	Political Stability Index	The likelihood that the government will be destabilised or overthrown by unconstitutional means.	0.75	0.33	−0.99	0.83	Global Economy.com
9.	Heating Degree Days	A metric used to track energy use.	2569.76	5888.99	1.34	3.01	WorldBank.org
10.	Fossil_CO_2 per GDP	CO_2 emission totals of fossil fuel use and industrial processes	0.13	0.23	1.18	0.23	European Commission
11.	GBP/EUR in euro	The value of 1 British Pound in Euro	0.06	1.20	1.27	0.14	Yahoo Finance. com
12.	EUR/CHF in CHF	The value of 1 Euro in Swiss Francs	0.11	1.19	1.40	1.88	Yahoo Finance.com
13.	EUR/SEK in SEK	The value of 1 Euro in Swedish Kronor	0.62	9.49	0.24	−1.36	Yahoo Finance.com
14.	Economic Globalisation Index	Actual economic flows and restrictions to trade and capital	10.66	73.82	−0.48	−0.72	Global Economy.com
15.	Political Globalisation Index	Number of international organisations, international treaties and treaty partner diversity	10.76	87.22	−1.02	−0.11	Global Economy.com
16.	Global CBOE Volatility Index (^VIX)	Real-time index which indicates the market's expectations for the relative strength of the S&P 500 Index (SPX) short-term price fluctuations.	4.26	17.84	0.19	−0.95	Yahoo Finance.com
17.	World Pandemic Uncertainty Index (WPU)	The percent of the word 'uncertain', and its variants, that appear near the pandemic terms in EIU country reports, multiplied by 1000	0.11	0.08	1.85	2.67	Federal Reserve Banks of St. Louis (fred.stlouisfed.org)

and the Global CBOE Volatility Index (^VIX) provided by Yahoo Finance, to illustrate the influence of the financial market dynamics on tourism demand. Environmental factors have significant influence on the formation of tourists' activities. Therefore, we used the Global fossil CO2 emissions per unit of GDP (Fossil_CO2) and Heating Degree Days sourced from the European Commission and World Bank, respectively. These factors include environmental trends that could affect tourist travel behaviours, such as a desire for warmer or colder climates for leisure (Falk & Lin, 2018).

Finally, in order to evaluate the impact of global uncertainty events, we included the World Pandemic Uncertainty Index (WPU) in our research. This index, compiled by the Federal Reserve Banks of St. Louis as it was constructed by the Economist Intelligence Unit (EIU), provides valuable insight into the level of uncertainty surrounding pandemics in different countries.

According to the above descriptive statistics, the target variable, Tourist Arrivals exhibits a positive Skewness (2.63) and high Kurtosis (7.27), indicating a heavy-tailed distribution. Following the criteria outlined by Darren and Mallery (2010), Hair et al. (2010), and Byrne (2010) for normal distribution, variables such as Heating Degree Days, Political Stability, Government Spending, Fiscal Balance, Economic Globalisation Index, Political Globalisation Index, GBP/EUR in euro, EUR/CHF in CHF, and the Global CBOE Volatility Index (^VIX) have normal distributions with acceptable skewness and kurtosis. Economic indicators with considerable positive Skewness and extremely high Kurtosis, such as Inflation and Government Dept, indicate volatility in the data due to uncertainty in the economy.

4. Methodology

4.1. Support vector regression (SVR)

Support Vector Regression (SVR) is a technique that has its roots in the principles of Support Vector Machines (SVM), a popular machine learning method known for its use in classification problems. The foundational SVM algorithm was introduced by Boser et al. (1992) and was later refined by Cortes and Vapnik (1995), emerging from the realm of statistical learning theory.

While SVM is used for classification tasks, SVR is its adaptation for regression problems. The main objective of SVR is to find a function that approximates the given data points such that the deviation of this function from the actual data values is minimised. However, not all deviations are considered critical. SVR operates on the idea that deviations below a certain threshold, denoted by ε (epsilon), are acceptable and do not contribute to the overall error. This means that SVR aims to fit the data in a manner where errors that fall within this ε band are not penalised. Only deviations exceeding this threshold will incur a penalty.

This tolerance band, defined by ε, is crucial to SVR's functionality. The data points that reside on or outside the boundaries of this ε-tolerance band are called 'Support Vectors'. They play a central role in defining the linear regression, as the algorithm primarily focuses on these points and not on those that fall within the band. It uses a combination of linear and non-linear kernel to transform the input space, enabling the derivation of optimal linear regressions. In our models, we test two kernels: the linear and the non-linear (Radial Basis Function – RBF). The mathematical representation of each kernel is:

$$\text{Linear: } K_1(x_1, x_2) = x_1^T x_2, \tag{1}$$

$$\text{RBF: } K_2(x_1, x_2) = e^{-\gamma x_1, x_2^2}. \tag{2}$$

The construction of the model unfolds in two phases: the training sample and the test sample. During the training sample, the bulk of the dataset is employed to determine the Support Vectors that delineate the band. In the test sample, we assess the model's adaptability by examining its efficacy on the minor subset that was reserved during the training sample. By employing cross-validation methods, we obtain a solution that's broad-based and not merely tailored to a specific sample, mitigating the risk of overfitting.

4.2. Decision trees

In the field of data mining, particularly for pattern classification tasks, decision trees and neural networks often emerge as top contenders. Both are celebrated for their efficacy and accuracy. However, in terms of interpretability, decision trees outperform neural networks and are thought to be faster and easier to implement. A decision tree is computed by posing questions related to various attributes at each node. In contrast to classification trees, which generally reflect binary or discrete values at the leaves, in regression trees the end of nodes or leaves represent clusters of instances that share a continuous numeric value. To predict the value of an unknown instance, it's guided down the tree, following paths determined by its attributes. Upon reaching a leaf, the instance is determined based on the mean of that leaf (value). Ideally, leaves (predicted values) should be as homogenous as possible in their classifications.

4.3. Random forest models

The random forest algorithm is a sophisticated ensemble technique rooted in the principles of decision trees. Enhanced by bootstrapping and aggregation procedures, it generates a diverse collection of individual regression systems (Breiman, 1997). This method stands out in the academic realm, with numerous scholars like Lang et al. (2021) and Mishina et al. (2015), highlighting its efficacy in circumventing the overfitting issues often associated with singular decision trees. In essence, a random forest is a confluence of multiple decision trees, each cultivated from a distinct set of features chosen at random from the primary dataset.

An interesting aspect of the random forest methodology is the use of the 'out-of-bag' (OOB) set, which comprises observations excluded during the bootstrapping phase. This OOB set offers a practical means to gauge the model's generalisation capability. Further adding to its robustness, each decision tree within the random forest is designed to operate on a randomised subset of explanatory variables or features. Typically, the number of features chosen is the square root of the total available. Ultimately, the predictions rendered by individual trees are amalgamated, with the mean value emerging as the final prediction.

4.4. Gradient-boosting trees

The gradient-boosting tree (GBT) model, introduced by Breiman's (1997) insight that boosting can be viewed as an optimisation algorithm applied to a suitable function, stands out as a premier machine learning technique, especially adept at modelling nonlinear relationships between a target variable and its predictors. Gradient-boosting technique is used to deal with missing values, outliers, and high cardinality categorical values on the features without any special treatment. Gradient boosting is one of the variants of ensemble methods where you create multiple weak models and combine them to obtain superior performance (Wang et al., 2021). The weak models are the individual's decision trees, which are connected in series and each tree tries to minimise the error of the previous tree. Due to this sequential connection, boosting algorithms are usually slow to learn, but also very accurate. Weak models are adjusted in such a way that each new model fits into the residuals of the previous step as the model is improved. By combining the outcomes of every step, the final model produces a powerful learner. A loss function is used to detect the residuals. It is important to note that adding a new tree to the model has no effect on the existing trees. The additional decision tree fits the residuals from the previous model.

In random forests, the addition of too many trees won't cause overfitting. While the accuracy of the model doesn't improve beyond a certain point, no overfitting issues are faced. On the other hand, in gradient-boosting trees having too many weak learners in the model may lead to overfitting of data. Therefore, gradient-boosting trees require careful tuning of the hyperparameters.

4.5. K-Nearest neighbours (KNN) model

KNN is one of the most fundamental algorithms in the machine learning domain developed by Fix and Hodges (1955) and then expanded by Cover and Hart (1967). It operates by identifying the k closest data points with known target values when faced with an input whose target is uncertain. Typically, the forecasted target value is calculated as either the average or the median of these neighbours. Applying KNN regression to univariate time series data entails using past time series values as features, where the aim is to predict a historical time series value based on previous studies (Suhel & Bashir, 2018; Tajmouati et al., 2024). Given that KNN is sensitive to the scale of the data, we use a process called feature scaling. This means that we adjust the scale of the different features in our data. We do this by making sure that each feature has a mean of zero and a standard deviation of one in the training data. This ensures that all the features are on a similar scale. This adjustment is important because KNN works by measuring distances between data points. When the features are on the same scale, it helps the KNN perform better and provide more accurate results.

4.6. Feature scaling

Feature scaling is a vital step in preparing data for Machine Learning models. Feature scaling is used to standardise the range of independent variables or features in the data. Many machine learning algorithms compute distances between data points, so they are sensitive to features being on different scales. Having features on a similar scale can improve the performance of these algorithms. Standardisation transforms the feature into a distribution with zero mean and unit variance using the formula:

$$z = \frac{x - \mu}{\sigma}, \tag{3}$$

Where z is the standardised value, x is the original feature value, μ is the mean of the feature, and σ is its standard deviation.

4.7. Over-fitting

Cross-validation is a technique employed during the training phase to mitigate overfitting. This concern is described in the literature as the 'low bias-high variance' (Mehta et al., 2019; Russo & Zou, 2020) problem, or the bias-variance trade-off. Essentially, it involves splitting the training data into n subsets. The training process begins with a chosen initial setup where the model is trained on $n-1$ of these subsets, reserving one subset for validation. This process is iteratively performed, with a different subset reserved for validation each time. The model's in-sample accuracy is then determined by averaging the forecast results across all n subsets. To refine the model further and reduce prediction error, this iterative training process is repeated with different parameter configurations. This strategy is termed 'n-fold cross-validation'. In our study, we employed a 3-fold cross-validation approach, a representation of which is illustrated in Figure 2. To evaluate the model's ability to generalise to new, unseen data, its out-of-sample accuracy is tested using 20% of the dataset that was excluded from the cross-validation-based training process (Figure 3).

4.8. Performance metrics

Three metrics are used to evaluate the effectiveness of forecasting approaches: the Mean Absolute Percentage Error (MAPE), the Root Mean Squared Error (RMSE), and the R^2. The MAPE is a measure for regression models which provides the error as a percentage, allowing for a more accurate estimate of

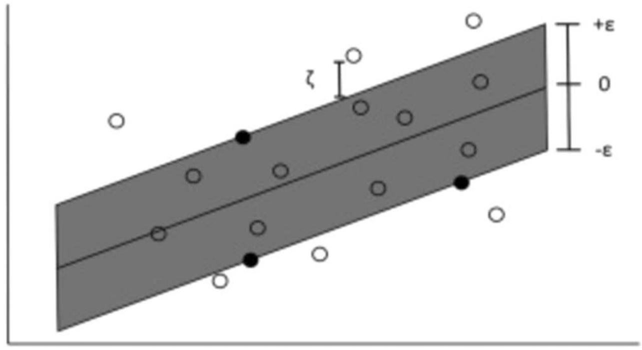

Figure 2. Support vector regression. The upper and lower error tolerance threshold are presented by the letter ε. The black filled points define the boundaries of the error tolerance band, Support Vectors (SVs). Forecasting values greater than ε incur a penalty according to their distance from the tolerance allowed band.

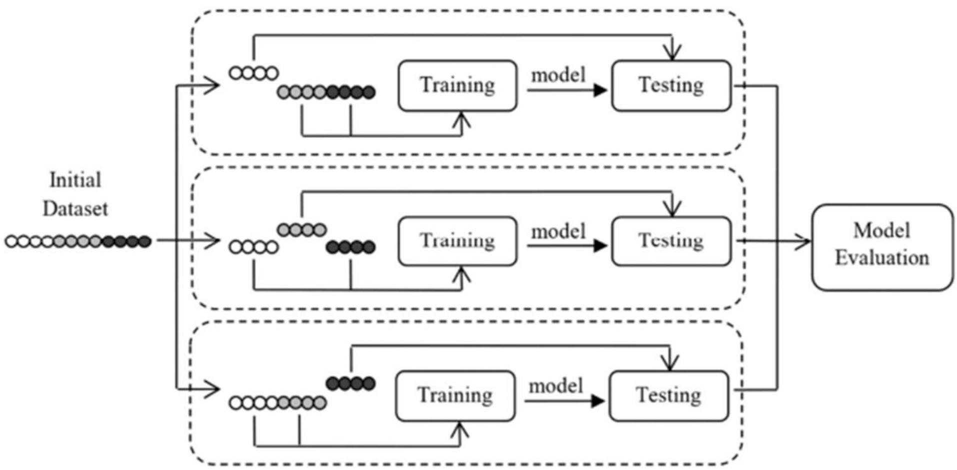

Figure 3. A three-fold cross-validation training method in overview. It demonstrates that each fold serves as a testing sample, while the remaining folds are utilised to train the model for each possible combination of the parameters' values.

error and it offers a scale-independent view of the error. This is especially beneficial in situations where it is critical to understand the extent of the error in relation to the real value. Moreover, RSME is the square root of the mean square error, which makes the scale of errors equal to the scale of targets. The R^2 calculates what percent of the variance in the target (dependent) variable is explained by the forecasting model.

They are calculated in the following standard way:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}, \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{yi})^2}. \tag{5}$$

Where n is the number of observations, y_i is the real value, and $\hat{y}_i$ is the predicted value. The smaller the two criteria are, the more accurate the forecast of tourism demand. The smaller the two criteria are, the more accurate the forecast of tourism demand.

Similarly, the R^2 is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{6}$$

where $\hat{y_i}$ is the model's forecasted value and y_i represents the actual value of the dependent variable. Also, $\bar{y}_i$ represents the mean of all values and thus, the denominator represents the total sum of squares (TSS), which represents the total variance in the dependent variable. In contrast, the numerator calculates the sum of squared residual error (SSE), which denotes the difference between the true values and forecasted values, and thus, represents the variation that the regression model is unable to explain.

5. Empirical results

The study's primary goal was to develop the optimal (more accurate) single model to out-of-sample (OOS) forecast the aggregative tourism demand for the 24 countries. This was achieved by training and testing six alternative machine learning algorithms, namely, Support Vector Regression with both a linear and an RBF kernel, Random Forests, Decision Trees, K-Nearest Neighbours, and gradient boosting.

The first step was to implement a feature scaling technique to standardise the range of all the independent variables in our dataset. This step ensures that each feature has the same scale, making gradient descent algorithms converge more quickly and distance-based algorithms, like KNN, to be more accurate. Feature scaling is essential since unscaled features with larger ranges can disproportionately influence the model, rendering it biased or leading to suboptimal performance. In the training step of all our models we employed a 3-fold cross-validation procedure to avoid over-fitting, and a grid-search coarse-to-fine technique to explore a wide range of parameters values and optimise the models' hyperparameters. The hyperparameters for each model were meticulously fine-tuned to yield the best predictive outcomes.

Figures 4–9 provide a simple visual evaluation of the quality of the forecasts. They depict with blue dots a scatter plot of the predicted (on the vertical axis) against the actual (on the horizontal axis) values. The better the forecasting model, the closer the dots are to the 45-degree red line. In these figures, we are only depicting the out-of-sample (OOS) subset ('out-of-bag' set for the Random Forest model) for each algorithm. According to these, the Support Vector Regression models, with the RBF and the linear kernel in Figures 4 and 5, appear to have a very poor performance in forecasting tourist arrivals as they seem to lie consistently below the reference red line, indicating a systematic tendency to underestimate the actual value of tourist arrivals. In contrast, the rest of the models in Figures 6–9 exhibit a significantly better forecasting performance, as evidenced by their predicted points closely aligned with the red line. Thus, the predicted values closely approximate the actual values. In Figure 7 and 9 we can see that although the accuracy is high, the models – KNN and RF respectively – seem to underestimate the actual extremely high values of tourist arrivals.

On the other hand, the decision trees and the gradient-boosting models seem to provide a high fit consistently throughout the whole range of values. This is evident in Figure 10, where we plot the forecasted tourism demand values from the Gradient-Boosting Tree Model (red dashed line with circles) against the actual values (black line with circles). It is evident visually, that the predicted values closely correspond to the actual ones, providing evidence of the quality of fit of our best trained model.

It is important to remind the reader here, which all Figures 4–10 refer to the out-of-sample part of our dataset that is unknown to our models.

The dots represent the actual/true data points in our dataset, while the red line represents the fitted line which minimises the overall distance between the dots and the predicted values on the line (Figures 4–9).

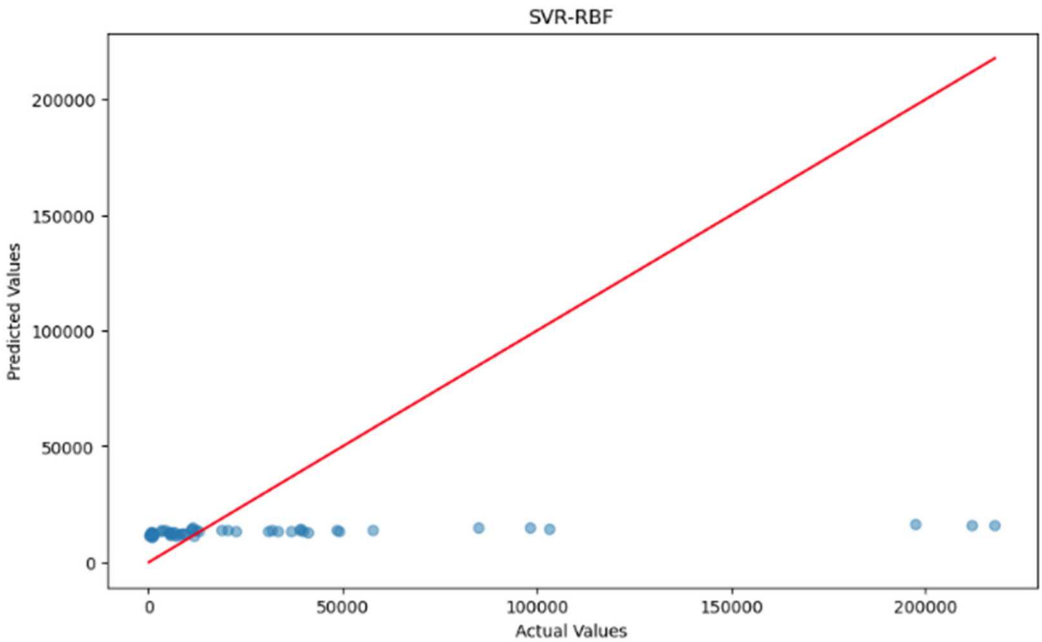

Figure 4. Support vector regression with RBF kernel ($C = 245$, gamma = 0.063).

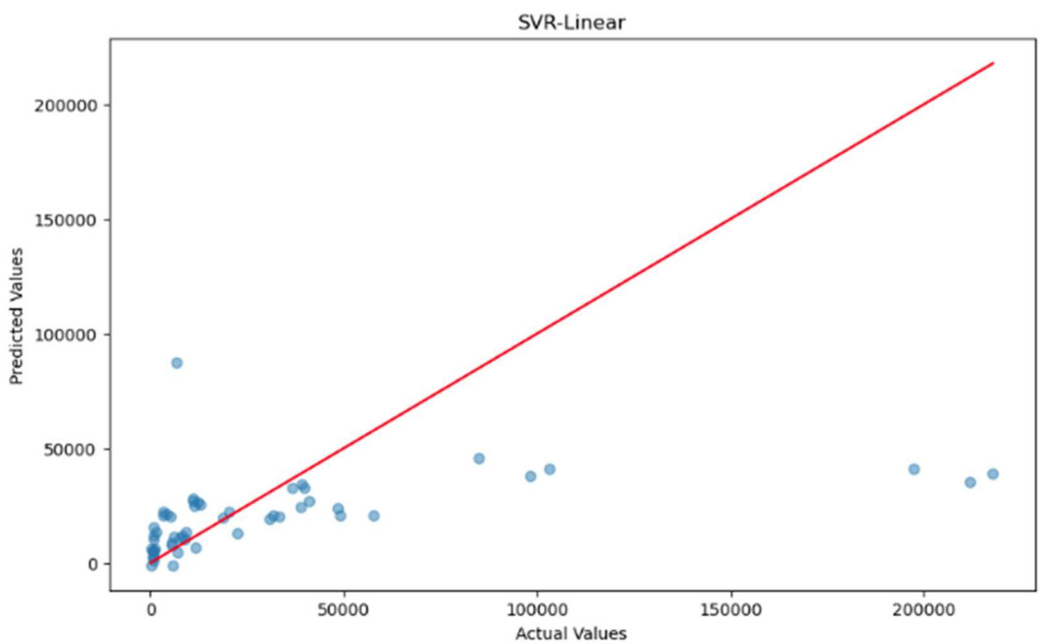

Figure 5. Support vector regression with linear kernel ($C = 199.99$).

To assess the forecasting accuracy of our trained models more formally, in Table 2 we present their respective forecasting metrics in terms of the MAPE, the RMSE, and the R^2 both in the training set and the out-of-sample data. According to the training set forecasting metrics in 2-7, the best

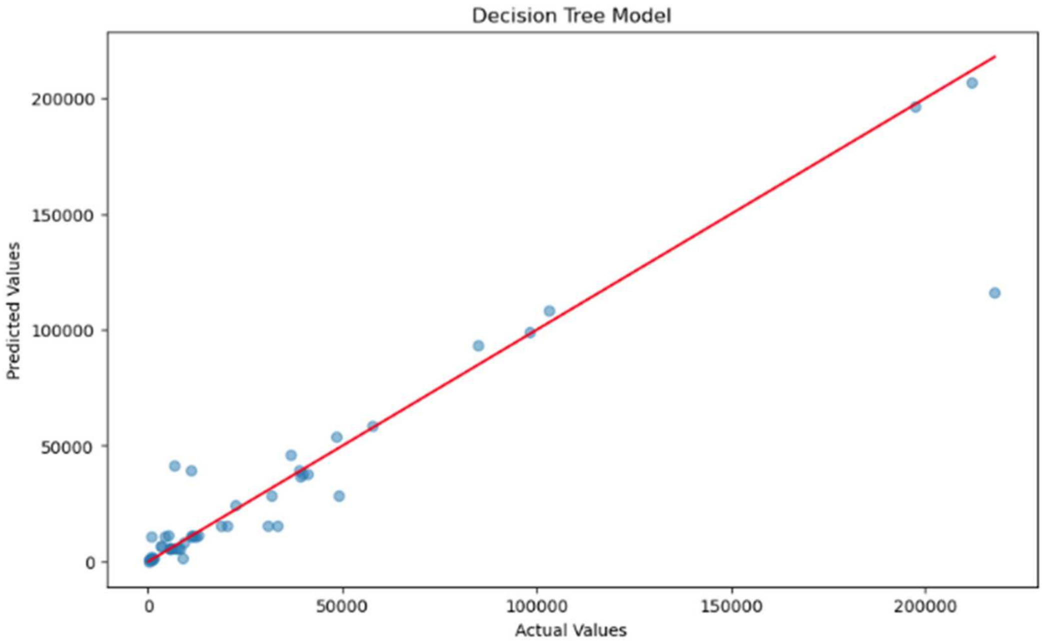

Figure 6. Decision tree-predicted model. (criterion: absolute error, max depth: 5, max leaf nodes: 22, min samples split: 4).

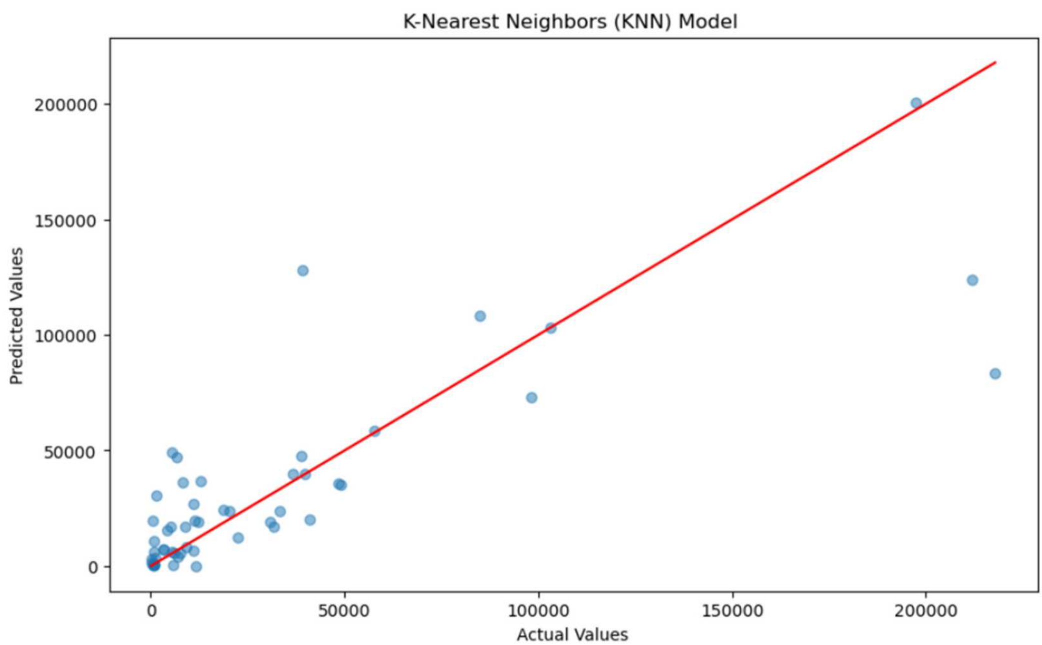

Figure 7. K-Neighbours predicted regression model (n_neighbours: 2).

model is the gradient-boosting tree (GBT), which significantly outperforms the other models in all three metrics, both in the training set and the out-of-sample test set. The GBT exhibits a MAPE of 0.10%, an RMSE of 244.01 and an R^2 of 1.000 (rounded) in the training set. The same metrics in the out-of-sample test are 1.36%, 15,463.20 and 0.90. the Decision Trees model ranks second in

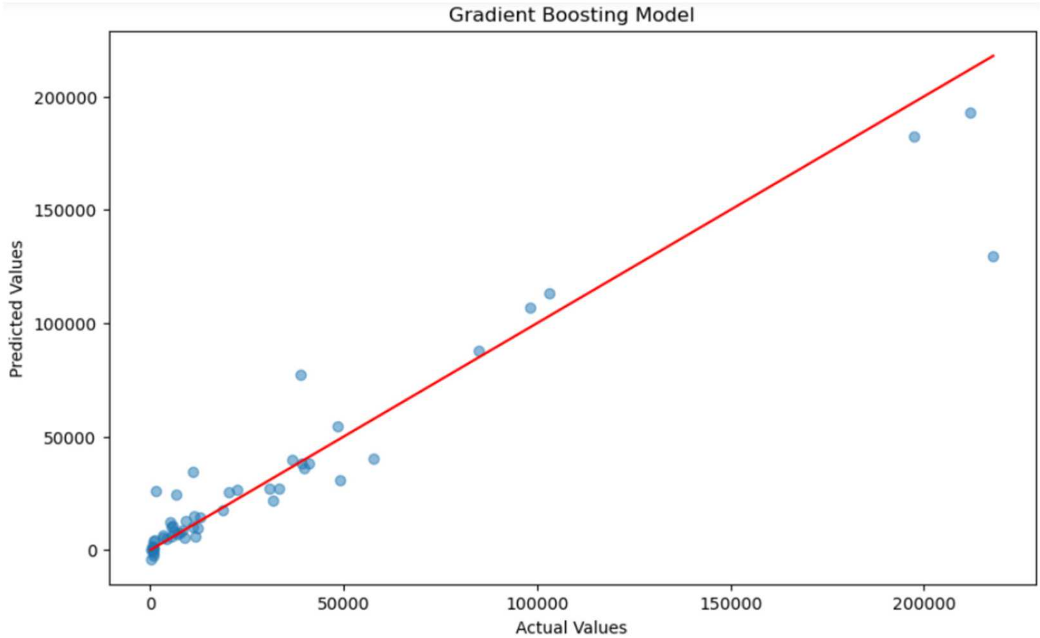

Figure 8. Gradient boosting-predicted model (learning rate: 0.1, max depth: 4, n_estimators: 100).

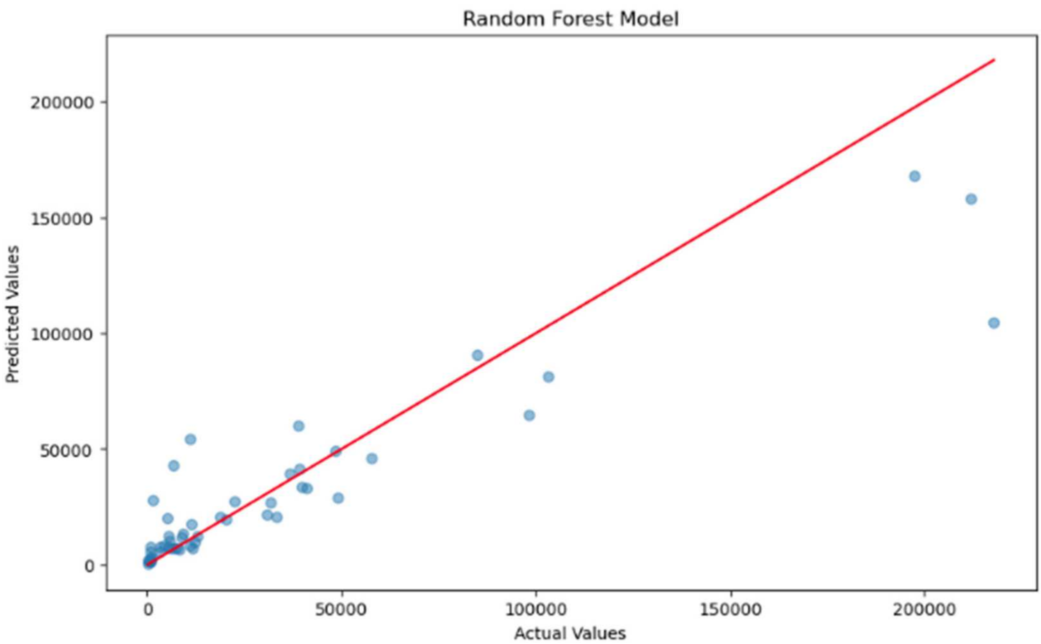

Figure 9. Random forest-predicted model (max depth = 12, min samples leaf = 5, random state = 50).

forecasting performance, as it outperforms the KNN, Random Forest, SVR-linear and SVR-RBF, in terms of the MAPE and the RMSE in the training set and in all three metrics in the out-of-sample set.

Next, we extract the Variable Importance Measure (VIM) that is presented in Table 3. The VIM ranks the independent variables according to their importance in forecasting the dependent

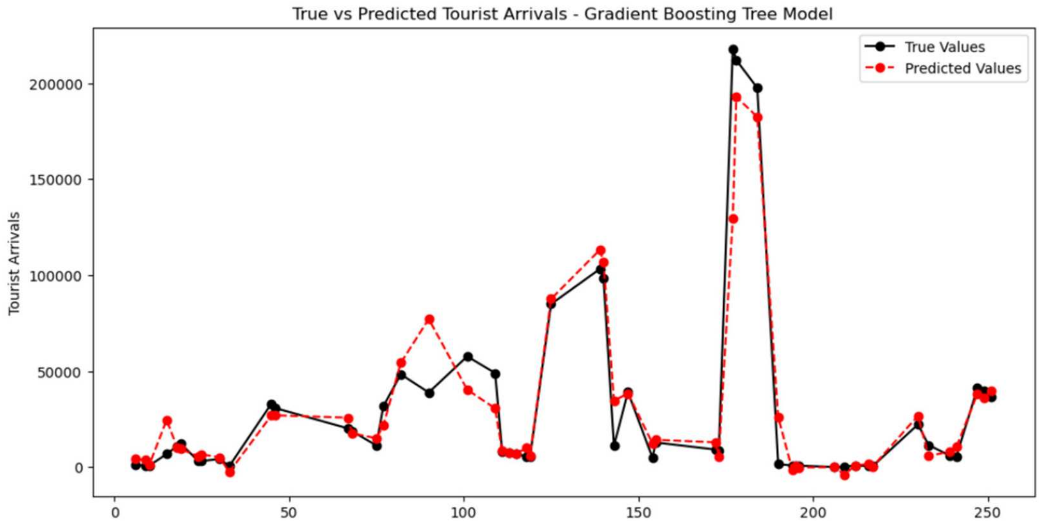

Figure 10. Comparison of true and predicted tourist arrivals – optimal gradient boosting tree model.

Table 2. Forecasting performance metrics of the machine learning models for tourism demand in the training set and the out-of-sample data.

Model	Training set MAPE	RMSE	R^2	Out-of-sample set MAPE	RMSE	R^2
SVR-RBF	8.03%	52,781.31	0.11	7.99%	52,789.87	0.11
SVR-Linear	4.21%	37,023.33	0.23	3.80%	45,602.25	0.17
Decision Tree	0.21%	22,401.45	0.80	0.73%	16,191.47	0.89
KNN	1.54%	19,885.33	0.78	2.99%	28,656.98	0.67
GBT	**0.10%**	**244.01**	**1.00**	**1.36%**	**15,463.20**	**0.90**
Random Forest	1.03%	12,355.01	0.91	1.52%	21,294.44	0.82

Note: The table presents the Mean Absolute Percentage Error (MAPE), the Root Mean Square Error (RMSE) and the R^2 for various machine learning models used in predicting tourism demand. Lower values of MAPE and RMSE indicate superior model performance while higher values of the R^2 indicate a better fit model. Best models are shown in bold.

variable. It is evident that certain features distinctly influence tourist demand. The Political Globalisation Index (PGI), with an importance of 57.62%, emerges as the most important predictor. This suggests that tourism demand is intricately linked with how countries integrate themselves on the global stage. Furthermore, delving into the intertwined nature of global politics, economics, and tourism, it becomes clear that international mobility and choices of potential tourists are not solely influenced by a country's attractions, but also by its political and economic orientation in the global market. A strong PGI indicates that a country is an active member of the global community, which can create a perception of openness and acceptance of foreign visitors that is of course attractive to foreign visitors. This finding aligns with the study of Bayar et al. (2021), where they found that the PGI positively affects inbound tourism, further highlighting the importance of political globalisation in shaping tourism trends.

Government spending as a percentage of GDP ranks as the second most important forecaster of tourist arrivals with a VIM of 15.22%. This is somewhat surprising at first; why government spending plays such an important role in tourist arrivals? One way to interpret this result, is that there exists a potential correlation between government spending in terms of investment on infrastructure, spending on promotional campaigns, or security, and a surge in tourism demand as Cannonier and Burke (2019), also find. Quality infrastructure is appealing, since it can translate into efficient transportation services and reliable public utility facilities during the tourists' stay. These two

Table 3. Variable importance results for forecasting tourism demand.

Variables	Variable importance
Political Globalisation Index (PGI)	0.576207
Government spending as percent of GDP	0.152214
Unemployment rate	0.087058
Government dept, % of GDP	0.077282
Political Stability	0.035592
Heating Degree Days	0.031632
Economic globalisation Index	0.014229
Fossil_CO2 per GDP	0.009037
Economic growth %	0.008577
Fiscal balance, percent of GDP	0.005102
EUR/SEK in SEK	0.003136
Global CBOE Volatility Index (^VIX)	0.001173
World Pandemic Uncertainty Index (WPU)	0.000803
GBP/EUR in Euro	0.000697
Inflation	0.000498
EUR/CHF in CHF	0.000363

Note: The table presents the importance scores of selected factors derived from the variable selection process. Variables are more likely to predicting the target variable when they receive higher significance scores. Some variables have a zero-significance score, which indicates that they have no impact on the predictive model when using the current methodology.

variables, the Political Globalisation Index and Government Spending as a percentage of GDP, account for 15.22% of all information used to forecast tourist arrivals and the rest 14 variables account for the remaining 42.33%.

Another noteworthy observation is that economic indicators like the Unemployment rate and Government debt to GDP ratio were also significant, though their influence was of lower magnitude at 8.70% and 7.72%, respectively. The unemployment rate was in third place with a contribution of 8.70%. This result suggests that countries with high unemployment have potential challenges, possibly resulting in decreased tourist satisfaction (Sánchez López, 2019). The government's debt to GDP ratio is the fourth most important variable provided by the VIM 7.72%; it provides evidence of each country's fiscal health and financing patterns. A high Government debt may impact tourist arrivals in two ways: (a) directly, as tourists might perceive a country's excessive debt burden as evidence of economic instability, which could raise concerns about its stability, and (b) indirectly, as a significant debt burden may limit government spending on initiatives that promote tourism, and result in the implementation of fiscal policies that discourage travellers as they may negatively impact safety, cleanliness, security, transportation and other important infrastructure.

Conversely, the importance of Political stability and Heating Degrees Days, are ranked fifth and sixth with a VIM of 3.56% and 3.16%, respectively. These variables may indicate that tourists are prioritising safety, a factor heavily influenced by a region's political stability and may be very sensitive to climate factors. This aligns with the study by Falk and Lin (2018), which found a four per cent increase in arrivals due to a decrease of one degree Celsius. In our study, this importance could be related to travellers' personal choices or behaviours influenced by the need for colder or warmer climates, making this as an important factor in understanding and predicting tourist demand.

Moreover, while the Economic Globalisation Index's importance of 1.42% reflects the aggregated economic openness of the 24 EU countries, its relatively lower importance suggests that its influence on attracting tourist demand through economic and trade dynamics is very limited. Finally, the VIM demonstrates that the rest of the variables have a negligible influence on tourist arrivals, with percentages ranging from 0.9% to 0%. These variables, economic growth, Fiscal Balance, Fossil OC2, exchange rates, Global CBOE Volatility Index, Inflation and the WPU Index, have a negligible impact on the aggregated predictive model for tourist demand.

While several studies have indicated a relationship between exchange rates and tourist demand (Adeleye et al., 2022; Tung, 2019), our findings reveal a weak such relationship. In our case, where we study tourism demand for the 24 European countries, this might be the result of strong tourism demand that originates within the Euro area where exchange rates are irrelevant.

6. Conclusions and policy implications

In this study, we collected and combined a comprehensive dataset, including tourist arrivals, various relevant macroeconomic indicators, political stability indices, environmental and energy variables, foreign exchange prices, globalisation metrics, and financial market indicators for 24 EU countries spanning the period from 2010 to 2020. The dataset included a total of 264 observations, divided into two subsamples: the in-sample set used to train our machine learning models, and the out-of-sample set used to validate and measure the accuracy of the optimal models in new and unknown (to the trained models) data. Six state-of-the-art machine learning algorithms, the Support Vector Regression with both the Linear and the RBF kernel, the Random Forest, the Decision Trees, the gradient-boosting trees, and the K-Nearest Neighbours were employed to train the models using the in-sample dataset. For each Machine Learning algorithm, the optimal values of the respective hyperparameters were initially obtained using a three-fold cross-validation process and a feature scaling technique, to avoid overfitting and provide robust forecasts.

Our research highlights the superior performance of the gradient-boosting tree models, which achieved a high training accuracy with a MAPE of 0.1%, outperforming the rest of the machine learning models (the same model achieved 1.36% MAPE on the out-of-sample).

Accurately predicting tourism demand is crucial for policymakers and stakeholders worldwide. As a result, a multitude of studies employs different models and techniques (Peng et al., 2014; Periera & Cerqueira, 2022; Sulong et al., 2023). While the existing literature focuses on traditional forecasting models and takes into account factors such as economic policy uncertainty, geopolitical risks, and global events (Ongan & Gozgor, 2018; Saha & Yap, 2014; Sharma & Khanna, 2023; Wujie, 2023), there is a need for more accurate forecasting techniques (Chen et al., 2015; Sulong et al., 2023). This study offers significant insights into forecasting tourism demand amidst global uncertainties and provides a topical contribution to both the existing literature and also to relevant policymakers and stakeholders in the tourism industry.

In this paper, we, first forecast tourism demand in 24 European countries using a single universal model, enabling policymakers to make informed decisions, optimising resource distribution, and streamlining cross-border activities, fostering better coordination and cooperation within the EU. Second, by employing six alternative state-of-the-art machine learning (ML) algorithms, we provide stakeholders with important information about the efficiency of various forecasting techniques combining a grid-search cross-validation method. Third, our investigation into the impact of uncertainty on international tourism, including five indices that measure various forms of uncertainty, provides valuable insights for local governments, offering strategies for coping with global uncertainties that may reduce tourism-related income and employment. Fourth, by applying the Variable Importance Measure, we identify and rank the most important variables in predicting tourism demand. This provides insights into the key determinants that drive tourism trends. This variable importance ranking reveals the importance of the Political Globalisation Index, as it is ranked as the most significant factor in predicting tourism demand, emphasises the crucial role of political factors in shaping tourism demand for 24 EU countries, encouraging policymakers to consider political dynamics and geopolitical risks more carefully when developing tourism policies and plans.

This research has important implications for both business and government decision-making processes. Employing the proposed machine learning techniques, we have created highly accurate predictive models that can lead to more effective strategies, particularly for local governments in distributing tourism benefits to the local economies. By implementing the appropriate policies based on the forecast, governments can better manage employment and revenue influx, associated with tourism demand fluctuation. This promotes economic stability and resilience even in the face of global uncertainties that affect local tourism demand. Essentially, the ability to accurately anticipate tourism demand during uncertain times arms businesses, governments, and other stakeholders with vital insights to counteract adversities, fostering tourism's consistent and sustainable growth and resilience to exterior shocks in the demand.

Future research could attempt to refine these models by obtaining more granular data related to political globalisation and government spending. For instance, data on specific types of political treaties or particular sectors of government spending might provide sharper insights. The implications of these findings are substantial, providing a roadmap for stakeholders in the tourism sector to navigate and strategise in an ever-evolving global landscape.

Note

1. Variance Importance Factor (VIF): VIF is used to measure multicollinearity by calculating how much the variance of an estimated regression coefficient rises when predictors are correlated. If the VIF equals to 1, there is no multicollinearity among regressors. If the VIF is more than 1, then the regressors may be slightly correlated. A VIF of 5–10 suggests significant correlation. If the VIF exceeds 10, then the regressions coefficients are underestimated due to multicollinearity.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Adeleye, B. N., Ogede, J. S., Rabbani, M. R., Adam, L. S., Mazhar, M., & Trinidad Segovia, J. E. (2022). Moderation analysis of exchange rate, tourism and economic growth in Asia. *PLoS One*, *17*(12), e0279937. https://doi.org/10.1371/journal.pone.0279937

Adnan Hye, M. Q., & Ali Khan, R. E. (2013). Tourism-Led growth hypothesis: A case study of Pakistan. *Asia Pacific Journal of Tourism Research*, *18*(4), 303–313. https://doi.org/10.1080/10941665.2012.658412

Ahir, H., Bloom, N., & Furceri, D. (2022). *The world uncertainty index w29763*. National Bureau of Economic Research, p. w29763. https://doi.org/10.3386/w29763

Bayar, Y., Odabas, H., & Sakar, E. (2021). Impact of globalization on inbound tourism in Mediterranean countries: A panel regression analysis. In Y. Bayar (Ed.), *Handbook of research on institutional, economic, and social impacts of globalization and liberalization* (pp. 87–95). IGI Global.

Blake, A., & Sinclair, M. T. (2003). Tourism crisis management. *Annals of Tourism Research*, *30*(4), 813–832. https://doi.org/10.1016/S0160-7383(03)00056-2

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). *A training algorithm for optimal margin classifiers, in Proceedings of the fifth annual workshop on Computational learning theory*. COLT92: 5th Annual Workshop on Computational Learning Theory, Pittsburgh Pennsylvania USA: ACM (pp. 144–152). https://doi.org/10.1145/130385.130401

Boukas, N., & Ziakas, V. (2013). Impacts of the global economic crisis on Cyprus tourism and policy responses: Impacts of the global economic crisis. *International Journal of Tourism Research*, *15*(4), 329–345. https://doi.org/10.1002/jtr.1878

Breiman, L. (1997). *Arcing the edge*. Technical report 486. Statistics Department, University of California at Berkeley.

Burger, C. J. S. C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting — a case study of Durban, South Africa. *Tourism Management*, *22*(4), 403–409. https://doi.org/10.1016/S0261-5177(00)00068-6

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.

Cannonier, C., & Burke, M. G. (2019). The economic growth impact of tourism in small island developing states—evidence from the Caribbean. *Tourism Economics*, *25*(1), 85–108. https://doi.org/10.1177/1354816618792792

Chase, L. C., Phillips, R. G., & Amsden, B. (2023). Stakeholder engagement in tourism planning and development. In M. Uysal & M. J. Sirgy (Eds.), *Handbook of tourism and quality-of-life research II. International handbooks of quality-of-life* (pp. 475–490). Springer.

Chen, R., Liang, C. Y., & Gu, D. X. (2015). Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. *Applied Soft Computing*, *26*, 435–443. https://doi.org/10.1016/j.asoc.2014.10.022

Chen, D., Sun, F., & Liao, Z. (2024). Forecasting tourism demand of tourist attractions during the COVID-19 pandemic. *Current Issues in Tourism*, 27(3), 445–463. https://doi.org/10.1080/13683500.2023.2165482

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Darren, G., & Mallery, P. (2010). *SPSS for windows step by step: A simple guide and reference* (16th ed.). Routledge.

Falk, M., & Lin, X. (2018). Sensitivity of winter tourism to temperature increases over the last decades. *Economic Modelling*, *71*, 174–183. https://doi.org/10.1016/j.econmod.2017.12.011

Fix, E., & Hodges, J. L. (1955). Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Revue Internationale de Statistique* [*International Statistical Review*], *57*(3), 238. https://doi.org/10.2307/1403797.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

Guo, W. (2007). Inbound tourism, an empirical research based on gravity model of international trade'. *Tourism Tribune*, *22*(3), 30–34.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson Educational International.

Hall, C. M. (2010). Crisis events in tourism: Subjects of crisis in tourism. *Current Issues in Tourism*, *13*(5), 401–417. https://doi.org/10.1080/13683500.2010.491900

Hu, M., Li, H., Song, H., Li, X., & Law, R. (2022). Tourism demand forecasting using tourist-generated online review data. *Tourism Management*, *90*, 104490. https://doi.org/10.1016/j.tourman.2022.104490

Jiao, X., Li, G., & Chen, J. L. (2020). Forecasting international tourism demand: A local spatiotemporal model. *Annals of Tourism Research*, *83*, 102937. https://doi.org/10.1016/j.annals.2020.102937

Khadaroo, J., & Seetanah, B. (2008). The role of transport infrastructure in international tourism development: A gravity model approach. *Tourism Management*, *29*(5), 831–840. https://doi.org/10.1016/j.tourman.2007.09.005

Kreishan, F. M. M. (2010). Tourism and economic growth: The case of Jordan. *European Journal of Social Sciences*, *15*(2), 229–234.

Kulendran, N., & King, M. L. (1997). Forecasting international quarterly tourist flows using error-correction and time-series models. *International Journal of Forecasting*, *13*(3), 319–327. https://doi.org/10.1016/S0169-2070(97)00020-4

Kumar, S. P. K., & Sanjeev, M. A. (2020). Terrorism–tourism–economic growth nexus in India: An NARDL evidence. *FIIB Business Review*, *9*(4), 300–308. https://doi.org/10.1177/2319714520965281

Lang, L., Tiancai, L., Shan, A., & Xiangyan, T. (2021). An improved random forest algorithm and its application to wind pressure prediction. *International Journal of Intelligent Systems*, *36*(8), 4016–4032. https://doi.org/10.1002/int.22448

Li, X., Zhang, X., Zhang, C., & Wang, S. (2024). Forecasting tourism demand with a novel robust decomposition and ensemble framework. *Expert Systems with Applications*, *236*. https://doi.org/10.1016/j.eswa.2023.121388

Lim, C., & McAleer, M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management*, *23*(4), 389–396. https://doi.org/10.1016/S0261-5177(01)00098-X

Martins, L. F., Gan, Y., & Ferreira-Lopes, A. (2017). An empirical analysis of the influence of macroeconomic determinants on world tourism demand. *Tourism Management*, *61*, 248–260. https://doi.org/10.1016/j.tourman.2017.01.008

Mehta, P., Bukov, M., Wang, C.-H., Day, A. G., Richardson, C., Fisher, C. K., & Schwab, D. J. (2019). A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, *810*, 1–124. https://doi.org/10.1016/j.physrep.2019.03.001

Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted random forest.. *IEICE Transactions on Information and Systems*, *E98.D*(D(9)), 1630–1636. https://doi.org/10.1587/transinf.2014OPP0004

Neumayer, E. (2004). The impact of political violence on tourism – dynamic econometric estimation in a cross-national panel. *Journal of Conflict Resolution*, *48*(2), 259–281. https://doi.org/10.1177/0022002703262358

Ongan, S., & Gozgor, G. (2018). Tourism demand analysis: The impact of the economic policy uncertainty on the arrival of Japanese tourists to the USA. *International Journal of Tourism Research*, *20*(3), 308–316. https://doi.org/10.1002/jtr.2182

Pattie, D. C., & Snyder, J. (1996). Using a neural network to forecast visitor behavior. *Annals of Tourism Research*, *23*(1), 151–164. https://doi.org/10.1016/0160-7383(95)00052-6

Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management*, *45*, 181–193. https://doi.org/10.1016/j.tourman.2014.04.005

Peng, B., Song, H., Crouch, G. I., & Witt, S. F. (2015). A meta-analysis of international tourism demand elasticities. *Journal of Travel Research*, *54*(5), 611–633. https://doi.org/10.1177/0047287514528283

Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, *25*(17), 2733–2750. https://doi.org/10.1080/13683500.2021.1999397

Russo, D., & Zou, J. (2020). How much does your data exploration overfit? Controlling bias via information usage. *IEEE Transactions on Information Theory*, *66*(1), 302–323. https://doi.org/10.1109/TIT.2019.2945779

Saha, S., & Yap, G. (2014). The moderation effects of political instability and terrorism on tourism development: A cross-country panel analysis. *Journal of Travel Research*, *53*(4), 509–521. https://doi.org/10.1177/0047287513496472

Sánchez López, F. (2019). Unemployment and growth in the tourism sector in Mexico: Revisiting the growth-rate version of okun's Law. *Economies*, *7*(3), 83. https://doi.org/10.3390/economies7030083

Scarlett, H. G. (2021). Tourism recovery and the economic impact: A panel assessment. *Research in Globalization*, *3*, 100044. https://doi.org/10.1016/j.resglo.2021.100044

Sequeira, T. N., & Maçãs Nunes, P. (2008). Does tourism influence economic growth? A dynamic panel data approach. *Applied Economics*, *40*(18), 2431–2441. https://doi.org/10.1080/00036840600949520

Sharma, C., & Khanna, R. (2023). Does global economic policy uncertainty drive tourism demand? A cross-country analysis. *Journal of Policy Research in Tourism, Leisure and Events*, *15*(1), 106–113. https://doi.org/10.1080/19407963.2021.1916510

Skerritt, D., & Huybers, T. (2005). The effect of international tourism on economic development: An empirical analysis. *Asia Pacific Journal of Tourism Research*, *10*(1), 23–43. https://doi.org/10.1080/1094166042000330209

Sofianos, E., Zaganidis, E., Papadimitriou, T., & Gogas, P. (2024). Forecasting east and west coast gasoline prices with tree-based machine learning algorithms. *Energies*, *17*(6), 1296. https://doi.org/10.3390/en17061296

Sönmez, S. F. (1998). Tourism, terrorism, and political instability. *Annals of Tourism Research*, *25*(2), 416–456. https://doi.org/10.1016/S0160-7383(97)00093-5

Suhel, S., & Bashir, A. (2018). The role of tourism toward economic growth in the local economy. *Economic Journal of Emerging Markets*, *10*(1), 32–39. https://doi.org/10.20885/ejem.vol10.iss1.art4

Sulong, Z., Abdullah, M., & Chowdhury, M. A. F. (2023). Halal tourism demand and firm performance forecasting: New evidence from machine learning. *Current Issues in Tourism*, *26*(23), 3765–3781. https://doi.org/10.1080/13683500.2022.2145458

Tajmouati, S., Wahbi, B. E., Bedoui, A., Abarda, A., & Dakkoun, M. (2024, February 25). Applying k-nearest neighbors to time series forecasting: Two new approaches. *Journal of Forecasting*, https://doi.org/10.1002/for.3093

Thompson, C. G., Kim, R. S., Aloe, A. M., & Becker, B. J. (2017). Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*, *39*(2), 81–90. https://doi.org/10.1080/01973533.2016.1277529

Tiwari, A. K., Das, D., & Dutta, A. (2019). Geopolitical risk, economic policy uncertainty and tourist arrivals: Evidence from a developing country. *Tourism Management*, *75*, 323–327. https://doi.org/10.1016/j.tourman.2019.06.002

Tung, L. T. (2019). Does exchange rate affect the foreign tourist arrivals? Evidence in an emerging tourist market. *Management Science Letters*, *9*(8), 1141–1152. https://doi.org/10.5267/j.msl.2019.5.001.

Valentinas, N., Petrokė, I., Urbański, M., & Soboleva, O. (2022). Macroeconomic factors influencing the development of the sharing economy in the Lithuanian tourism sector. *Journal of Tourism and Services*, *13*(25), 248–266. https://doi.org/10.29036/jots.v13i25.437

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer.

Wang, T., Hu, S., & Jiang, Y. (2021). Predicting shared-car use and examining nonlinear effects using gradient boosting regression trees. *International Journal of Sustainable Transportation*, *15*(12), 893–907. https://doi.org/10.1080/15568318.2020.1827316

Wu, D. C. W., Ji, L., He, K., & Tso, K. F. G. (2021). Forecasting tourist daily arrivals With A hybrid SARIMA–LSTM approach. *Journal of Hospitality & Tourism Research*, *45*(1), 52–67. https://doi.org/10.1177/1096348020934046

Wujie, X. (2023). The impact of geopolitical risks and international relations on inbound tourism - evidence from China and key source countries. *Cogent Social Sciences*, *9*(2), 2285244. https://doi.org/10.1080/23311886.2023.2285244

Xu, T., & Lv, Z. (2023). The effect of tourism on the shadow economy: The level of economic development is key. *Tourism Economics*, *29*(8), 2129–2142. https://doi.org/10.1177/13548166221126577