Journal of Digital Economy 1 (2022) 32–43

Contents lists available at ScienceDirect

Journal of Digital Economy

journal homepage: www.keaipublishing.com/en/journals/journal-of-digital-economy

Using social media big data for tourist demand forecasting: A new machine learning analytical approach

Yulei Li, Zhibin Lin *, Sarah Xiao

Durham University Business School, Millhill Lane, Durham, DH1 3LB, UK

ARTICLE INFO

Keywords:
Tourism demand forecasting
Tourist arrival
Tourist attitude
Social listening
Machine learning
Natural language processing

ABSTRACT

This study explores the possibility of using a machine learning approach to analysing social media big data for tourism demand forecasting. We demonstrate how to extract the main topics discussed on Twitter and calculate the mean sentiment score for each topic as the proxy of the general attitudes towards those topics, which are then used for predicting tourist arrivals. We choose Sydney, Australia as the case for testing the performance and validity of our proposed forecasting framework. The study reveals key topics discussed in social media that can be used to predict tourist arrivals in Sydney. The study has both theoretical implications for tourist behavioural research and practical implications for destination marketing.

1. Introduction

Accurate forecasting of tourist arrivals is important yet a well-recognised challenge for both scholars and destination managers (Zhang et al., 2021), because numerous factors could potentially influence tourist arrivals, and there is a time lag between events and their effects on arrivals (Law et al., 2019; Li et al., 2020). Inaccurate forecasting could result in an under- or over-supply of food, hotel rooms, and infrastructures in the destination, which could have a significant impact on the sustainability of the destination's tourism industry. Accordingly, identifying those potential factors, collecting relevant data, and adopting appropriate analytical methods are essential for demand forecasting.

Various methods of tourism demand forecasting have been adopted in the literature, for example, the non-causal time-series models (Lim and McAleer, 2001; Wu et al., 2017), causal econometric models (Fourie and Santana-Gallego, 2011; Song and Witt, 2006), and the recent advances in machine learning and artificial intelligence offer new approaches to forecasting tourism demand with greater accuracy (Bi et al., 2021; Hu et al., 2021; Law et al., 2019; Sun et al., 2019; Zhang et al., 2021; Zheng et al., 2021).

The proliferation of the internet and social media in recent years provides a new source of data, which is large, valuable, and publicly available for market research (Ghimire et al., 2022). Recently Park et al. (2021) use news topics in online newspapers to predict tourism demand. Sun et al. (2019) have explored the use of an online search index for predicting tourist arrivals. Gunter et al. (2019) show that Facebook “likes” can be used for the prediction of tourist arrivals along with Google Trends. Moreover, researchers have increasingly used social media data to analyse tourist satisfaction (Guo et al., 2017) or to establish tourists' use of social media and tourist well-being (Chen et al., 2021). Recently, Ghimire et al. (2022) explore the link between online reviews and the size of the customer base. However, whether the content discussed in social media can be used for predicting tourist arrivals remain unknown.

In this study, we examine the consumer perceptions and attitudes towards the destination as expressed on social media to predict

* Corresponding author.
E-mail addresses: yulei.li@durham.ac.uk (Y. Li), zhibin.lin@durham.ac.uk (Z. Lin), hong.xiao@durham.ac.uk (S. Xiao).

https://doi.org/10.1016/j.jdec.2022.08.006
Received 18 January 2022; Received in revised form 13 July 2022; Accepted 16 August 2022

tourist arrivals. To our best knowledge, our research is one of the first to examine tourist perceptions of destination attributes by using social media user-generated content to predict tourist arrivals. We show the possibility to provide timely prediction through listening to social media's consumer voice, i.e. social listening. For this purpose, we develop a methodological framework that allows researchers to extract and analyse consumer attitudes towards the destination for demand forecasting. The study provides both theoretical implications for destination marketing research and practical implications for destination marketing organizations by adding social listening a new feature for demand forecasting.

2. Literature review

2.1. Factors influencing tourism demand

2.1.1. Economic factors

Economic models of tourism demand forecasting are based on conventional economic theory, such as the demand-supply model, utility theory and consumption behavioural theory (Goh, 2012). Those theories suggest that economic factors, such as price and income, can influence the decision making of potential tourists, which further influences a destination's tourism demand. The five most commonly examined economic factors are: income (Lanouar and Goaied, 2019; Martins et al., 2017; Shafiullah et al., 2019; Tavares and Leitão, 2017); population (Cho, 2010; Lanouar and Goaied, 2019; Pham et al., 2017); relative price (Dogru et al., 2017; Yang et al., 2019); exchange rate (Demiralay, 2020; Martins et al., 2017; Tavares and Leitão, 2017) and travel costs (De Vita and Kyaw, 2013; Demiralay, 2020).

However, the conventional economic theory receives numerous critiques when it is used to explain demands. The theory assumes that all tourists always make rational decisions and that their tastes are static rather than dynamic. These two assumptions impede the interpretability of the generalisability of pure economic theory in the real world. Lipsey et al. (1987) argue that, for typical products and services, price and income can only explain approximately 70% of the variation in demand. The tourism ecosystem is much more complex than typical products. Non-economic factors can also influence tourists' destination choices and further affect the tourism demand of the destination. For example, Tavares and Leitão (2017) find that tourists tend to avoid destinations with similar languages and choose to visit those countries which speak distinct languages.

The static assumption of traditional economic demand theory requires itself to reflect tourists' preferences on a utility map and this impedes economic theory to reveal the intrinsic properties of a particular product or service (Goh, 2012). Lancaster (1966) posits that consumers possess goods' characteristics, and it is these characteristics, rather than the goods themselves that give consumers utility. He continues to point out that traditional economic models may fail to predict the demand for a product when the situation changes without knowing how the properties affect consumers' preferences at the beginning (Lancaster, 1966). Compared with typical products, tourists do not possess or consume travel destinations to derive utility (Rugg, 1973). Rather, they derive utility from being in a particular destination for a certain period of time. Various incidents can occur during this period, which makes it challenging for the conventional demand models to accurately forecast tourism demand.

2.1.2. Non-economic factors

Apart from economic determinants, there are numerous other potential determinants of tourism demand. Uysal (1998) splits non-economic determinants into two categories: exogenous and social-psychological factors. The exogenous determinants include general economic growth; political stability; technological advancement; epidemics; terrorism, among others. One of the most studied exogenous factors is the occurrence of terrorism attacks (Lanouar and Goaied, 2019; Liu and Pratt, 2017; Neumayer and Plümper, 2016), and the second most studied one is the impact of economic policy (Balli et al., 2018; Demir and Gözgör, 2018; Tsui et al., 2018).

Social-psychological determinants, such as travel preferences, perceptions, and attitudes about the destinations, cultural similarities, and tourists' demographic factors, play a vital role in the understanding of tourism demand (Dogru et al., 2017; Uysal, 1998). Um and Crompton (1990) outline the factors that can influence tourists' perception of a destination as internal and external inputs. The internal inputs are psychological factors such as an individual's values and motives. External inputs refer to the stimuli from social and marketing environments, including *significative stimuli* (the attributes of a destination); *symbolic stimuli* (marketing and promotional messages perceived); and *social stimuli* (the influence from the interaction with other people, such as electronic word of mouth or eWOM). The external stimuli help people be aware of the existence of a potential destination. People consider their preferences and situational constraints to decide on the list of potential destinations, and the external inputs are then re-assessed for making the ultimate destination decision (Um and Crompton, 1990).

The *significative stimuli* are the most commonly examined, among which a destination's environment is seen as a powerful predictor for tourism demand. Previous research provides empirical evidence on the impact of climate in destination demand (Cho, 2010; Goh, 2012; Law et al., 2019; Li et al., 2017, 2018) and the negative effect of pollution on destination choice (Volgger et al., 2019). In addition, the spatial relationship between origin and destination also plays an essential role for travellers in deciding destination. Cho (2010) and Tavares and Leitão (2017) argue that the geographic distance between origin and destination has a negative impact on tourism demand. Finally, heritage sites are a key reason for visiting and have been largely examined as a pull factor of tourism demand (Gao and Su, 2019; Ribaudo and Figini, 2017; Richards, 2018; Yang et al., 2019).

The empirical studies on the impacts of *symbolic stimuli* focus mainly on exploring how marketing and promotion affect destination tourism demand (Balli et al., 2015; Kulendran and Divisekera, 2007; Tsui and Balli, 2017). For example, Kulendran and Divisekera (2007) report a significant link between marketing expenditure and tourism demand. Zhang et al. (2010) show how marketing effectiveness is associated with inbound tourism flows. Tsui and Balli (2017) further confirm that marketing expenditure can positively

influence the inflows of airports' passengers.

Social stimuli arise from both online and offline interaction with other people, which can influence potential travellers' decision choices. Social stimuli include recommendations from family and friends, online reviews, and electronic word of mouth on social media. Especially with the prevalence of social media, such as Facebook and Twitter; and online review platform, such as TripAdvisor, electronic word of mouth play an increasingly important role in tourists' travel decision (Filieri et al., 2021; Wen et al., 2021; Weng et al., 2021; Zhang et al., 2020; Zhang et al., 2021).

2.1.3. Missing pieces in the jigsaw puzzle

Our review of the literature identifies three aspects that the current tourism demand literature has certain shortcomings: a) an unlimited number of determinants; b) dynamic nature of tourist preferences; and c) tourist attitude.

An unlimited number of influential factors. The tourism ecosystem is a highly complex system that involves nearly unlimited factors. Thus, it is virtually impossible for any researcher to exhaust all possible determinants and include them in one single model. In fact, it is not necessary to consider all determinants either, because of two reasons. First, not all determinants significantly influence tourism demand. Second, from an econometric perspective, those determinators are most likely to correlate with each other, and including those determinators into one single model may cause the multicollinearity issue. As a result, selecting a few key factors from the unlimited determinants is vital. Most scholars extract determinants based on a review of previous studies, personal experience, or data availability, however, this effort may not be as efficient as expected, given the nature of the tourism ecosystem.

Dynamic nature of tourist preferences. People's tastes and preferences have a dynamic nature. First, the determinants of destination selection may differ across destinations. Second, the factors driving tourists to a certain destination may change over time. Third, the importance of a certain determinant may be different across time. Using one single model of determinants to forecast tourism demand of across destinations and time is not realistic.

Tourist attitudes. The last missing piece in current tourism literature is the psychological links between external stimuli and internal inputs. According to Um and Crompton (1990), before external stimuli can influence tourists' destination decisions, cognitive constructs can integrate those external stimuli, such as destination attributes, with internal inputs, such as value and preferences, to generate attitudes. It is those attitudes, not stimuli or inputs themselves, influence whether or not they go to a certain destination. However, the majority of tourism demand studies merely examine the direct relationship between inputs and tourism demand. Without considering tourists' attitudes towards various determinants, the literature may fail to provide insights into what factors matter the most to the tourists when selecting a destination and how those aggregated attitudes influence tourism demand.

The omission of the intrinsic links in tourism forecasting can be attributed to the lack of available data and the challenge of measurement. Most researchers collect attitudinal data by using qualitative methods, such as interviews. For example, Kock et al. (2016) conduct 50 semi-structured interviews to collect destination-specific and salient incorporation that tourists link to Germany and Spain. Moreover, predicting tourism arrivals requires aggregated data and thus traditionally it is not common nor practical to collect qualitative data on such a large scale.

2.2. Approaches to tourism demand forecasting

Scholars have studied tourism demand forecasting since the 1960s (Gerakis, 1965; Gray, 1966), by adopting various methods. According to Song et al. (2019), the three most common approaches are time series; causal econometrics; and machine learning or artificial intelligence.

2.2.1. Time series approach

Time series models utilise previous data in the series, such as historical tourist arrivals data, to predict future trends (Peng et al., 2014). Various models have been adopted, including Naïve I (Gunter and Önder, 2016; Long et al., 2019; Martin and Witt, 1989), Naïve II (Chu, 1998; Long et al., 2019; Wu et al., 2017), ARIMA (Hassani et al., 2017; Hu and Song, 2019; Xie et al., 2020), Exponential Smoothing methods (Cho, 2003; Goh and Law, 2002; Lim and McAleer, 2001), Holt-Winter (Chu, 1998; Lim and McAleer, 2001). The advantage of these approaches is obvious: they can forecast reasonable results without using more than one data series. The time series approach, however, fails to provide insights into the influential factors or customer behaviour that decision-makers wish to understand and alter (Peng et al., 2014).

2.2.2. Causal econometric approach

Unlike time series models which attempt to predict tourist arrivals only, econometric approaches focus on examining the causal relationships between potential independent variables and tourist arrivals. The most common approaches are classical regression models (Balli et al., 2016; Darani and Asghari, 2018; Fourie and Santana-Gallego, 2011; Li et al., 2018). Moreover, as a type of dynamic econometric model, the vector autoregressive (VAR) performs well for a medium to long-term period (Gunter and Önder, 2016; Liu et al., 2018; Song and Witt, 2006). In addition to the two broad approaches, there still are many other important models, such as the Autoregressive Distributed Lag Model (ADLM) model (Hu and Song, 2019; Wan and Song, 2018) and the time-varying parameter (TVP) model (Witt et al., 2003).

2.2.3. Machine learning approach

The machine learning approach has become increasingly applied in tourism studies, thanks to the advances in algorithms and the availability of big data. It is argued that this approach can provide more accurate results than the traditional approaches (Wang, 2004).

Specifically, the artificial neural network (ANN) can deal with the nonlinear relationship between independent variables and tourist arrivals and this makes the ANN the most popular approach in tourism forecasting (Peng et al., 2014). Nevertheless, ANN cannot establish the causal relationships between those independent variables and tourism demand (Wu, 2010). In addition to ANN, scholars have also adopted other machine learning models such as the rough set approach, the support vector regression (Xu et al., 2016), the fuzzy time series method (Wang, 2004), and the grey theory (Hu et al., 2019).

3. Methods

This section describes a) how we collected data including social media data and tourism demand data; b) how we analyse the data using natural language processing (NLP) techniques; and c) how we built the optimal model for tourism demand forecasting.

3.1. Sample data collection

Sydney in Australia was chosen as the exemplar destination to test our proposed social media listening predictive modelling. Sydney has been one of the top 100 popular English-speaking tourism destinations (Brilliant Maps, 2015). We chose Twitter as the social media platform because of three main reasons. First, Twitter has been one of the most popular social media platforms for tourists to share their tourism experiences and feelings (Dillette et al., 2019). Second, destination marketing organizations (DMOs) have been using Twitter to communicate destination information (Hays et al., 2013). Third, the 140-character restriction contributes to the instantaneity of tweets (Hutchins, 2011). This characteristic is helpful in exploring the dynamic nature of tourist preference.

This study used an open-source python package called 'Twitterscraper' (Taspinar & Schuirmann, 2017). Unlike the official Twitter API which has a restriction of tweets collection for only the recent seven days, 'Twitterscraper' enable researchers to collect all tweets posted since 2006. Twitterscraper uses a command prompt to crawl desired tweets by defining keywords; the number of tweets; the start date; the end date; the language; and the output format (Taspinar & Schuirmann, 2017). The authors used 'Sydney' and 'sydney' as the search keywords to collect all relevant tweets for ten years from 2009 to 2018. Since most algorithms in NLP is for English only, we collected only tweets written in English. After we collected all the data, we randomly sampled 10,000 tweets using the built-in sample method within Pandas in Python (McKinney, 2010) for each year and 100,000 tweets in total to speed up the followed NLP modelling and release the computation pressure.

Tourist arrivals data in Sydney across ten years are collected from the report by Destination NSW, a lead Government agency for the tourism and major events sector (Destination, 2019).

3.2. Natural language processing (NLP)

3.2.1. Pre-processing

Only actual tweet texts were extracted. Because we are exploring the sentiments of the tweet, we are only interested in the tweet textual variable. We excluded other variables which are not helpful for analysing sentiments, such as likes, replies, and retweets. URLs were removed (Pak and Paroubek, 2010). We then pre-process the textual tweets by conducting tokenisation and lemmatisation for the following topic modelling. Tokenisation is a fundamental step for many natural language processing tasks and it refers to the process of splitting texts into individual words (Mullen et al., 2018; Webster and Kit, 1992). Lemmatisation aims to convert a word into its base or dictionary form (Balakrishnan and Lloyd-Yemoh, 2014).

3.2.2. Latent Dirichlet Allocation (LDA)

The latent Dirichlet Allocation algorithm was then used to extract topics from the pre-processed textual corpus. LDA is an unsupervised topic model to extract potential topics from texts (Blei et al., 2003). It is based on the Bayesian approach as below and it assumes that the words in each text are drawn from a mixture of baskets independently while each basket contains a set of words and the generative process for each tweet, D (Blei et al., 2003):

1) Choose $N \sim$ Poisson (ζ), N represents the length of documents;
2) Choose $\sim$ Dir (α), where α is the parameter of the Dirichlet prior on the per-review topic distributions; and
3) For each of N words w_n:
   (a) Choose a topic $z_n \sim$ Multinomial (); and
   (b) Choose a word w_n from, a multinomial probability conditioned on the topic z_n.

LDA is able to identify latent and new topics from a large number of unlabelled texts, such as the 100,000 tweets in our research, because it is an unsupervised algorithm (Blei et al., 2003). The only parameter we need to provide is the K which is the number of topics that we propose to detect. The coherence score will be used to decide the best K value.

3.2.3. Sentiment analysis

This section aims to analyse the sentiment for each topic. TextBlob, an open-source Python library based on NLTK and Pattern library, was adopted to analyse the sentiment of all the textual data (Loria et al., 2014). This library returns a sentiment polarity score for each tweet. If the result is positive (negative), it means the sentiment is positive (negative). The same rule applies to negative numbers. If the polarity result is 0, it means the sentiment is neutral. We used the TextBlob library to detect the sentiment for all tweets, followed by

averaging sentiment scores for each topic.

3.3. Forecasting models

3.3.1. Extreme Gradient Boosting (XGBoost)

This study used the Extreme Gradient Boosting (XGBoost) as the forecasting model. Gradient tree boosting, also known as gradient boosting machine (GBM) or gradient boosted regression tree (GBRT), is a popular technique among machine learning algorithms (Friedman, 2001). Chen and Guestrin (2016) made improvements to the traditional gradient tree boosting to achieve faster speed and fewer resources requirement. Given a dataset with n examples and m features, a tree ensemble model using K additive functions is shown as,

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(x_i),\ f_k \in F,$$

where F is the space of the regression tree (CART). Unlike decision trees, each regression tree has a continuous value on each leaf. The main goal of this algorithm is to minimise the following function,

$$\mathscr{L}(\varphi) = \sum_i l(\widehat{y}_i, y_i) + \sum_k \Omega(f_k)$$

where l is the loss function measuring the difference between the predicted $\widehat{y}_i$ and the actual y_i. The Ω is the penalisation function of the complexity of the model. One advantage of XGBoost is it can generate an importance score for each independent variable which can overcome the shortfalls of non-causal time-series models and ANN models. This can provide important insight for policymakers to identify the influencing factors for tourist arrivals.

To apply the Extreme Gradient Boosting method, we first randomly split the dataset into training (80% of total samples) and test sets (20% of total samples). Then, XGBoost was used to train the model. The dependent or response variable is the number of tourist arrivals to Sydney. The independent variables or features are the sentiment scores of social media data in the previous year. Finally, the model was used to predict the tourist arrivals in the test centre and analyse the accuracy of the model by comparing the predicted tourism arrivals and the actual tourism arrivals.

3.4. Performance evaluation

3.4.1. Baseline models

To evaluate the performance of the proposed XGBoost model, we chose the following four most commonly used models as the baseline models:

a) Support Vector Regression (SVR) with Gaussian kernel. SVR is a popular model for tourism demand forecasting (Law et al., 2019; Shen et al., 2019; Sun et al., 2019). Drucker et al. (1997) proposed the Support Vector Regression (SVR), which is a version of Support Vector Machines (SVM). SVR aims to minimise an upper bound of the generalisation error and this improves its potential to predict better for new data (Chen and Wang, 2007). This model requires defining the kernel function before the regression. For the first baseline model, we chose Gaussian or 'rbf' kernel.

b) SVR with linear kernel. This baseline model used a linear kernel in the SVR model discussed above.

c) Random Forest Regression (RFR). RFR can improve prediction accuracy without significantly increasing the amount of computation (Breiman, 2001). It is commonly used in classification and regression problems. For instance, Feng et al. (2019) adopted this model to predict the inbound tourist arrivals to China and achieved high accuracy.

d) Classical Linear Regression. The simplicity and interpretability of classical linear regression make it one of the most commonly adopted methods in tourism demand forecasting literature (Chu, 2009; Frechtling, 2012; Lim, 1997; Uysal and Crompton, 1984).

We used two popular error metrics, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to compare the performance of the proposed model and the baseline models.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y}_i|}{y_i}$$

The model with the lowest values of the two measures is the best performing model.

3.4.2. K-fold cross validation

To deal with the potential over-fitting issue and to increase the generalisability of our proposed model, we apply the 10-fold cross validation. K-fold cross validation divides the whole data into K subsets. Each subset is used as a test dataset and the rest is used as a training dataset. To apply this method, the number K is required. This study chose 10-fold because, in most previous tourism literature, 10-fold cross validation is the most popular one and this study followed the tradition (Feng et al., 2019).

4. Results

4.1. Extracted topics by LDA

As explained in the section, the K, which means the number of topics to be extracted, should be defined before applying LDA. According to Fig. 1, the highest coherence score is 0.5218 at 38 latent topics. We, therefore, chose 38 as the number of topics, which produced the following results.

After the process of LDA using the K defined above, the results were visualised as shown in Fig. 2.

The 38 circles on the left indicate the 38 topics identified by the LDA model. The sizes of those circles stand for their frequencies appearing in the tweets. On the top-right, the Lambda (λ) value can be customised and we chose 0.6 following the suggestion by Sievert and Shirley (2014). On the right-hand side are the top 30 most salient terms for each topic. The naming process for each topic is based on identifying a logical connection between the most frequent words for each topic. We then examine the relevance of the topic name to tourism. If a logical connection was found in a topic and the topic is related to tourism, the topic was retained. Otherwise, the topic was abandoned. This process yielded 23 topics that are relevant to tourism (as shown in Table 1 below).

After analysing the mean sentiment score for each topic for each year, we examine how the attitudes about certain topics are associated with the change in the tourist arrivals in Sydney. Table 2 shows the descriptive information of all sentiment changes for each topic during the ten years. The results show that all sentiment scores, except the topic 'game' are positive, meaning that people were generally satisfied when they discussed those topics. Averagely, people are happier with the sporting softball and beach than any other topics. The topic 'game' is the least average satisfied topic discussed regarding Sydney.

Fig. 3 shows the relationship between the mean sentiment scores for the 23 topics (grey lines) and the changes in actual tourist arrivals in Sydney (the orange line). Due to the sentiment scores (ranging from 0 to 1) and tourist arrivals (ranging from 25,000 to 37,000) are not at the same magnitude, the feature scaling method for both sentiment scores and tourist arrivals was conducted. Fig. 3 shows that the sentiment scores of certain topics can contribute more to the changes in Sydney's tourist arrivals.

4.2. Performance of the estimation model

Table 3 presents the values of MAE and MAPE for XGBoost model and baseline models, which show indicate that the XGBoost achieves the least errors on both MAE and MAPE measurements compared with the benchmark models. The forecasting MAE of the proposed model reaches 1869 and the MAPE reaches 6.06%. Therefore, our proposed estimation model can be considered the best model when predicting Sydney's tourist arrivals.

4.3. Identifying the influencing variables

One of the advantages of XGBoost, compared with other machine learning models is that it can summarise the variables with the most important for forecasting the demand. As shown in Fig. 4, the ten most influencing variables when forecasting the tourist arrivals for Sydney includes significative stimulus (Christmas; security; travel; music events; landscape; soccer events), exogenous factors (business; news report; security), and economic factors (exchange rate).

5. Discussion and conclusions

The main objective of this study was to examine the possibility of using social media data for tourism demand forecasting through

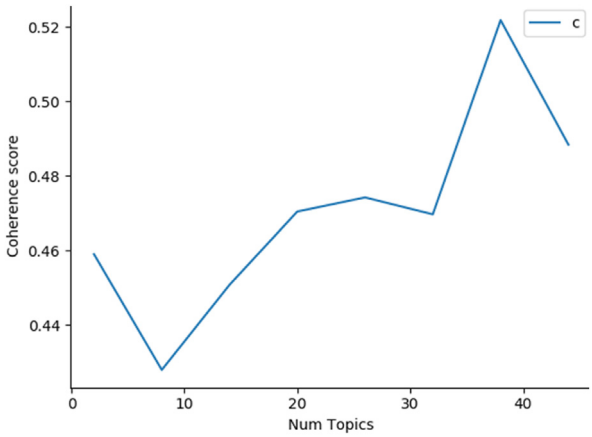

Fig. 1. Coherence score for each Ka.

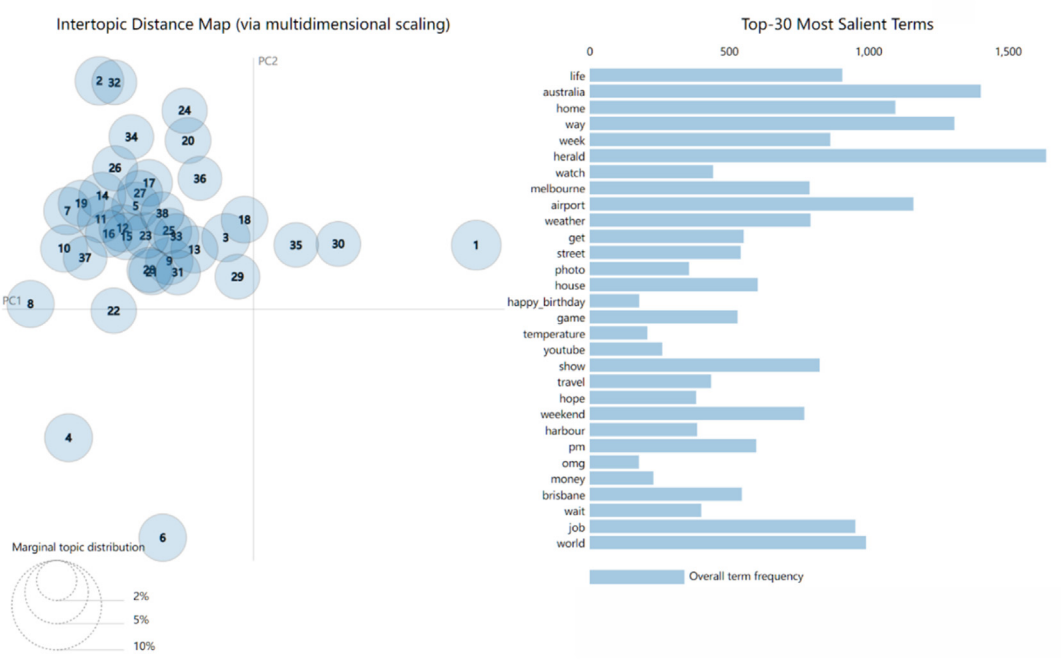

Fig. 2. Inter-topic Distance Map (via multidimensional scaling).

Table 1
Topics identified using LDA after filtering.

Topics
Weather
Landscape
Beach
History/Culture
Marketing
Travel
General news
Terrorism
Christmas
Crime
Vivid Sydney
Music
Game
Film
Exchange rate
Business environment
Employment
Rugby
Softball
Soccer
Traffic accident
General security
Drug security

natural language processing and machine learning techniques. We chose an important tourism destination Sydney, Australia to test the performance and validity of our proposed forecasting framework. The study reveals several topics discussed in social media that can be used to predict tourist arrivals in Sydney.

5.1. Theoretical implications

The study shows several topics discussed on social media that are important in tourism demand forecasting literature. For example, the weather is considered as one of the most important factors when deciding travel destination (Becken, 2013; Martín, 2005) but this factor is found insignificant in this study. One of the possible reasons is the weather condition in Sydney is relatively consistent and people tend not to show too much emotion about it.

Table 2
Descriptive statistics of sentiment scores and tourist arrivals to Sydney.

	weather	*event$_n$ews*	*travel*	*exchange$_r$ate*	*marketing*	*history*	*ent$_f$ilm*	*landscape*	*business*	*sport$_r$ugby*	*event$_t$errorism*	*sport$_o$ftball*	*event$_c$hristmas*	*event$_c$rime*	*event$_v$ividsydney*	*traffic$_a$cc*	*ent$_g$ame*	*security*	*employment*	*beach*
mean rowhead	0.12	0.12	0.09	0.1	0.11	0.12	0.09	0.12	0.16	0.11	0.1	0.13	0.11	0.11	0.09	0.11	0.05	0.1	0.09	0.13
std rowhead	0.02	0.03	0.04	0.02	0.03	0.02	0.03	0.02	0.04	0.03	0.03	0.02	0.02	0.02	0.03	0.02	0.04	0.02	0.02	0.05
min rowhead	0.1	0.08	0.0	0.07	0.04	0.09	0.06	0.07	0.09	0.07	0.06	0.1	0.09	0.08	0.06	0.08	−0.02	0.06	0.06	0.07
25% rowhead	0.11	0.09	0.07	0.09	0.09	0.1	0.07	0.11	0.13	0.08	0.08	0.12	0.1	0.09	0.07	0.11	0.04	0.08	0.07	0.1
50% rowhead	0.13	0.11	0.09	0.11	0.11	0.11	0.09	0.12	0.15	0.12	0.1	0.14	0.12	0.11	0.09	0.12	0.05	0.1	0.09	0.12
75% rowhead	0.13	0.13	0.12	0.12	0.13	0.13	0.11	0.14	0.19	0.13	0.12	0.15	0.13	0.12	0.11	0.13	0.08	0.12	0.12	0.14
max rowhead	0.16	0.16	0.14	0.13	0.15	0.15	0.13	0.15	0.21	0.17	0.16	0.16	0.14	0.15	0.13	0.14	0.11	0.13	0.12	0.23

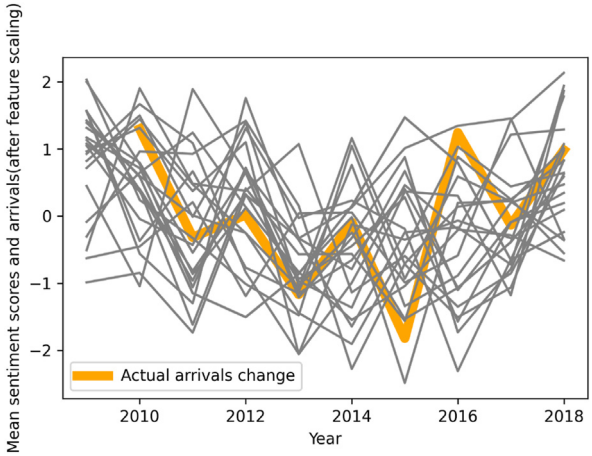

Fig. 3. Sentiment scores of 23 topics vs. tourism arrivals change.

Table 3
Performance of XGBoost and other baseline models.

Model	MAE (mean)	MAPE (mean)
XGBoost	1432	3.93%
SVR (linear)	2425	7.82%
SVR (rbf)	2439	7.87%
Random Forest Regressor	2398	7.92%
Linear Regression	2987	9.75%

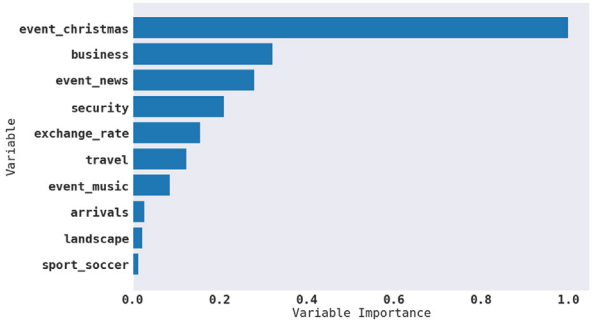

Fig. 4. Top 10 influential variables.

The study reveals topics regarding Christmas, business environment, general news, security and exchange rate are the five main factors to predict tourist arrivals. Of those five variables, the discussion on two significative stimulus, Christmas and security, play an essential role in attracting tourist arrivals to Sydney. The positive effects of significative stimulus, such as events and security, have been empirically verified by tourism scholars (Cho, 2010; Gao & Su, 2019; Goh, 2012; Ribaudo & Figini, 2017; Richards, 2018; Yang et al., 2019).

Apart from the tweets discussing Christmas events, social media users' attitudes towards the business environment are another important factor influencing tourist arrivals to Sydney. This finding is consistent with the business literature, which highlights that the business environment in both destinations and origins can influence whether tourists choose the destination. For leisure tourism, a better environment usually means better tourism facilities that may attract tourists (Yap & Allen, 2011). Moreover, a stable business environment can attract business travellers because when firms feel more confident about the business environment in the destination, they are more likely to authorise the travel request (Njegovan, 2005).

Social media users' attitudes towards general news events are another important predictor. The result aligns with the research by Stepchenkova and Eales (2011) and Imison and Schweinsberg (2013). As Költringer and Dickinger (2015) explain, online news and other information sources, such as travellers' blogs and videos, can communicate certain aspects of the destination and further influence tourists' decision making.

The last influential topic of tourist arrivals to Sydney refers to an economic factor, exchange rate. Exchange rate determines the relative price of a destination. Traditional economic theory postulates that the increase in price may lead to the decrease in demand.

Empirical studies has also confirmed the significant relationship between exchange rate and tourist arrivals (Demiralay, 2020; Martins et al., 2017; Tavares & Leitão, 2017).

5.2. Methodological implications

In this study, we present a methodological framework for tourism demand analysis. This framework utilises topic modelling, sentiment analysis, and XGBoost forecasting model to forecast tourist arrivals and to explore the potentially influential factors of tourism demand. As argued in the literature review, it is challenging to include appropriate independent variables for tourism demand modelling because of the dynamic nature of tourists' tastes and attitudes. The NLP analysis of social media big data within the methodological framework provides a novel approach to extracting the user attitudes towards those important factors. Attitudes towards destinations play an essential role in various paradigms of the marketing discipline. Extracting attitudes using the topic modelling and sentiment analysis proposed in this study provides a tool of social listening for forecasting and managing tourism demand.

5.3. Managerial implications

This study provides three primary implications for policymakers and managerial teams in the marketing and tourism sectors. First, social media is a powerful source for collecting attitudinal data of consumers and social listening enables the forecasting of consumer demand. This study provides a tool to extract the main topics discussed on Twitter and calculate the mean sentiment score for each topic as the proxy of the general attitudes towards those topics. This enables the managerial team to take advantage of the power of social media big data and gain invaluable insights into how their consumers evaluate their products and services.

Second, the application of the XGBoost model may help managers and policymakers extract the important factors that could influence their bottom line. For example in the case of our empirical study, our model suggests that tourism marketers in Australia should allocate more resources to soccer events to manage and improve the general attitudes towards soccer. For policymakers, the model suggests that they should focus on building a better business environment to attract both business and leisure tourists.

Third, the developers of commercial social listening applications such as Google Alerts, Twitter Analytics, Brandwatch, NetBase, Pulsar, can adapt and integrate our framework into their current ones to offer their customers enhanced social listening functions, such as visual dashboards, display of trends, alert of potential demand hype or crisis, so as to enable tourism organizations to monitor their market and demand in real-time.

5.4. Limitations and further research

The testing of our forecasting framework is limited to one destination and data are based on a single social media Twitter, thus the findings regarding the demand influencing factors may not be generalisable to other destinations, and data collected from other social media may reveal different significant factors. In addition, the analysis of this study is based on English tweets which may miss out social media posts in other languages. Therefore, researchers are encouraged to adopt more advanced topic modelling and sentiment analysis algorithms, such as BERTopic (Grootendorst, 2020) and Multilingual BERT (Libovický et al., 2019), to conduct topic extraction and sentiment analysis on multilingual social media data. Using those advanced techniques, future research may apply our approach to examine alternative destinations using a combination of data collected across multiple social media platforms. For example, China is the biggest source market for Sydney and thus, future studies may extract data from Chinese social media such as Weibo, Ctrip, and WeChat. In this study, the frequency is based on yearly data for 10 years, which may also limit the application of this method. After averaging sentiment scores, the sample size of the forecasting model becomes only 10 which may restrict the performance of forecasting algorithms. Therefore, researchers are encouraged to collect high-frequency (e.g. weekly or daily) data to increase the sample size and therefore, performance for the forecasting model.

Moreover, further research could integrate and triangulate social media data with other data sources for demand forecasting, for example, the destination tourism organizations' sales data, customer interviews and surveys, industry updates, and news releases. Qualitative analytical techniques such as netnography can be used to supplement the quantitative analysis to provide a rich and in-depth understanding of the consumers and market trends. Chen et al., 2019, Lee and Taylor, 2005.

References

Balakrishnan, V., Lloyd-Yemoh, E., 2014. Stemming and lemmatization: a comparison of retrieval performances. Lect. Notes Software Eng. 2 (3), 262–267.
Balli, F., Balli, H.O., Jean Louis, R., 2016. The impacts of immigrants and institutions on bilateral tourism flows. Tourism Manag. 52, 221–229.
Balli, F., Balli, H.O., Tangaroa, N., 2015. Research note: the impact of marketing expenditure on international tourism demand for the Cook Islands. Tourism Econ. 21 (6), 1331–1343.
Balli, F., Shahzad, S.J.H., Uddin, G.S., 2018. A tale of two shocks: what do we learn from the impacts of economic policy uncertainties on tourism? Tourism Manag. 68, 470–475.
Becken, S., 2013. Measuring the effect of weather on tourism: a destination-and activity-based analysis. J. Trav. Res. 52 (2), 156–167.
Bi, J.-W., Li, H., Fan, Z.-P., 2021. Tourism demand forecasting with time series imaging: a deep learning model. Ann. Tourism Res. 90, 103255.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brilliant Maps, 2015. Top 100 International Tourist Destination Cities by Country. Retrieved AUG from. https://brilliantmaps.com/top-100-tourist-destinations/.
Chen, J.L., Li, G., Wu, D.C., Shen, S., 2019. Forecasting seasonal tourism demand using a multiseries structural time series method. J. Trav. Res. 58 (1), 92–103.
Chen, K.-Y., Wang, C.-H., 2007. Support vector regression with genetic algorithms in forecasting tourism demand. Tourism Manag. 28 (1), 215–226.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining.
Chen, Y., Lin, Z., Filieri, R., Liu, R., 2021. Subjective well-being, mobile social media and the enjoyment of tourism experience: a broaden-and-build perspective. Asia Pac. J. Tourism Res. 26 (10), 1070–1080.
Cho, V., 2003. A comparison of three different approaches to tourist arrival forecasting. Tourism Manag. 24 (3), 323–330.
Cho, V., 2010. A study of the non-economic determinants in tourism demand. Int. J. Tourism Res. 12 (4), 307–320.
Chu, F.-L., 1998. Forecasting tourism demand in Asian-Pacific countries. Ann. Tourism Res. 25 (3), 597–615.
Chu, F.-L., 2009. Forecasting tourism demand with ARMA-based methods. Tourism Manag. 30 (5), 740–751.
Darani, H.R., Asghari, H., 2018. Study of international tourism demand in Middle East by panel data model. Int. J. Cult. Tourism Hospit. Res. 12 (1), 80–88.
De Vita, G., Kyaw, K.S., 2013. Role of the exchange rate in tourism demand. Ann. Tourism Res. 43, 624–627.
Demir, E., Gözgör, G., 2018. Does economic policy uncertainty affect Tourism? Ann. Tourism Res. 69 (C), 15–17.
Demiralay, S., 2020. Political uncertainty and the us tourism index returns. Ann. Tourism Res. 84, 102875.
Destination, N.S.W., 2019. About Us. Retrieved July from. https://www.destinationnsw.com.au/about-us.
Dillette, A.K., Benjamin, S., Carpenter, C., 2019. Tweeting the black travel experience: social media counternarrative stories as innovative insight on #TravelingWhileBlack. J. Trav. Res. 58 (8), 1357–1372.
Dogru, T., Sirakaya-Turk, E., Crouch, G.I., 2017. Remodeling international tourism demand: old theory and new evidence. Tourism Manag. 60, 47–55.
Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J., Vapnik, V., 1997. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9, 155–161.
Feng, Y., Li, G., Sun, X., Li, J., 2019. Forecasting the number of inbound tourists with Google Trends. Procedia Comput. Sci. 162, 628–633.
Filieri, R., Lin, Z., Pino, G., Alguezaui, S., Inversini, A., 2021. The role of visual cues in eWOM on consumers' behavioral intention and decisions. J. Bus. Res. 135, 663–675.
Fourie, J., Santana-Gallego, M., 2011. The impact of mega-sport events on tourist arrivals. Tourism Manag. 32 (6), 1364–1370.
Frechtling, D., 2012. Forecasting Tourism Demand. Routledge.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Gao, Y., Su, W., 2019. Is the World Heritage just a title for tourism? Ann. Tourism Res. 78, 102748.
Gerakis, A.S., 1965. Effects of exchange-rate devaluations and revaluations on receipts from tourism. Staff Papers 12 (3), 365–384.
Ghimire, B., Shanaev, S., Lin, Z., 2022. Effects of official versus online review ratings. Ann. Tourism Res. 92, 103247.
Goh, C., 2012. Exploring impact of climate on tourism demand. Ann. Tourism Res. 39 (4), 1859–1883.
Goh, C., Law, R., 2002. Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. Tourism Manag. 23 (5), 499–510.
Gray, H.P., 1966. The demand for international travel by the United States and Canada. Int. Econ. Rev. 7 (1), 83–92.
Grootendorst, M., 2020. *BERTopic: Leveraging BERT and C-TF-IDF to Create Easily Interpretable Topics* in. https://doi.org/10.5281/zenodo.4381785. Zenodo.
Gunter, U., Önder, I., 2016. Forecasting city arrivals with Google analytics. Ann. Tourism Res. 61, 199–212.
Gunter, U., Önder, I., Gindl, S., 2019. Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. Tourism Econ. 25 (3), 375–401.
Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tourism Manag. 59, 467–483.
Hassani, H., Silva, E.S., Antonakakis, N., Filis, G., Gupta, R., 2017. Forecasting accuracy evaluation of tourist arrivals. Ann. Tourism Res. 63, 112–127.
Hays, S., Page, S.J., Buhalis, D., 2013. Social media as a destination marketing tool: its use by national tourism organisations. Curr. Issues Tourism 16 (3), 211–239.
Hu, M., Qiu, R.T., Wu, D.C., Song, H., 2021. Hierarchical pattern recognition for tourism demand forecasting. Tourism Manag. 84, 104263.
Hu, M., Song, H., 2019. Data source combination for tourism demand forecasting. Tourism Econ. 26 (7), 1248–1265.
Hu, Y.-C., Jiang, P., Lee, P.-C., 2019. Forecasting tourism demand by incorporating neural networks into Grey–Markov models. J. Oper. Res. Soc. 70 (1), 12–20.
Hutchins, B., 2011. The acceleration of media sport culture: Twitter, telepresence and online messaging. Inf. Commun. Soc. 14 (2), 237–257.
Imison, M., Schweinsberg, S., 2013. Australian news media framing of medical tourism in low-and middle-income countries: a content review. BMC Publ. Health 13 (1), 109.
Kock, F., Josiassen, A., Assaf, A.G., 2016. Advancing destination image: the destination content model. Ann. Tourism Res. 61, 28–44.
Költringer, C., Dickinger, A., 2015. Analyzing destination branding and image from online sources: a web content mining approach. J. Bus. Res. 68 (9), 1836–1843.
Kulendran, N., Divisekera, S., 2007. Measuring the economic impact of Australian tourism marketing expenditure. Tourism Econ. 13 (2), 261–274.
Lancaster, K.J., 1966. A new approach to consumer theory. J. Polit. Econ. 74 (2), 132–157.
Lanouar, C., Goaied, M., 2019. Tourism, terrorism and political violence in Tunisia: evidence from Markov-switching models. Tourism Manag. 70, 404–418.
Law, R., Li, G., Fong, D.K.C., Han, X., 2019. Tourism demand forecasting: a deep learning approach. Ann. Tourism Res. 75, 410–423.
Lee, C.-K., Taylor, T., 2005. Critical reflections on the economic impact assessment of a mega-event: the case of 2002 FIFA World Cup. Tourism Manag. 26 (4), 595–603.
Li, H., Goh, C., Hung, K., Chen, J.L., 2018. Relative climate index and its effect on seasonal tourism demand. J. Trav. Res. 57 (2), 178–192.
Li, H., Hu, M., Li, G., 2020. Forecasting tourism demand with multisource big data. Ann. Tourism Res. 83, 102912.
Li, H., Song, H., Li, L., 2017. A dynamic panel data analysis of climate and tourism demand: additional evidence. J. Trav. Res. 56 (2), 158–171.
Libovický, J., Rosa, R., Fraser, A.M., 2019. How Language-Neutral Is Multilingual BERT?, 03310 ArXiv, abs/1911.
Lim, C., 1997. An econometric classification and review of international tourism demand models. Tourism Econ. 3 (1), 69–81.
Lim, C., McAleer, M., 2001. Forecasting tourist arrivals. Ann. Tourism Res. 28 (4), 965–977.
Lipsey, R.G., Steiner, P.O., Purvis, D.D., 1987. Economics, Eight Edition (New York).
Liu, A., Pratt, S., 2017. Tourism's vulnerability and resilience to terrorism. Tourism Manag. 60, 404–417.
Liu, Y.-Y., Tseng, F.-M., Tseng, Y.-H., 2018. Big Data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. Technol. Forecast. Soc. Change 130, 123–134.
Long, W., Liu, C., Song, H., 2019. Pooling in tourism demand forecasting. J. Trav. Res. 58 (7), 1161–1174.
Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., 2014. Textblob: simplified text processing. Secondary TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/en/dev/. (Accessed 18 January 2022).
Martin, C.A., Witt, S.F., 1989. Forecasting tourism demand: a comparison of the accuracy of several quantitative methods. Int. J. Forecast. 5 (1), 7–19.
Martín, M.B.G., 2005. Weather, climate and tourism a geographical perspective. Ann. Tourism Res. 32 (3), 571–591.
Martins, L.F., Gan, Y., Ferreira-Lopes, A., 2017. An empirical analysis of the influence of macroeconomic determinants on World tourism demand. Tourism Manag. 61, 248–260.
McKinney, W., 2010. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference.
Mullen, L.A., Benoit, K., Keyes, O., Selivanov, D., Arnold, J., 2018. Fast, consistent tokenization of natural language text. J. Open Source Softw. 3, 655.
Neumayer, E., Plümper, T., 2016. Spatial spill-overs from terrorism on tourism: western victims in Islamic destination countries. Publ. Choice 169 (3), 195–206.
Njegovan, N., 2005. A leading indicator approach to predicting short-term shifts in demand for business travel by air to and from the UK. J. Forecast. 24 (6), 421–432.
Pak, A., Paroubek, P., 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. LREc.
Park, E., Park, J., Hu, M., 2021. Tourism demand forecasting with online news data mining. Ann. Tourism Res. 90, 103273.
Peng, B., Song, H., Crouch, G.I., 2014. A meta-analysis of international tourism demand forecasting and implications for practice. Tourism Manag. 45, 181–193.
Pham, T.D., Nghiem, S., Dwyer, L., 2017. The determinants of Chinese visitors to Australia: a dynamic demand analysis. Tourism Manag. 63, 268–276.
Ribaudo, G., Figini, P., 2017. The puzzle of tourism demand at destinations hosting UNESCO World Heritage Sites: an analysis of tourism flows for Italy. J. Trav. Res. 56 (4), 521–542.
Richards, G., 2018. Cultural tourism: a review of recent research and trends. J. Hospit. Tourism Manag. 36, 12–21.

Rugg, D., 1973. The choice of journey destination: a theoretical and empirical analysis. Rev. Econ. Stat. 64–72.
Shafiullah, M., Okafor, L.E., Khalid, U., 2019. Determinants of international tourism demand: evidence from Australian states and territories. Tourism Econ. 25 (2), 274–296.
Shen, M.-L., Liu, H.-H., Lien, Y.-H., Lee, C.-F., Yang, C.-H., 2019. Hybrid approach for forecasting tourist arrivals. In: Proceedings of the 2019 8th International Conference on Software and Computer Applications.
Sievert, C., Shirley, K., 2014. LDAvis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces.
Song, H., Qiu, R.T.R., Park, J., 2019. A review of research on tourism demand forecasting: launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. Ann. Tourism Res. 75, 338–362.
Song, H., Witt, S.F., 2006. Forecasting international tourist flows to Macau. Tourism Manag. 27 (2), 214–224.
Stepchenkova, S., Eales, J.S., 2011. Destination image as quantified media messages: the effect of news on tourism demand. J. Trav. Res. 50 (2), 198–212.
Sun, S., Wei, Y., Tsui, K.-L., Wang, S., 2019. Forecasting tourist arrivals with machine learning and internet search index. Tourism Manag. 70, 1–10.
Taspinar, A, Schuirmann, L, 2017. Python Package Index. https://pypi.org/project/twitterscraper/0.2.7. (Accessed 18 January 2022).
Tavares, J.M., Leitão, N.C., 2017. The determinants of international tourism demand for Brazil. Tourism Econ. 23 (4), 834–845.
Tsui, W.H.K., Balli, F., 2017. International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. Tourism Econ. 23 (2), 403–428.
Tsui, W.H.K., Balli, F., Tan, D.T.W., Lau, O., Hasan, M., 2018. New Zealand business tourism: exploring the impact of economic policy uncertainties. Tourism Econ. 24 (4), 386–417.
Um, S., Crompton, J.L., 1990. Attitude determinants in tourism destination choice. Ann. Tourism Res. 17 (3), 432–448.
Uysal, M., 1998. The Determinants of Tourism Demand, vol. 79. The economic geography of the tourist industry: A supply-side analysis.
Uysal, M., Crompton, J.L., 1984. Determinants of demand for international tourist flows to Turkey. Tourism Manag. 5 (4), 288–297.
Volgger, M., Taplin, R., Pforr, C., 2019. The evolution of 'Airbnb-tourism': demand-side dynamics around international use of peer-to-peer accommodation in Australia. Ann. Tourism Res. 75, 322–337.
Wan, S.K., Song, H., 2018. Forecasting turning points in tourism growth. Ann. Tourism Res. 72, 156–167.
Wang, C.-H., 2004. Predicting tourism demand using fuzzy time series and hybrid grey theory. Tourism Manag. 25 (3), 367–374.
Webster, J.J., Kit, C., 1992. Tokenization as the Initial Phase in NLP. COLING.
Wen, J., Lin, Z., Liu, X., Xiao, S.H., Li, Y., 2021. The interaction effects of online reviews, brand, and price on consumer hotel booking decision making. J. Trav. Res. 60 (4), 846–859.
Weng, L., Zhang, Q., Lin, Z., Wu, L., 2021. Harnessing heterogeneous social networks for better recommendations: a grey relational analysis approach. Expert Syst. Appl. 174, 114771.
Witt, S.F., Song, H., Louvieris, P., 2003. Statistical testing in forecasting model selection. J. Trav. Res. 42 (2), 151–158.
Wu, C., 2010. *Econometric Analysis of Tourist Expenditures* the Hong Kong. Polytechnic University.
Wu, D.C., Song, H., Shen, S., 2017. New developments in tourism and hotel demand modeling and forecasting. Int. J. Contemp. Hospit. Manag. 29 (1), 507–529.
Xie, G., Li, X., Qian, Y., Wang, S., 2020. Forecasting tourism demand with KPCA-based web search indexes. Tourism Econ. 27 (4), 721–743.
Xu, X., Law, R., Chen, W., Tang, L., 2016. Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs. CAAI Trans. Intelligence Technol. 1 (1), 30–42.
Yang, Y., Xue, L., Jones, T.E., 2019. Tourism-enhancing effect of World heritage sites: panacea or placebo? A meta-analysis. Ann. Tourism Res. 75, 29–41.
Yap, G., Allen, D., 2011. Investigating other leading indicators influencing Australian domestic tourism demand. Math. Comput. Simulat. 81 (7), 1365–1374.
Zhang, H.Q., Kulendran, N., Song, H., 2010. Measuring returns on Hong Kong's tourism marketing expenditure. Tourism Econ. 16 (4), 853.
Zhang, K., Chen, Y., Lin, Z., 2020. Mapping destination images and behavioral patterns from user-generated photos: a computer vision approach. Asia Pac. J. Tourism Res. 25 (11), 1199–1214.
Zhang, K., Lin, Z., Zhang, J., 2021. Tourist gaze through computer vision: differences between Asian, North American, and European tourists. Ann. Tourism Res. 88, 103039.
Zhang, Y., Li, G., Muskat, B., Law, R., 2021. Tourism demand forecasting: a decomposed deep learning approach. J. Trav. Res. 60 (5), 981–997.
Zheng, W., Huang, L., Lin, Z., 2021. Multi-attraction, hourly tourism demand forecasting. Ann. Tourism Res. 90, 103271.