EAI Endorsed Transactions
on Creative Technologies

Review Article EAI.EU

A Review of the Methods and Techniques Used in Tourism Demand Forecasting

R. Shakir Al Jassim [1,*], K. Jetly[1], A.Abushakra [2] , Sh Al Mansori [1]

[1]School of Computing and Information Sciences, UTAS University, Oman
[2]Business Information Technology, Princess Sumaya University for Technology, Jordan.

Abstract

The purpose of this paper is to discuss the methodology and results of researchers who conducted a study concerning the forecasting of tourism demand. In more detail, this study aims to examine and assess various studies about search engines, web traffic data, and social media data, specifically. Using an extensive database of indexed articles, we conducted the review with the goal of providing a solid understanding of the literature. The findings of our study revealed that few researchers integrate different data sources when forecasting tourism demand. Therefore, the authors of this paper decided to conduct a systematic review to provide researchers with a comprehensive overview of the importance of such data. This paper may inspire Omani researchers to undertake similar research based on its findings, which is currently lacking. Thus, this paper will improve understanding of how data sources affect forecasting accuracy and how modern technologies can support economic growth.

Received on 13 April 2022; accepted on 29 December 2022; published on 13 January 2022

Keywords: Social media, Forecasting Tourism Demand,Web traffic, Search Engines, Forecasting Methods

doi:10.4108/eetct.v9i31.2986

1. Introduction

The travel and tourism sector represents a complex set of social, cultural, human, technological, and economic ingredients. The scope of the tourism industry remains limited without having all the forces working positively. Tourism demand forecasting plays a significant role in this sector and has important consequences for destination decision-makers and tourism professionals. Traditional tourism demand forecasting lacks predictions of inaccurate data because such methods rely on structured statistical data [15]. Internet big data has opened valuable opportunities to gain additional insight into knowledge and information. This information can measure and monitor tourist behaviors while avoiding the failure of traditional forecasting methods. Internet big data is [21], therefore, an efficient supplement to traditional data sources [15].

Many studies have investigated the use of internet big data to forecast tourism demand, including data from search engines or website traffic. Studies on the use of search query data to forecast tourism demand have shown the usefulness of such data in improving the forecasting process. In addition to search query data, studies conducted by [6] [18] [19] have demonstrated that web engine data can be used to enhance the precision of forecasting hotel room demand, visitor demand, and tourism demand [17].

Social media is another vital source of data that contributed significantly to the development of new methods for understanding social and economic phenomena [35]. The impact of these data on the tourist industry's overall performance comes from tourists' intentions and sentiments. Many researchers contributed to gaining valuable insights from social media data analysis, such as web reviewers[35]. It is worth mentioning that social media data can also help practitioners provide information on the changes that occur within the tourist's performance and specify popularity trends related to the destination and local attraction.

There are a number of primary objectives underlying this study. A comprehensive review of internet data research in tourism forecasting is presented, along

*Corresponding author. Email: rashas.soh@cas.edu.om

with a comparison of three different types of internet data (such as web traffic, social networks, and search engines). The study will also examine various forecasting approaches on different internet data sets, including time series analysis, artificial intelligence techniques, and hybrid methods. Moreover, this study explores how differences in internet data affect the accuracy of tourist forecasting and how Internet data can be integrated into forecasting algorithms to optimize forecasting accuracy.

Additionally, in this paper, we will examine the importance of forecasting tourism demand in sustaining development in a number of countries. As a result of modern technologies, developing countries have access to new information that can be used for their development in the future. By analyzing web traffic data, for instance, economic transformation can be enhanced. Actually, the scope of the tourism industry remains limited here in Oman and the ingredients for building market information about tourism demand are still lacking despite their financial endowments. Academics, businesses and policy makers remain concerned about the digital divide. Investing in knowledge capacity powered by modern technology could contribute to the closing of the knowledge gap and the sustainability of the development of Oman. A number of countries will be examined in this paper regarding the importance of forecasting tourism demand and its role in sustaining their development. It is possible for Oman to maintain its growth if it adapts methods that will be discussed in detail in this paper to produce tourism demand information and create new methods to deal with challenges academics, businesses, and policymakers may face. As a result of this research paper and as a first step to investigate, researchers in this field will have many new opportunities to explore in the future.

This paper has several sections: in section 2, the selection of literature has been explained in detail. Section 3 includes an extensive discussion of the types of data available on the Internet. Throughout sections 4, 5, and 6, all studies are examined in terms of search engines, website traffic, and social media sources. We discuss the implications of existing forecasting models in the tourism industry, and the paper findings are in sections 7 and 8, respectively. The end of this paper offers a conclusion, potential future directions, and a list of limitations associated with the research presented here.

2. Problem Statement and Objectives

Recently, all various tourism industry sectors must ensure that tourism demand forecast remains accurate to prevent the risks of misplanning and improve business decision-making. Besides, policymakers also seek to have good predictions to determine pricing policies and implement a good business plan. Indeed, the tourism sector had a vital role in improving economic activities. The impact of this process on countries' economic development helped to create job opportunities and enhancing investment and government earnings. All the above explanations clarify why researchers and politicians are contributed to analyzing tourism's future evolution.

Data from multi source are essential for increasing tourism demand forecasting accuracy because of their contribution to enriching data with valuable information and avoiding problems the researchers aim to alleviate[10]. One of the developed strategies in [17] argued that combining data from various Internet platforms can influence forecasting models' performance. On the other hand, the importance of using social media data to forecast tourist arrivals was addressed in [25]. Indeed, social media data are rich with data that reflect tourist sentiment, but only using such data may hurt prediction accuracy. As part of the tourism demand forecasting process, integrating social media with web engine data will be superior to increase the performance of forecasting model accuracy.

It should be noted that the nature of tourism data is unclear in Oman and that no study has examined the factors involved in selecting a suitable data source and the consequences of using it for forecasting tourism demand. Further, it may be difficult to obtain reliable data when big internet data is only derived from web query data or web traffic data. Furthermore, using only social media data may reduce the validity of predictions. This study is therefore focused on identifying the crucial factors that will encourage Oman's researchers to conduct similar research in the future, something that is currently lacking in the country due to a lack of such research findings. Furthermore, the purpose of this study is to carry out a systematic review in order to provide researchers with a comprehensive overview of the importance of such a data source in their research and how it impacts tourism forecasting accuracy.

3. Methodology

This paper primarily discusses studies related to the issue of forecasting tourism demand, including forecasting methods and models. Several studies used Internet data sources and examined their effects on the tourism industry. However, studies written in English were the only ones considered in this paper.

To select the articles that will be used for this study, we used a Scopus database. It was selected for its status as one of the largest and most popular databases of abstracts and citations of peer-reviewed literature as well as its widespread use. Table 1 illustrates the main

selection criteria that were applied when selecting the papers.

Table 1. Methodology's criterion

Criteria	Approach/Method
Language	English
Year of Publication	2006-2021
Type of Publication	Full-text Publication

As part of this study, multiple search strings were used to collect data. Table 2 provides an overview of the primary keywords that were used to perform the search. In Figure 1, a description of the screening process can be found.

4. Aspects of Tourism Data

The key to forecasting tourism demand is to take a closer look at the type of data that represents tourists' attention, interests, and sentiments, which can play a role in influencing industry demand. These data have been classified into two categories based on their collection methods: structured and non-structured data.

Structured data is a collection of data that results from human interaction with a search engine or other online service such as Google search. A variety of studies incorporate these data to better forecast visitors' attention or interests. This type of data is beneficial for the DMO website as it can help allocate resources and detect tourist behavior alike. Data such as this can provide valuable information about how many visitors a site receives, how long they stay, and what materials they view. Unfortunately, such data is subject to certain restrictions that are not public and may result in additional challenges for collecting such data. Search engines, too, provide structured data that can be retrieved from them. Data is provided by actual searches entered by users on search engines such as Google, Baidu, and Bing provided in real-time throughout the day, week, and month. It has proven more cost-effective than other types of data and more effective in improving forecasting and supporting the management of the tourism industry. In any case, noises and irrelevant data can affect the accuracy of an analysis. In general, these data are helpful to management sectors in understanding tourists' focus quickly.

In addition to structured data, unstructured data is generated, either by software programs or by different methods based on real-world feedback, such as Facebook, Twitter, and Instagram. Those kinds of data do not conform to a prescribed model or schema. 80% of the data within organizations is unstructured. Data that is unstructured grows faster than structured data in the tourism sector. Crawlers are usually required for processing and storing unstructured data. This type of data cannot be directly processed or queried using SQL. On the other hand, the NoSQL database is a non-relational database that can be utilized to store both unstructured and structured data. This kind of data is used in a lot of tourism-related research. The reason is that the data contain information about tourism behavior. By collecting such data, managers and marketers can make timely decisions based on what is expressed in reviews and photos.

The use of multiple sources of data for forecasting tourism demand has recently been explored by researchers, such as Li et al. in 2018, Sun et al. in 2018, and Volchek in 2018 [19][27][31]. Often, this approach helps to enrich data with valuable information and avoid problems the researchers aim to alleviate[10]. One of the developed strategies in [17] argued that combining data from various Internet platforms can influence forecasting models' performance. In the context of tourism demand forecasting, integrating social media with web engine data will be superior to increasing forecasting model accuracy. Few researchers have considered the need to create a linkage between these two sources of data to predict tourism demand. Indeed, social data networks are rich with data that reflect tourist sentiment, but only using such data may hurt prediction accuracy. Despite the many benefits that can be gained when data is obtained from multiple sources, there are still several challenges such as:

1. Data might be owned or managed by organizations that refuse to share their data with the public.
2. A limited number of tools are available for processing and accessing multiple data sources.
3. The data obtained has various characteristics, formats, and sizes. As a result, the data processing and preparation require a significant amount of time and effort.

There is a wide variety of data available to monitor and forecast tourist behavior from various sources. A traveler's data is generated at different stages of the journey: before, during, and after their trip, as well as when planning their trip through Google searches and web visits. They can be used to track tourists in real time at the destination, and others to forecast future tourist activity. There are a number of online tools that can be used to forecast tourism behavior based on search engine queries and website traffic, while social media is becoming more popular as a tool of monitoring service quality.

By using Internet data in tourism forecasting, DMOs will better understand tourists' behavior, allocate

Table 2. Keywords

Topic	Keywords
Data on Web Traffic	(“Web Traffic Data”) AND ("Forecasting Tourism")
Social Network	(“Social Media" OR "Facebook" OR "Twitter" OR "Semantic Analysis”) AND ("Forecasting") AND ("Tourism") AND ("Demand")
Multi Source Data	(“Multi Source Data”) AND ("Forecasting Tourism")
Google Trends Data	(“Google Trends” OR “Google Insights”) AND ("Forecasting Tourism")

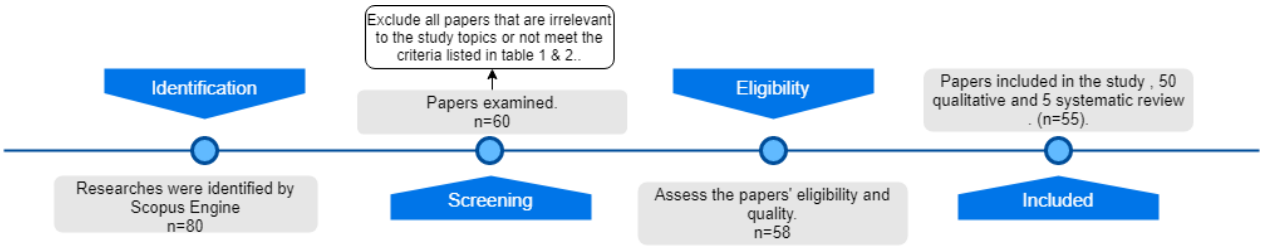

Figure 1. Diagram of the PRISMA process for conducting literature searches.

resources, and make more effective decisions to increase tourism demand. While this is true, search keywords selection, noise, and irrelevant information on the internet limit the accuracy of the data. Currently, studies of social media and multi-source data are limited, and researchers need to explore which data sources can be compared and analyzed rigorously and robustly. In fact, internet-based data sources can also provide more precise forecast accuracy than a single source.

5. Web Traffic Data for Forecasting Tourism Demand

Data on web traffic sheds light on who visits a website, how long visitors stay, and what materials or pages are viewed. Once this information has been accumulated over time, it can be analyzed to establish a set of metrics that can be used to optimize stakeholders’ e-business investments. In several peer-reviewed studies, web traffic data has been used to identify potential tourist interests [27][37]. These papers used traffic volume data to forecast hotel room demand and select the best econometric forecasting model.

In an attempt to forecast tourism demand and hotel demand using website traffic data, [37] has analyzed web traffic data. In their study, Yang et al noted that visitors’ interests might be used as a resource for predicting tourist arrivals in the future. Charleston, United States, South Carolina was chosen as the study destination by the authors. Visitors statistics for the 21st week of 2007 to the 17th week of 2011 were used as input variables in the ARMAX model. ARMAX models have lower errors than ARMA models when web traffic is added as an explanatory variable. Study findings suggest that data on website traffic for a given destination can be used as a tool to predict room demand for a given region’s hotels.

A different approach was taken by Pan et al in 2012) [27] , who predicted hotel demand in the US based on an aggregate search volume of five keywords. This study mainly compares the predictive performance of three autoregressive moving average families in two scenarios: one that uses search volume data as an explanatory variable and one that does not. As a result of the first scenario, search volume data proved more valuable than most other measured variables for forecasting. Furthermore, three widely used forecasting models were tested for their superiority over ARMA-type models. In terms of accuracy, the ARMAX models are superior to all three of their counterparts from the ARMA family. According to the study, search volume data is essential for forecasting hotel demand accurately.

As a means of improving forecasting accuracy, Gunter and Onder in 2016 [12] also used web traffic data. The data collected for this study were obtained from the Vienna Convention and Visitors Bureau’s website. From 2008 to 2014, data for this study were collected. As part of the data analysis, ten different Google Analytics metrics were used. The metrics included the number of sessions, the average duration of each session, the time spent on the site, the bounce rate, the number of new sessions, the number of

unique views, the number of returning visitors, and the number of social media network referrals. As part of the analysis, the data were aggregated to include unlimited sessions, unique views, and the number of users. Researchers conducted a benchmark study and used Bayesian VAR models in their study, which concluded that short-term and long-term performance varied significantly across models.

The usage of web traffic data in forecasting tourism is still a relatively new area of research. But this kind of information can assist in forecasting tourism in many ways. It is necessary to seek approval from the websites of DMOs to gather this information, which poses additional challenges in the collection process. Examples of such data collection applications include Google Analytics.

6. Social Media Data for Forecasting Tourism Demand

Nowadays, the Internet is one of the most significant sources of information when planning a trip to a popular tourist location. Unlike other tangible products, potential tourists cannot touch or feel tourism products and services before they travel to the destination, so information is the only thing they are left with to reduce the uncertainty related to traveling.

The Internet as a medium is characterized by a tremendous amount of information about travel destinations. Finding the data of interest in such a massive volume of data can also be challenging. Therefore, tourists turn to Facebook, Twitter, and other online travel forums, such as Trip Advisor, to minimize the amount of work they need to do.

For digital marketing, social media networks such as Facebook are commonly used by DMOs and hospitality organizations [24]. Compared to many other platforms, Facebook has the advantages of being accessible and having many users. In 2018, there were 2.9 billion accounts on Facebook (Kemp). Tourists may choose to minimize their efforts by searching Facebook for the information they need regarding their travel destinations rather than Google since they can easily find the information they need. Consequently, it becomes increasingly important to observe and evaluate the impact of these DMOs' efforts in improving the number of tourists arriving at the tourist attractions. According to [1], researchers used Facebook likes as a primary indicator to predict tourists' willingness to travel to various destinations. Moreover, it was observed that when present and past preferences are combined with past arrival numbers, tourism demand models are more accurate.

Recent studies have examined how DMOs' social media activity affects hotel reservations during short-stay holidays [5]. Researchers identified several factors that might affect hotel reservations in tourist destinations. The occupancy rate of hotels was predicted based on data collected from 10 Spanish DMOs. Researchers collected data on DMO's tweets and retweets and likes from users using Twitter's API. DMO's tweet data was analyzed utilizing Text Mining tools to differentiate between tweets regarding attractions, events, socialization, and marketing. Analyzing the data was also carried out using artificial neural networks. Researchers determined the rate of hotel occupancy for any destination can be successfully predicted based on the number of replies and retweets received by Twitter users and tweets from destination marketing organizations regarding events and tourist attractions.

Another common practice among tourists is filling out reviews for different properties where they stay during their vacation. An empirical study in [16] has examined how these reviews influence the tourists' preferences for hotels. Using pattern mining techniques on 118000 reviews on the TripAdvisor website, the researchers identified various hotel features and amenities that might appeal to international tourists. Hotel managers can make these adjustments to improve their properties to attract more international travelers.

Apart from the analysis of the user reviews on Trip Advisor, researchers also focused on the different posts posted in the Trip Advisor forums in seven cities across Europe over the last decade and collected data on approximately 2,660,000 postings made by 147,000 users [1]. The study presents a novel method of analyzing big data related to tourism. Additionally, it also included some variables that could be incorporated into traditional forecasting models. By using factor augmented autoregressive and bridge models, the authors were able to better forecast media trends than a single-variate model of Google Trends. Moreover, the study concluded that the forum language's complexity and the communication network's centralization play a significant role in predicting international arrivals at airports.

A lot of research has been done to design and develop a forecasting model for tourism demand based on the consumer data from internet platforms, such as search engines and social media. Most researchers have shown an increased accuracy of forecasting models based on data collected from the Internet (search engines and social media). However, these models still have some weaknesses in their design due to a lack of appropriate benchmarks or thorough evaluation. So this puts a question mark on the usefulness of data available on online platforms for assisting decision making at operational levels[30].

According to a study conducted in [36], social networks were utilized as a measure to predict inbound tourist arrivals. As a part of this study, the GA-LSSVR model was used to predict the monthly inbound

tourist traffic rates from the most frequent four countries and some other destinations. An evolutionary algorithm (GA) was used to discover the optimal variables for LSSVR, and a MAPE method was later applied to test the fitness function. The study analyzed data regarding the monthly arrival of tourists from four countries based on traditional predictors, Google indexes, and social network metrics. Based on the experiment results, the GA-LSSVR is more accurate than conventional forecasting methods when measured using social network data.

Among the approaches exploiting social media data from Twitter for the analysis of region- or culture-specific tourism trends, [2] presented a study that utilized social media data from Twitter to explore the trends of Halal tourism. Halal tourism trends were investigated by analyzing Twitter tweets from a social networking site. As part of this study, they examined 85,259 records using 19 keywords from both English and Bahasa Malaysia to extract the data. Based on the extracted tweets, the researchers attempted to determine the origin of tweets associated with halal tourism worldwide. Halal tourism was assessed from the perspective of several key populations. The tweets also indicate which are the most popular halal tourism destinations. Among halal tourism destinations tweeted, Japan ranks first, followed by Malaysia and Indonesia. The findings of this study may be useful to various stakeholders in the halal tourism industry for planning and decision-making in the future.

Zheng Xiang et al. [35], published a study on consumer-generated content and social media. Researchers have found that big data analytics can help address previously unresolved problems in hospitality research. This study uses big data analytics to identify and demonstrate the benefits of addressing major hospitality issues. In particular, the authors examine the relationship between guest satisfaction and hotel experiences. Using a text analysis technique, the reviews collected from Expedia.com were mined and examined to determine the impact of the guest experience on guest satisfaction ratings. There were various degrees of the guest experience, each weighing differently and consisting of different semantic components. Guests' experience and satisfaction are inherently linked, and these two elements are closely related to consumer behavior. The findings of this study suggest that big data analytics could offer new insights into variables that have been extensively investigated in the literature.

Several studies have also examined traveler interactions at hotel bookings by using information collected from online sources, such as blogs, travel websites, and social media platforms. Yang Li, et al [4] discuss the possibility that hotel managers might gain insights into travelers' interests. They identified four factors that impact customer decisions regarding online reviews, including the target audience of the review, the overall value of the set of reviews, how the reviews were framed, and whether a numerical rating was included with the written text. It was found that potential guests are most affected by the foremost negative information, especially when there is an overall negative review. In contrast, consumers are more likely to decide when the data is presented positively and numerical ratings. it was further concluded that consumers are more comfortable relying on information that is easy to process and understand when evaluating a hotel based on reviews. Additionally, consumers showed a high level of trust when they saw a positive review based on interpersonal service experience.

In conclusion, social media is an excellent platform to forecast tourist demand. It is noted that many researchers have utilized multiple social media platforms such as Facebook likes, tweets and likes on Twitter, and customer reviews on websites like TripAdvisor to develop mythologies to predict tourism demand with significant accuracy. Researchers have improved the accuracy of already available forecasting models by adding additional factors and variables obtained from multiple social media sources.

7. Google Trends Data for Forecasting Tourism Demand

Throughout 2006, Google released its data on search queries intensity (for any keyword) to academic and business researchers. The majority of the emerging research using data generated by Google searches is in the field of epidemiology [11], and economics disciplines [8]. Recent studies have investigated predicting future tourism trends using data from internet searches [25]. They used Google data to forecast arrivals and hotel occupancy of tourists from Western Europe and Asia. Google Trends data is real-time structured data, offering people access to a global view of how events are being handled globally. Google Trends data represents a sample of Google search data. The information provided is anonymous, categorized, and aggregated. It is possible to use such data to measure interest in a particular topic across searches conducted around the world, down to an individual city level. There is a wide range of data in the Google Trends data set. In addition to search terms, the data also includes images and videos (youtube).

The use of search terms on the internet is crucial in predicting tourism demand since the internet is now the leading source industry insiders use to gather information about those needs. Throughout the last decade, the global tourism industry has experienced unprecedented growth. Consequently, as noted by Song and Li (2008) [32], the number of articles and research

papers that focus on forecasting tourism demand has increased.

Researchers in [9] forecast tourist arrivals in Hong Kong using Google trends data in a study conducted in 2012, which has been suggested that Google Trends data about destinations could help determine how many people will visit Hong Kong. A quantitative analysis of their Google data was presented in detail, specifically focusing on the first two weekly observation periods to estimate the total monthly number of visitors. In this study, a simple seasonal AR model was used to predict the total number of visitors for each month of the year. However, contrary to Choi and Varian (2012), Bangwayo [6] and others extend their work in a bit of a different manner. Several countries in the Caribbean that are not included in Google Trends' list of destinations are used in their analysis. In this study, Mixed-data sampling (MIDAS) approach was used to avoid information loss.

A study published in Sangkon [29] suggests that short-term forecasts can be improved using data from an internet search. They further contributed to the literature by systematically constructing the google variable. A set of travel-related keywords was selected and combined by the authors according to their significance to minimize the mean squared and absolute forecasting error. As part of this study, many standard time-series models were used to examine the impact of the Google variable on forecasting.

By using Google Trends, the researchers at [25] also analyzed how the data provided by Google Trends could help in predicting the tourist demand for different countries, and in particular for specific cities. A comparison is made between two forecasting models using Web search indices or image indices from Austria and Belgium and two cities (Vienna, Barcelona). As a result of this study, web searches seem to be a good indicator of tourism demand, particularly in cities. Additionally, the author states that despite the potential of using Google Trends alone as a predictive indicator, this may not prove to be the best method since user motivation can be influenced by both internal and external factors, which are difficult to control. In this regard, it is clear that one cannot assume that every person searching for a destination on Google intends to travel to that destination. Still, their search does indicate a personal interest in that location. Thus, researchers should be careful when using Google Trends to assist them in forecasting research.

Search query volume data from both Google and Baidu permit researchers in [38] to estimate tourism volumes to a specific location. Specifically, the researchers sought to determine if the search engine data is a reliable indicator of tourist activity in China. This study also examined which data from search engines are more accurate predictors, Baidu or Google? Furthermore, the authors discuss in their research how a researcher might select among perhaps hundreds or thousands of possible queries the most appropriate queries for forecasting tourist volumes. A study found that search volume data from Google can improve forecasting accuracy significantly. Compared with ARMA, both types of search engine data significantly reduced forecasting errors.

By using Google Trends data, the author of the paper in [26] demonstrated that short-term forecasting could be made more accurate. This study uses a carefully selected set of travel-related keywords to minimize mean squared errors or mean absolute forecasting errors. In addition, the authors compare various standard time-series models with the Google variable to determine if it has an impact on forecasting performance. Upon analysis, the authors concluded that Google-augmented models are much better at short-term forecasting and produce more accurate predictions based upon out-of-sample data than from in-sample data.

Another study is conducted in [33] to assess how relevant websites, captured by search platforms, impact the consumption of tourism products in Cyprus. Cyprus was chosen as one of the candidates for developing a reliable predictor of international arrivals . It was decided to focus on Google as a search engine due to its global popularity and the fact that Google provides historical data on the search intensity of executed queries through its Google Trends platform. Two non-causality testing procedures were employed to analyze monthly data for Cyprus from 2004 to 2015. Researchers examined the arrival patterns of various destinations using various search engines and languages. Compared to the non-adjusted methodology, the adjusted approach was significantly more accurate in forecasting international visitor numbers.

Based on the study published in [7], it seems possible to forecast short-term tourist arrivals in various German holiday regions based on Google Trends data. As well, it is stated in the paper that, to facilitate the development of statistically sound forecasts for tourism demand in the future, cleaning up the Google Trends results was proposed to minimize their adverse impact on predictability.

This section of the paper summarizes the general principles of foreseeing tourism demand using Google Trends data. There has been a substantial increase in the popularity of tourism forecasting research based on Google Trends data since 2012. Google Trends data can be classified as Low-cost data and structured time series as determined by the findings in this section, making it relevant to various research fields. Google Trends might not be the most suitable tool to predict the future, but it can certainly be helpful when it comes to predicting the present, according to Choi and Varian[9].

The Google Trends database serves as one of the most reliable real-time information databases. It serves as a leading indicator in a variety of different industries, as Havranek [13] states. In general, Google trends help enterprises determine tourists' interests quickly and improve forecasting accuracy with the addition of a new driving factor. Even though Google Trends offers free, vast, and near real-time information, it does have some disadvantages, which include the following:

- Irrelevant data : Using Google trends may lead to inaccurate forecasts because the data contains noise and irrelevant information.
- Unreliable data: Google Trends data alone cannot be used to predict future events because it is not possible to assume that every individual who searches for a destination via Google intends on visiting that destination [25].

8. Tourism Demand Forecasting Methods and Approaches

Different methods of predicting tourism demand have been proposed by examining the research methodologies commonly used to predict traveler demand. Two general types of research are being conducted in this field: prediction models that utilize machine learning algorithms and prediction models that use numerical forecasting algorithms.

A new method for constructing a tourism demand forecasting model has been developed by Rob et al.[28]. This study utilized a deep learning model to predict monthly Macau tourist arrival numbers. This study shows that deep learning outperforms both SVRs and APNs in accuracy and predictive power.

A similar study was conducted by Sun [31] in which he proposed a new conceptual framework for Internet search indexing, machine learning, and forecasting tourist arrivals. This study used a kernel-based extreme learning machine (KELM). The proposed model was examined for its predictive capacity using data obtained from tourist arrivals from Beijing. The results show that KELM proves to be helpful, and the authors have confirmed that their proposed methodology is superior to other methods.

As part of Alvarez [3]'s research, genetic programming (GP) is incorporated into the study to predict the arrival of tourists from the United Kingdom and Germany on the Balearic Islands. This study found that the proposed model was highly accurate and stable compared to the NC model, the MA model, and the ARIMA model. According to the findings, GP is helpful and can be considered a promising method for predicting tourist arrivals.

Volchek et al.[34] conducted another study to determine the potential tourist value of the five most important museums in London. Different models were used to estimate London museum tourism demand for the period 2-6 months ahead. In comparison, the ANN model forecasts tourist demand more accurately one month in advance. Furthermore, the study concludes that policymakers and decision-makers should use different techniques according to their forecasting purposes. In the end, the SARMAX model has a lower overall forecasting error. If short-term decisions have to be made, the ANN model may be a better option.

Researchers have concluded that there is no likely chance of one approach beating another on every occasion. Time-series models, econometric strategies, and artificial intelligence techniques are some of the methods used in [32]. There are several methods of forecasting tourism demand available at the moment. However, it is necessary to note that although the new ways of forecasting have been proved more accurate when in some circumstances, there is no conclusive evidence that one method is better than the other in the forecast. Therefore, to achieve a better forecast, it is necessary to integrate both the new and older forecasting methods.

9. Forecasting tools for Toursim Demand

A big part of making informed decisions will be made through the collection, distribution, and storage of data, which will enable individuals, corporations, and governments to make better decisions, as well as giving them the ability to make more informed decisions. Tourism is one of the industries that develops a range of tools, known as forecasting tools, so that they can meet the challenges facing their societies with regards to tourism demand. It is no secret that web technology has evolved quickly over the last decade, supporting an increasing number of activities over the internet. Traditionally, forecasting tourism demand was done using standalone software, but in recent years Web-based applications have become increasingly popular [22]. Based on the results of a comprehensive survey conducted by Nikolopoulos et al (2003), five categories of Web-based forecasting services have been identified, and a general framework for the establishment of such services has been simplified. Additionally, the authors in [32] developed a new and innovative approach to generating tourism demand forecasts by utilizing Web-based technologies and advanced forecasting techniques. about how the In [14], Microsoft Time Series algorithms were used to predict tourist arrivals. Based on this algorithm, we were able to calculate how many tourists arrived in Sri Lanka in the last few decades based on the number of visitors that arrived. Additionally, this study provides an evaluation of the shortcomings of the current Microsoft Time Series algorithm. This is in relation to the consideration

Table 3. Literature on tourism forecasting using Google Trends

Year	Author	Approach/Method	Region
2012	Choi and Varian	AR model	Hong Kong
2014	Prosper F et	MIDAS	Caribbean
2015	Yang,Pan,Evans,James, Benfu	ARMA	China
2016	Sangkon Park,Jungmin Lee,Wonho Song	SARIMA,HW	Japan
2016	Park,Sangkon,Jungmin,Wonho	SARIMA,SARIMA with interventions,SARIMA with exogenous variables,HW	South Korea
2017	Önder	ADL, HW, Naïve-1	Vienna, Barcelona, Austria, and Belgium
2019	Bokelmann and Lessmann	SARIMA, DLM	German

of external factors that may affect the accuracy of predictions made using this algorithm. As part of this study, some possible solutions to this problem are also discussed.

10. Finding and Discussion

Tourism is a sector that depends heavily on information. It is becoming increasingly possible for researchers, managers, and policymakers to develop new forms of evidence through big data collection, which can be analyzed to make decisions based upon quantitative evidence rather than anecdotes, guesses, intuition, or past experience. Due to the wealth of big data emerging from the tourism industry, the sector will likely gain new insights into human behavior and activity that are both more precise and valuable. This will be a big boost for the industry and benefit both its customers and the people involved in its operations.

Big data and information technologies have a significant impact on tourism industry growth. The rise of the internet, mobile telephone, and various other computer applications led to the emergence of a new knowledge-based about tourists and one of the essential sources of tourism demand. Some academic papers have recently aimed at integrating big data into the tourism industry to empower the growth of forecasting tourism demand applications.

Tourism research has taken advantage of big data sources in recent years, leading to various technological advances. The data in big data can come from different sources: the users with their actions produce data such as online textual data and transaction data . Sentiment analysis, tourism behavior analysis, tourism marketing, and tourism recommendation are among the most useful types of tourism data. Moreover, there are few research studies on tourism that use transaction data. This could be because it is difficult to obtain such private information, which is usually available only through tourism associations or government agencies.

In recent years, there has been an increase in the number of scientific journals publishing articles examining data from Google Trends in the tourism sector, which is becoming increasingly relevant in the field of big data in tourism. Nonetheless, it should be noted that while there has been considerable progress in the literature on Google Trends and tourism, many articles fail to discuss the data's potential and limitations, the purpose for which the data was collected, and the methodology employed by the authors. The existence of the current limits may be attributed to several reasons. One of the reasons for this problem lies in the lack of knowledge among researchers on properly documenting their methodologies. Researchers involving Google Trends data also raise how search terms are selected. Researchers recommend more systematic methods based on the testing of search terms, with keyword selection as a starting point and strategies based on other statistical methods of predicting search results on the web.

Recent research in the tourism sector has examined social media as a potential data source. The majority of existing studies use very small sample sizes for their datasets, resulting in inaccurate results. For better results, many researchers recommend using a larger sample size. Researchers also encountered challenges when examining texts written in multiple languages, which is crucial in order to avoid information loss. Sentiment analysis of tourist reviews typically focuses on nouns, adjectives, and negative adverbs. Research has also emphasized the importance of examining the degree of polarity and intensity of language in tourist sentiment when studying tourist sentiment.

Web traffic data, as we mentioned previously, and according to the research presented in Section 5, have been introduced to tourism research and have proven to be useful. Despite the availability of such informative data, there are not many articles pertaining

to them. This could be because most transaction data are held by government agencies and tourism organizations and are challenging to acquire. In light of these factors, establishing such reciprocal cooperation between academia and the industry may be a possible way to promote emerging research and address practical issues associated with the tourism sector.

11. Research limitations

It is significant to note that this study relied heavily on research papers as primary sources. As far as this study is concerned, there are limitations in that it does not address the topics discussed in the books. A search on keywords highlighted in table 2 resulted in a selection of published papers. It is worth mentioning that the data were taken from Scopus. We recommend that future versions of the study include additional keywords and data from a variety of sources.

12. Conclusion

In its simplest form, forecasting is the process of organizing past information in order to predict one or more future events[20]. As the authors of this paper, we are particularly interested in saying something intelligent about the future of the tourism industry in order to reduce the risks associated with running a tourism business in the future.

Recently, all various tourism industry sectors must ensure that tourism demand forecast remains accurate to prevent misplanning risks and improve business decision-making. Besides, policymakers also seek to have good predictions to determine pricing policies and implement a good business plan. Indeed, the tourism sector had a vital role in improving economic activities. The impact of this process on countries' economic development helped create job opportunities and enhance investment and government earnings. All the above explanations clarify why researchers and politicians are contributed to analyzing tourism's future evolution.

In the context of tourism demand forecasting, integrating social media with web engine data will be superior to increase the performance of forecasting's model accuracy. Notably, few researchers have emphasized on creating an interplay between these two data sources to forecast tourism's demand. The main idea of integrating data from various sources is to enhance the proposed forecasting model's robustness.

Tourism demand forecasting is an essential component of destination decision-making and has significant implications for local tourism professionals. In fact, traditional tourism forecasting methods do not accurately predict data as they are based on structured statistical data. In the age of big data, there are many opportunities for individuals to gain additional insight into knowledge and information. As a result of this information, we can measure and monitor tourist behavior while avoiding the failure of traditional forecasting methods. As such, Internet big data is a complementary source of information to traditional data sources.

Recently, all various tourism industry sectors must ensure that tourism demand forecast remains accurate to prevent the risks of misplanning and improve business decision-making. Besides, policymakers also seek to have good predictions to determine pricing policies and implement a good business plan. Indeed, the tourism sector had a vital role in improving economic activities. The impact of this process on countries' economic development helped to create job opportunities and enhancing investment and government earnings. All the above explanations clarify why researchers and politicians are contributed to analyzing tourism's future evolution.

Data from multi source are essential for increasing tourism demand forecasting accuracy because of their contribution to enriching data with valuable information and avoiding problems the researchers aim to alleviate[10]. One of the developed strategies in [17] argued that combining data from various Internet platforms can influence forecasting models' performance. On the other hand, the importance of using social media data to forecast tourist arrivals was addressed in [25]. Indeed, social media data are rich with data that reflect tourist sentiment, but only using such data may hurt prediction accuracy. As part of the tourism demand forecasting process, integrating social media with web engine data will be superior to increase the performance of forecasting model accuracy.

Additionally, it has been noted that the mentioned learning algorithms in the literature are not always suitable for forecasting tourism demand. However, for any given pair of learning algorithms, there may be instances in which one algorithm is more effective than the other. It is more beneficial to determine factors that influence the performance of the models under consideration instead of identifying the best learning model.

References

[1] A.F. Colladon, B. Guardabascio and R. Innarella,(2019) Using social network and semantic analysis to analyze online travel forums and forecast tourism demand,Decision Support Systems,113075, https://doi.org/10.1016/j.dss.2019.113075.

[2] Ainin, Sulaiman; Feizollah, Ali; Anuar, Nor Badrul; Abdullah, Nor Aniza (2020) Sentiment analyses of multilingual tweets on halal tourism. Tourism Management Perspectives,34, 100658–. doi:10.1016/j.tmp.2020.100658

[3] Álvarez Diaz, M., Mateu Sbert, J. and Rossello Nadal, J. (2009) Forecasting tourist arrivals to Balearic Islands using genetic programming, Int. J. Computational Economics and Econometrics, Vol. 1, No. 1, pp.64–75.
[4] Beverley A. Sparks; Victoria Browning (2011) The impact of online reviews on hotel booking intentions and perception of trust,32(6),1310–1323. doi:10.1016/j.tourman.2010.12.011.
[5] Bigne, Enrique; Oltra, Enrique; Andreu, Luisa (2019) Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. Tourism Management,71, 490–503. doi:10.1016/j.tourman.2018.10.013.
[6] Bangwayo Skeete, P. F., and Skeete, R. W. (2015) Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tourism Management, 46, 454–464. doi:10.1016/j.tourman.2014.07.014.
[7] Bokelmann, Björn; Lessmann, Stefan (2019) Spurious patterns in Google Trends data : An analysis of the effects on tourism demand forecasting in Germany. Tourism Management, 75, 1–12. doi:10.1016/j.tourman.2019.04.01.
[8] Carriere-Swallow, Y., Labb e, F. (2011) Nowcasting with Google Trends in an emerging market. Journal of Forecasting, 32, 289-298.
[9] Choi, H., Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88, 2-9.
[10] Geva, T., Oestreicher-Singer, G., Efron, N., and Shimshoni, Y. (2017) Using forum and search data for sales prediction of high-involvement products. MIS Quarterly, 41(1), 65–82.
[11] Ginsberg, J., Mohebbi, M., Patel, R. et al.(2009) Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014. https://doi.org/10.1038/nature07634.
[12] Gunter, U., and Onder, I. (2016) Forecasting city arrivals with Google analytics. Annals of Tourism Research, 61, 199–212.
[13] Havranek, Tomas; Zeynalov, Ayaz (2018) Forecasting Tourist Arrivals with Google Trends and Mixed Frequency Data, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg.
[14] Hisham Nawzer, Manoja Weerasekara, (2018) A Case Study On Customizing The Microsoft Time Series Algorithm: Tourist Arrival Prediction, International Conference On Business Innovation (ICOBI), 25-26 August 2018, NSBM, Colombo, Sri Lanka.
[15] Huang, X., Zhang, L., and Ding, Y. (2017) The Baidu Index: Uses in predicting tourism flows– A case study of the Forbidden City. Tourism Management,58, 301-306.
[16] Li, Gang; Law, Rob; Vu, Huy Quan; Rong, Jia; Zhao, Xinyuan (Roy) (2015) Identifying emerging hotel preferences using Emerging Pattern Mining technique. Tourism Management, 46, 311–321. doi:10.1016/j.tourman.2014.06.015.
[17] Li, H., Hu, M., Li, G. (2020) Forecasting tourism demand with multisource big data. Annals of Tourism Research, 83, 102912. doi:10.1016/j.annals.2020.102912.
[18] Li, S., Chen, T., Wang, L., Ming, C. (2018) Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. Tourism Management, 68, 116–126. doi:10.1016/j.tourman.2018.03.006.
[19] Li, X., Pan, B., Law, R., Huang, X. (2017) Forecasting tourism demand with composite search index. Tourism Management, 59, 57–66. doi:10.1016/j.tourman.2016.07.005.
[20] M. Mariani, R. Baggio, D. Buhalis, C. Longhi (2014) Tourism Management, Marketing, and Development, Copyright © Marcello M. Mariani, Rodolfo Baggio, Dimitrios Buhalis, and Christian Longhi.
[21] Miah, S. J., Vu, H. Q., Gammack, J., and McGrath, M. (2017) A Big Data Analytics Method for Tourist Behaviour Analysis. Information and Management, 54(6), 771–785. doi:10.1016/j.im.2016.11.011
[22] Nikolopoulos, K., Metaxiotis, K., Assimakopoulos, V., and Tavanidou, E. (2003), A first approach to e-forecasting: a survey of forecasting web-services', Information Management and Computer Security, Vol 11, No 3, pp 146–152.
[23] Önder, Irem (2017) Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. International Journal of Tourism Research,doi:10.1002/jtr.2137.
[24] Önder, I., Gunter, U., ans Gindl, S. (2019) Utilizing Facebook Statistics in Tourism Demand Modeling and Destination Marketing. Journal of Travel Research, 1-14, 004728751983596. doi:10.1177/0047287519835969.
[25] Önder, I., Gunter, U., and Scharl, A. (2019) Forecasting tourist arrivals with the help of web sentiment: A mixed-frequency modeling approach for big data. Tourism Analysis, 24(4), 437–452.
[26] Park, S., Lee, J., and Song, W. (2017) Short-term forecasting of japanese tourist inflow to south korea using google trends data. Journal of Travel Tourism Marketing, 34(3): 386–397. doi:10.1016/j.tourman.2014.07.019.
[27] Pan, B., Wu, D. C., and Song, H. (2012) Forecasting hotel room demand using search engine data. Journal of Hospitality and Tourism Technology, 3(3), 196–210.
[28] Rob Law, Gang Li, Davis Ka Chio Fong, Xin Han (2019) Tourism demand forecasting: A deep learning approach, Annals of Tourism Research, Volume 75,Pages 410-423,ISSN 0160-7383, https://doi.org/10.1016/j.annals.2019.01.014.
[29] Sangkon Park, Jungmin Lee and Wonho Song (2016) Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data, Journal of Travel and Tourism Marketing, DOI: 10.1080/10548408.2016.1170651
[30] Schaer, Oliver; Kourentzes, Nikolaos; Fildes, Robert (2018) Demand forecasting with user-generated online information. International Journal of Forecasting, S0169207018300505. doi:10.1016/j.ijforecast.2018.03.005.
[31] Sun, S., Wang, S., Wei, Y., Yang, X., and Tsui, K.-L. (2017) Forecasting tourist arrivals with machine learning and internet search index.IEEE International Conference on Big Data (Big Data).doi:10.1109/bigdata.2017.8258439.
[32] Song, H., and Li, G. (2008) Tourism demand modelling and forecasting : A review of recent research. Tourism Management, 29(2), 203–220. doi:10.1016/j.tourman.2007.07.016
[33] Theologos Dergiades, Eleni Mavragani, Bing Pan (2018), Google Trends and tourists' arrivals:

Emerging biases and proposed corrections,Tourism Management, Volume 66, Pages 108-120, ISSN 0261-5177, https://doi.org/10.1016/j.tourman.2017.10.014.

[34] Volchek, K., Liu, A., Song, H., Buhalis, D. (2018) Forecasting tourist arrivals at attractions: Search engine empowered methodologies. Tourism Economics, 135481661881155. doi:10.1177/1354816618811558

[35] Xiang, Z., Schwartz, Z., Gerdes, J. H., and Uysal, M. (2015) What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management, 44, 120–130. doi:10.1016/j.ijhm.2014.10.013.

Acknowledgement

The research leading to these results has recieved funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program. TRC Block Funding Agreement No. MoHERI/BFP/UTAS/2022

[36] Yuan, Fong-Ching (2020) Intelligent forecasting of inbound tourist arrivals by social networking analysis. Physica A: Statistical Mechanics and its Applications, 124944–. doi:10.1016/j.physa.2020.124944.

[37] Yang, Y., Pan, B., and Song, H. (2014) Predicting hotel demand using destination marketing organization's web traffic data. Journal of Travel Research, 53(4), 433–447.

[38] Yang, X., Pan, B., Evans, J. A., and Lv, B. (2015) Forecasting Chinese tourist volume with search engine data. Tourism Management, 46, 386–397. doi:10.1016/j.tourman.2014.07.019.