EMERALD POINTS

BIG DATA ANALYTICS FOR THE PREDICTION OF TOURIST PREFERENCES WORLDWIDE

Dr. N. PADMAJA
Dr. RAJALAKSHMI SUBRAMANIAM
Dr. SANJAY MOHAPATRA

BIG DATA ANALYTICS FOR THE PREDICTION OF TOURIST PREFERENCES WORLDWIDE

This page intentionally left blank

BIG DATA ANALYTICS FOR THE PREDICTION OF TOURIST PREFERENCES WORLDWIDE

BY

Dr. N. PADMAJA
SRI Padmavati Mahila Visvavidyalayam, India

Dr. RAJALAKSHMI SUBRAMANIAM
Talaash Research Consultants, India

AND

Dr. SANJAY MOHAPATRA
Batoi Systems Pvt Ltd, India

United Kingdom – North America – Japan– India
Malaysia – China

Emerald Publishing Limited
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2024

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-83549-339-7 (Print)
ISBN: 978-1-83549-338-0 (Online)
ISBN: 978-1-83549-340-3 (Epub)

CONTENTS

This page intentionally left blank

LIST OF FIGURES AND TABLES

This page intentionally left blank

LIST OF ABBREVIATIONS

BI	Business Intelligence
IoT	Internet of Things
LDA	Latent Dirichlet Allocation
NLP	Natural Language Processing
PMI	Ponitwise Mutual Information
TF-IDF	Term Frequency-Inverse Document Frequency
UGC	User Generated Content

This page intentionally left blank

PREFACE

The embracement of big data and machine learning techniques in tourism industry has been presumed as a mean to direct the challenges and issues of the smart development of tourism destinations and organizations by redesigning the margins of the competition and improving its pattern and design as information intensive business. In the process of their journeys and in their decision-making processes, people who travel contribute to the generation of a huge flow of data created by sensors and micro-devices disbursed on the urban and other areas of interest for travellers. All these information is a potential base for creating smart destinations and also for improving the tourism organizations' potential to customize their products and service offerings. Though, the real execution of such inventive forms of data-driven value generation in tourism continues to be more restricted to the theory or executed in a few excellent cases. Indeed, big data and machine learning techniques in tourism continues to be an unclear concept and a subject of investigation that necessitates closer analysis from an extensive range of field and research methods. Comprehensibly with this viewpoint, the current special issue aimed to deal with development of the discussion on Smart Tourism by concentrating on big data as a promising model which is re-shaping the concept and practice of tourism. In particular, the underlying principle of the study is to explore benefits, importance and challenges of big data in smart tourism with theoretical and experiential contributions and intend to derive meaning for smart tourism. Known the changes that digital era is introducing in the tourism industry and with the intention of supporting managers that function in this industry, it is significant to look into various challenges and issues included on how big data could contribute to decision-making processes in tourism organization. This issue makes an effort to cover these subjects and to consider the status of big data and smart tourism study. The study also intends to analyze the benefits of big data and various machine learning techniques evident for those in the area of tourism and to comprehend the effectiveness of, and opportunities presented within. The tourism sector is a wide business area, but the keystone of this sector is the ambience part of it and hotels are considered the influential in that field. Running a profitable business could be

millstone for every tourism organization and necessitate hotel organizations to hire more staffs in which they have to afford financially and manually. Three different techniques have been used in the current study, i.e. association rule mining, point wise mutual information technique (PMI) and neural network such as multiclass and multilabel classification algorithms. These three classification models have been used in order to check the accuracy of the results, i.e. TF-IDF, Doc2Vec and LDA. In order to generate tourists' profile and information, different big data algorithms and machine learning techniques have been recommended and applied to tourism text data assessment. The study concluded that accuracy is high for the TF-IDF approach. Successful cases are analyzed by elucidating the main opportunities and challenges for the generation of official tourist information. Throughout the growth and interpretation of the knowledge of tourism, every time as a result of the poor understanding of the leader or extreme dependence on big data, there might be a misunderstanding that big data could do anything. It results in blind decision-making, and consequently many technical inputs might not bring in the desired effect. The awareness and understanding of big data for users is considered important. The use of big data is a conventional application in the marketing policy. The study contributes to literature on value generation from big data and it's relating technologies. This paper also contributed to explain the role of big data in transforming the structure of tourism industry, which is dominated by deep-rooted business logics. The study offers original contributions which progress knowledge regarding challenges and importance for smart tourism with respect to big data. It also deals with state-of-the-art technique and related practice, and emerging challenges and issues and areas for future investigations are examined in this study.

1

INTRODUCTION

1.1 BIG DATA ANALYTICS IN TOURISM SECTOR

One of the major aspects of the digital era is the augmented growth of data. The amount of data generated by internet users these days is increasing unabated and IT people project 43% increase in yearly data generation by the end of 2020 (Reddy, 2016). Big data introduces undeniably great prospects to the tourism sector, by providing apprehensions to consumer preferences and facilitating them to restructure the complete travel experience (Akerkar, 2012). The period between 2005 and 2012, referred to as Analytics 2.0, is categorized by the employment of big data and analytics largely by online portals such as Google and Yahoo (Davenport, 2013). One of the most important aspects of big data indicates its large-scale bulk data sets are created every second of the day from a large number of sources, such as mobile phones and digital devices (Bello-Orgaza, 2015). For instance, reputed networks like Facebook produces over 500 terabytes of data (Maria, 2018), facilitating the growth of new data mining techniques.

The application of big data in the tourism sector has been presumed as a tool to direct the challenges of the smart development of tourism organizations (Alcántara-Pilar et al., 2017), by restructuring the extremities of the competition and improving its design as information intensive sector (Law et al., 2014). In their journeys and their decision-making processes, tourists play an important role in the generation of a huge flow of information created by sensors, micro-devices disseminated on the urban and developing regions of interest for travellers. Ardito et al. (2019) pointed out that all this information is a potential base for making better destinations and also for improving the tourism firms' competence to customize their deliverance of products and services.

Big data refers to data sets compiled from a large number of sources and generated in large amounts and high rate of recurrence. These records are so large and compound that traditional data analytical tools and techniques could not effectively handle them. Two different types of big data are involved in official statistics, i.e. classifications by type and classifications by size (Dornik & Hendry, 2015). Big data is obtained from different sources and this information is fundamentally unstructured and unregulated to a certain extent. These data are usually obtained from a variety of sources including Facebook, Twitter, YouTube, etc., i.e. presented in a digital form on social networks and are exemplified by high frequency rate. According to Upadhyaya and Kynclova (2019), the second classification includes conventional business systems such as administrative data types, records stored in public agencies, tax records and data generated by businesses, for example data on consumers, transactions, costs, proceeds and liability.

In the age of digital transformation, big data analytics have taken a critical role in transforming the universal travel and giving major challenges and opportunities for developed companies, and also new players into the tourism sector. Palomo (2016) put forward that the tourism organizations could get significant information on big data for forecasting tourists' demand, allowing better decision-making processes, dealing with knowledge flows and communication with tourists and delivering the best service in a more competent and efficient way.

In the present day, many tourism promoters are hyper-targeting customers with concurrent mobile advertisement promotions to drive conversions. They employ analytics to find out how external variables, such as the broader economy, competitive bidding and even the weather conditions, could influence their organizational performance (Nichols, 2013). They employ analytics to tailor their offerings and enhance their consumer engagement (Ransbotham & Kiron, 2018). The new benefits of crunching big data analytics are on the basis of the detection of how to enhance productivity levels and quickness to improve the financial performance of business (Wang et al., 2018). With the arrival of big data analytics, the most recent technological growths have empowered many tourism businesses, particularly airlines and hotels, to deal with their processes in a more effective and economical way (Camilleri, 2019). With data and analytics on the consumers' behaviours, numerous travel businesses are making use of channel-based distribution.

Data analytics encourage a customer-driven decision-making process. The extrapolative methods created by big data assist service providers and tourism zones in identifying leads, isolating them in real-time and concentrating on the ones with increased conversion probability. Almost 90% of travellers share

images, happenings and experiences of their travels on social media networks. Millions of travel-based opinions are shared on social media every single day. Handlowa (2019) stated that the application of big data from online networks might help to comprehend users' behaviour based on early warning, real-time monitoring and real-time feedback.

Big data management generally represents filter data for reliability, to combine data sets from various reliable sources and to encode acquired data set for the purpose of security and privacy. To be more precise, the underlying principle of big data management is to guarantee trustworthy data which are easily accessible, controllable, rightly stored and protected. Integration of manifold data sources, for example heterogeneous data sources, either in the form of structured or unstructured data set, for example consumer feedback, reservation and booking records and navigation data, in both consumer and dealer sides (Höpken et al., 2015), might exhibit a new perspective that is not capable of detecting with conventional approaches. In terms of the tourism sector, the volume of online data is increasing like never before. In particular, Australia (Stantic & Pokorný, 2014) was the first country that used online surveys in the form of structured data, and in 2013, around 40 established companies and 150 social research centres delivered services using online surveys.

Big data is large data sets now capable of being accumulated, stored and deciphered through contemporary and innovation technology. The tourism sector has broadly implemented internet technologies as a means of interaction with the targeted customers to improve their management effectiveness, enhance quality of service and consumer experiences. Information technology contributes to a developing managerial performance by providing information to set up marketing strategies that are compliant with customers, pricing and public relationships. The better growth of these services might gain from prediction on the basis of consumer data accumulated through their utilization of technology trend. As a result of the systematic decision-making with great accurateness level, the big data analytics gains lots of attention as decision-making is key in business. Sheoran (2017) specified that big data analytics can be considered more than just size; considerably, it is able to identify pattern, trends and perspective and make constructive predictions.

Big data is an innovative trend in itself that it emerges from the utilization of new sources of information, for example digital e-commerce forum associated with accommodation and food; social networks with tourists' remarks and viewpoints; record of electronic searches; mobile phone records from domestic and international travellers. The big data programme puts forward new sources of data for a better comprehension of the observable fact of

tourism. This endorses synchronization between the public and private enterprises to increase the profits of tourism and reduce its negative effect. As pointed out by Zerba (2018) in this way, big data allows quick and clear-cut answers to assist orientate government policies, creating a fundamental support for the local tourist board's policies, and also to endorse opportunities to local enterprises and businesses.

The surfacing of user-generated content on online forum and social media platform has paved the way for endless opportunities for practitioners to pay attention to market stakeholders like consumers, workforce, investors along with the media (Abbasi et al., 2016). Consequently, businesses are progressively collecting and exploring data from different sources to enhance their customer-centric market. This data collection might probably bring in new revenue flow and build long-standing loyalty. Ultimately, businesses can employ this database to provide promotional content to remind consumers regarding their offerings. As pointed out by Abbasi et al. (2016), big data has possibly significant inference for theory. From a perspective of information value chain, big data artefacts have different inference for both knowledge acquirement and for assessment and actions, which integrate system usage, operation and contentment.

In this IT age, tourism sector largely incorporate various information and communication technologies; consequently, different sources of big data have evolved to give a large amount of concurrent and synchronized information with space–time tags on travel activities (Pan & Yang 2016). Yang and Stienmetz (2018) demonstrated that the digitalization of all characteristics of tourism value-creation has brought in the opportunity for tourists to better comprehend perceptions, choices, behaviour and assessment of tourists and also the business performance of tourism-based transactions in a single destination (Stienmetz & Fesenmaier, 2013). The hype and anticipation of big data signify an unparalleled opportunity in tourism-based planning by giving more advanced, wider-scale and substantial knowledge of tourism demand and supply at various geographic points.

The use of big data applications to tourism planning and organization is complicated since it necessitates technological proficiency (Chen et al., 2012). Nonetheless, this is not the only aspect defining its application. Big data is a set of data accumulated from different sources with various formats, such as texts, images, voices, etc. They might be obtained from Instagram, Twitter, blogs and as well, from business databases and sensors.

Fig. 1.1 represents the most commonly considered sources of big data. Similar to other classifications, some elements might be instinctively found within different groups. The five most important properties of big data include

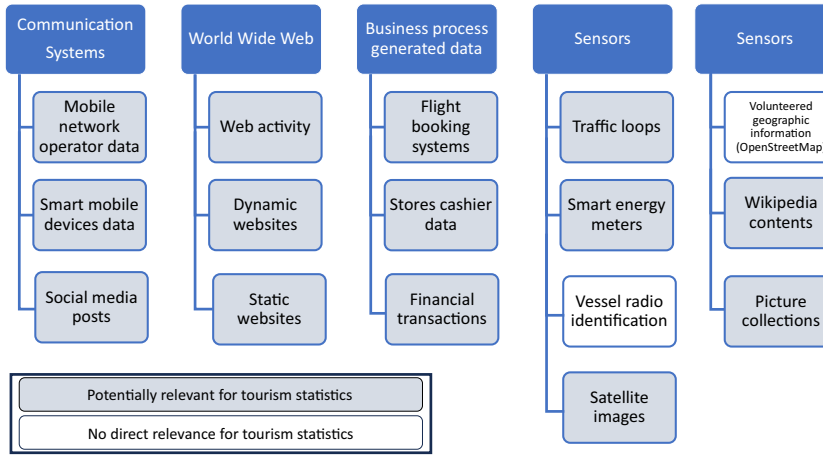

Fig. 1.1. Sources of Big Data.

5V, i.e. variety, velocity, volume, verity and value. As stated by Guilarte and Quintáns (2019) it does not only allow the measurement of tourists' physical movements, but as well financial transactions, thus becoming an essential tool for designing, executing and propagating invention systems within the domain of tourism.

Big data can be employed ultimately for Business Intelligence (BI) functions. More particularly, big data could empower business intelligence, but it should be expanded in an effective way. Every tourism company intends to control its decision-making and marketing approaches, strategies and tools to accomplish and retain sustained competitive edge. In this framework, as pointed out by Hopken and Fuchs (2016) big data could make a difference for the business intelligence of tourism organizations, assist them make enhanced strategic decisions and build value.

A tourism destination is a set of services that present travellers an integrated experience. Big data in tourism industry, in specific, has the prospective to include value appropriate to the aspects put forward by Brown et al. (2011) with regards to transparency and instant feedback, fast analysis, customers' division, decision-making and services innovation. Vecchio et al. (2015) pointed out that social big data are important sources for the growth of regional destinations in a social way. Social big data analytics has played an important role in deepened knowledge of what travellers demand, to generate involvement and trust. Vecchio et al. (2015) emphasized that the contribution of major influencers as authorized storytellers could optimistically play a role

in the success of initiative, build awareness and a communal culture of sustainable utilization.

Big data generally refers to the technological and methodological challenges in handling so large volumes of speedily varying sets of data. In addition to that, there is a demand for an enhanced set of particular competence and resources, and a various method is considered necessary relating to the one that for the last era has categorized the compilation and the analysis of data (Chen et al., 2014). Big data is expected to be a key for Business Intelligence endeavours. Advanced analytics and enhanced sources could give a deeper viewpoint on the data which could profit from more structured experience. The interpretation layer presented by business intelligence tools could therefore be critical to making highly developed big data analytics actionable (Liebowitz, 2013). Baggio (2016) specified that the big data analytics in tourism collect client information with the help of his/her web search activities, interactions in social media, locations tracking, personal interactions etc. which are directly associated his/her experience and choice.

In the smart tourism system, 'Big Data analytics is considered necessary for processing, developing and creating data in order that it could be employed to update operational decisions' (Gretzel et al., 2015a, p. 560). Big data analytics is proven to be a fundamental tool to generate real business intellect in the tourist destinations which changes the conventional information systems and contributes to create knowledge in the destination. It is the place where the fourth V of the big data is considered basic, i.e. value, and data should be evaluated in such a means that it could generate Value, to be exact, it facilitates informed decision making. Bernabeu et al. (2016) emphasized that one of the biggest challenges in employing the big data is the transformation of organizational cultures, i.e. the demand to spend considerable amounts of money, and to have specialized technical profiles, is also considered as another factor.

Internet of Things (IoT) can be categorized as one of the major markets of big data applications (Chen et al., 2014). The difficulties lay at various levels, for example data acquirement, storage, searching, analysis, organization and visualization. In addition, there are various security and privacy problems involved particularly in distributed data-driven management. Progressive data analysis is considered necessary in understanding the relationships amongst features and deals with data. For example, data analysis facilitates an organization to extort effective insight and examine the patterns which might influence positively or negatively the performance of the business (Wang et al., 2016). Oussous et al. (2016) emphasized that data scientists are going through various challenges when handling big data and one major difficulty is how to

collect, analyze, incorporate and store, with less hardware and software necessities, huge data sets created from distributed sources.

The tourist sector is extremely vast where in visitors are either from various locations or prefer to pay a visit to different locations. This involves numerous players and therefore makes the system more complicated and in such complex circumstances, big data analytics could be convenient for demand forecasting, inventory management, enhanced decision support system, travellers' trustworthiness, new products and services for visitors/travelers. Alaei et al. (2019) emphasized that the above mentioned functions are tackled by big data, which is not practicable with other existing techniques.

The implementation of big data in the tourism sector is evidently far behind other sectors, for example biological medicine, aerospace, e-commerce, and so on. As big data processing technology becomes increasingly mature and excellence, it is important for tourist enterprises to get better at employing big data analytics for transformation and innovation. In the age of big data analytics, the tourist businesses could objectively assess customer behaviour with increased accuracy. Simultaneously, tourist businesses can rely on big data analytics to keep up enterprise strategy development to get used to change. In the progression of planning development strategy, the tourist businesses could carry out customer data mining with special concentration on traveller demand, utilization need and customer feedback. Luo et al. (2016) pointed out that big data analytics could guide tourist business policy to serve custom.

The business-related value of big data integrates two characteristics. Primarily, it could assist tourist sector to lessen cost and enhance productivity. Secondly, it assists tourist ventures to better position tourist market, for example data analysis, product planning and better deal with customer demand, enhancing core competitiveness and value creation. Chen (2017) emphasizes that big data analytics is driving the amalgamation of tourism sector and also pointed out that technology innovation brings product market invention into tourist businesses and ultimately transform the mode of tourist business development.

The central part of big data is to forecast results; and 'four V, i.e. volume, variety, velocity and value' can be considered as its important characteristics. Intelligent tourism employs new technology of IoT, all through the Internet, with the assistance of the convenient terminal device, to dynamically identify tourism resources, related activities and other information that essentially makes individuals identify the information in due course, organize and change their journey plans, in order that to comprehend the impact of intelligent perception of different tourism information and expedient utilization of it. In

service-based organizations, they depend on big data to give adequate beneficial resources. Lingyun et al. (2015) pointed to that with the assistance of the related relationships in 'big data', intelligent tourism could possibly be 'intelligent' development.

Businesses using big data could develop their competitive advantage in a global market and make large amounts of information available regarding customers on the Internet (Verhoef et al., 2016). Some of the challenges included in big data are data sharing and security, extraction of data, collection, processing, examination and reporting (Gerard et al., 2016). Despite its challenges, big data have as well considered as a large or major source for Business Intelligence (BI) endeavours that intended at creating, distributing and capturing consumer value (Verhoef et al., 2016). Big data might result in better and more knowledgeable decision making for individual persons and managements and, therefore, build value for stakeholders and as well consumers.

1.2 PROBLEM STATEMENT

The most important issue in tourism industry for customers is contentment from services, while for service providers, it is to make a good name in fulfilling commitment and responsibility while keeping the expenditures at minimum. The machine learning algorithms and techniques in processing the big data is very essential nowadays in any sector. These days, big data is one of the main problems of information management in digital era and as well one of the major opportunities in tourism sector for optimal utilization of highest information. Big data could pave the way for smart travel. Remarkable development of these dataset sources has instigated new Strategies to comprehend the socio-economic occurrence in various fields. The big data analytical approach highlights the volume of data collection and exploration with an unparalleled extent, intensity and scale for resolving the issues of real life and employs it. The real execution of such inventive models of data-driven value generation in tourism sector remains more restricted to the theory or executed in only some commendable cases. In future Big data analytics may belong to those firms which are capable to shape and provide best and effective travel experience of tourist travelers. The main aim of the study is to examine the Application of Big Data Analytics in Predicting Travel Behaviour of International and Domestic Tourists – A study from Indian perspective.