INVESTIGACIONES TURÍSTICAS
REVISTA CIENTÍFICA
ISSN-2174-5609
Universitat d'Alacant
Universidad de Alicante
Institut Universitari d'Investigacions Turístiques
Instituto Universitario de Investigaciones Turísticas

Revista Investigaciones Turísticas, nº 26 (2023), pp 350-374.
ISSN: 2174-5609
DOI. https://doi.org/10.14198/INTURI.24629

Cita bibliográfica: Mendieta-Aragón, A. y Garín-Muñoz, T. (2023). Consumer behaviour in e-Tourism: Exploring new applications of machine learning in tourism studies. *Investigaciones Turísticas* (26), pp. 350-374. https://doi.org/10.14198/INTURI.24629

Consumer behaviour in e-Tourism: Exploring new applications of machine learning in tourism studies

Comportamiento del consumidor en e-Turismo: Explorando nuevas aplicaciones del aprendizaje automático en los estudios de turismo

Adrián Mendieta-Aragón, Programa de Doctorado en Economía Interuniversitaria (DEcIDE), Universidad Nacional de Educación a Distancia, España
amendieta@cee.uned.es

Teresa Garín-Muñoz, Universidad Nacional de Educación a Distancia, España
mgarin@cee.uned.es

ABSTRACT

Digital markets have altered how economic agents interact and have changed the behaviour of tourists. In addition, the COVID-19 pandemic has shown that it is necessary to constantly monitor the evolution of digital consumer behaviour and the factors that influence it, as they are dynamic elements that evolve over time. This paper analyses digital inequalities and validates the main factors influencing tourists to book online tourism services. This research uses a set of microdata with 69,752 and 23,779 observations to analyse the booking mode of accommodation and transportation services, respectively, obtained from the Resident Travel Survey of the National Statistics Institute of Spain during the period 2016-2021. The article confirms variations in the online consumer profile and in the trip's characteristics. One of the most relevant findings is the narrowing of the generational gap in the online contracting of tourist services. However, there are remaining digital inequalities, such as regional inequalities and others based on the education level and income of tourists. It is also highlighted that different types of trips, depending on the destination, the type of accommodation or transport have a different propensity to be booked through digital purchase channels. The accessibility to big data sources and recent advances in machine learning models have also made the methodologies for analysing digital consumer behaviour evolve and must be incorporated into tourism studies. This study compares the predictive performance of different methodologies in the context of e Tourism. In particular, we evaluate the potential predictive power that could be obtained using machine learning techniques to explain consumer behaviour in e-Tourism and use it as a benchmark to compare it with the results obtained using traditional statistical methods. The selected predictive evaluation metrics show that the logistic regression statistical model outperforms the predictive power of the Multilayer Perceptron neural network and presents values very close to the maximum predictive power achieved by the Random Forest algorithm.

Fecha de recepción: 20/02/2023   Fecha de aceptación: 02/06/2023

Keywords: e-Tourism, online travel booking, consumer behaviour, big data, logistic regression, machine learning, dynamic behaviour.

RESUMEN

Los mercados digitales han alterado la forma en que interactúan los agentes económicos y han cambiado el comportamiento de los turistas. Además, la pandemia de COVID-19 ha demostrado que es necesario monitorear la evolución del comportamiento del consumidor digital y los factores que influyen en él, ya que son elementos dinámicos que evolucionan en el tiempo. Este artículo analiza las desigualdades digitales y valida los principales factores que influyen en los turistas para reservar servicios turísticos en línea. Esta investigación utiliza un conjunto de microdatos con 69.752 y 23.779 observaciones para analizar el modo de reserva de los servicios de alojamiento y transporte, respectivamente, obtenidos de la Encuesta de Turismo de Residentes del Instituto Nacional de Estadística durante el periodo 2016-2021. El artículo confirma variaciones en el perfil del consumidor online y en las características del viaje. Uno de los hallazgos más relevantes es la reducción de la brecha generacional en la contratación online de servicios turísticos. Sin embargo, subsisten desigualdades digitales, como las desigualdades regionales y otras basadas en el nivel de estudios y los ingresos de los turistas. También se destaca que diferentes tipos de viajes, dependiendo del destino, el tipo de alojamiento o transporte, tienen una propensión diferente a reservarse a través de canales de compra digitales. La accesibilidad a las fuentes de big data y los avances recientes en los modelos de aprendizaje automático también han hecho evolucionar las metodologías para analizar el comportamiento del consumidor digital y deben incorporarse a los estudios de turismo. Este estudio compara el rendimiento predictivo de diferentes metodologías en el contexto del turismo electrónico. En particular, evaluamos la potencial capacidad predictiva que podría obtenerse usando técnicas de aprendizaje automático para explicar el comportamiento del consumidor en e-Tourism y lo usamos como punto de referencia para compararlo con los resultados obtenidos usando métodos estadísticos tradicionales. Las métricas de evaluación predictivas seleccionadas muestran que el modelo estadístico de regresión logística mejora la capacidad predictiva de la red neuronal Multilayer Perceptron y presenta valores muy cercanos la máxima capacidad predictiva alcanzada por el algoritmo Random Forest.

Palabras clave: turismo electrónico, reservas en línea, comportamiento del consumidor, big data, regresión logística, aprendizaje automático, comportamiento dinámico.

I. INTRODUCTION

The tourism industry has gained synergy effects from using Information and Communications Technology (ICT) (Buhalis et al., 2011; Femenia-Serra et al., 2019; Gratzer et al., 2004). This sector was a pioneer in adopting and developing ICT applications in online sales, which boosted confidence among consumers (Garín-Muñoz & Pérez-Amaral, 2011). Currently, the large number of available resources allows consumers to make travel plans and book tourist services through the Internet.

According to the Spanish National Commission of Markets and Competition (CNMC), electronic commerce (e-Commerce) in Spain has experienced continuous growth in recent years, tripling its turnover between the first quarter of 2016 and the fourth quarter of 2021. The tourism sector has led the number of transactions and the turnover of e-Commerce in

Spain until 2019 (CNMC, 2021). However, the COVID-19 pandemic had a very uneven impact on the different economic sectors, mainly affecting the tourism industry (del Rio-Chanona et al., 2020; Li et al., 2021) and reducing its market share (CNMC, 2021).

The importance of e-Commerce in tourism has generated a substantial body of knowledge focused on the theoretical and empirical development of ICT adoption in tourism named e-Tourism. Gratzer et al. (2002) were the first to name e-Tourism the research area in the field of ICT and tourism. Subsequently, Neidhardt and Werthner (2018) have named e-Tourism as the analysis, design, implementation, and application of IT/e-Commerce solutions in the travel and tourism industry, as well as the analysis of the impact of the respective technical/economic processes and market structures.

The analysis of consumer behaviour and tourism demand is a relevant topic in e-Tourism research and has been the subject of numerous studies (Coenders et al., 2016; Gardella et al., 2021; Garín-Muñoz & Pérez-Amaral, 2011; Reverte & Luque, 2021). This may be because identifying and interpreting the factors that make travellers more likely to book tourism services online is of great relevance to policymakers and companies. However, consumer behaviour and the analytical techniques for its study evolve over time, so its continuous analysis is necessary.

Traditionally, consumer behaviour research in e-Tourism has been based on statistical models, such as logit models (e.g., Boto-García et al., 2021; Chen & Hsu, 1999; Inversini & Masiero, 2014; Lyu & Hwang, 2021), but, following Pourfakhimi et al. (2019), e-Tourism researchers have to take a "hike" to alternative academic fields of studying human behaviour. In our "hike", we have observed that accessibility to big data sources and recent advances in machine learning (ML) models have motivated its implementation in consumer behaviour research in other academic fields (Dang & Pham, 2021; Liébana-Cabanillas & Lara-Rubio, 2017; Zhao et al., 2020). One of the advantages of applying ML techniques is the higher predictive capacity that can be achieved (Martín-Baos et al., 2021; Walker & Jiang, 2019). Furthermore, in line with van Nuenen and Scarles (2021), the implementation of the wide range of prediction and classification tasks offered by ML methods may be of significant benefit to tourism studies.

This study uses microdata for the period 2016–2021 from the Residents Travel Survey (RTS) of the National Statistics Institute of Spain (INE) to address the following research questions: (i) What factors and how do they influence the digital consumer behaviour in e-Tourism?; (ii) Have regional and socioeconomic inequalities increased in the online booking of tourism services in the leisure trips among residents in Spain? and (iii) How can consumers' decisions be forecasted more accurately?

The remainder of this paper is organised as follows. The second section provides a brief review of the literature on consumer behaviour in e-Tourism, as well as consumer behaviour modelling and the use of ML algorithms in academic research. The third section presents an exhaustive overview of e-Tourism in Spain and its recent evolution. The fourth section shows the methodology used in this study. The fifth section offers the results obtained. Finally, the article ends with the conclusions, including the corresponding policy recommendations, limitations and future avenues of research.

II. BRIEF LITERATURE REVIEW

The COVID-19 pandemic has been the accelerator of the process of integration of consumers into digital markets (Ali, 2020; Guthrie et al., 2021; Vollero et al., 2021), which started before the outbreak of the pandemic (Purwanto et al., 2021). The measures adopted by governments to contain the spread of the pandemic significantly altered the behavioural pattern of consumers around the world (Mendieta-Aragón, 2023; Pollak et al., 2022). As a consequence, recent literature confirms the shift in demand from traditional markets to digital markets in different economic sectors, such as the book market (Nguyen et al., 2020), the food sector (Chen et al., 2021; Din et al., 2022), including restaurants (Dsouza & Sharma, 2021; Mahmood et al., 2022) and clothing (Ong et al., 2021).

The COVID-19 pandemic has induced many consumers to adopt online shopping channels predominantly (Pejić-Bach, 2021). Although it is expected that most habits will return back to normal after the pandemic, some habits adopted by the consumer will inevitably remain because the consumer has discovered an alternative that is more convenient, flexible and accessible (Sheth, 2020; Timotius & Octavius, 2021).

The prominent scientific area of e-Tourism has penetrated mainstream tourism research and has attracted the attention of numerous scholars (Buhalis & Law, 2008; Desplas & Mao, 2014; Gretzel et al., 2020; Navío-Marco et al., 2018; Ukpabi & Karjaluoto, 2017).

Amaro and Duarte (2013), in their systematic review and meta-analysis of consumers' behaviour toward online travel purchasing, highlight three main categories of factors that encourage consumers to book tourist services online: (1) characteristics of the consumers, (2) perceived characteristics of the Internet as a sales channel and (3) characteristics of the website or products.

Previous studies support that sociodemographic and economic factors, such as gender, age, education, income level, and digital skills, among others, mainly determine the adoption of the Internet as a shopping channel in tourism (Aeknarajindawat, 2019; Amaro & Duarte, 2015; Garín-Muñoz et al., 2020). Some other research examines the effect that certain travel characteristics may have on online tourism consumption. For example, travel destination and type of transportation (Garín-Muñoz & Pérez-Amaral, 2011), length of stay (Coenders et al., 2016) or travel purpose and season of the year in which it took place (Boto-García et al., 2021).

On the other hand, multiple theories and models try to explain consumers' utilisation of new technologies. Ukpabi and Karjaluoto (2017) synthesised the theories, models and frameworks used to investigate consumers' adoption of ICT in tourism services. Most of these theories are based on the Theory of Reasoned Action (TRA; Fishbein and Ajzen, 1975) and its extensions, the Theory of Planned Behavior (TPB; Ajzen, 1985), and the Technology Acceptance Model (TAM; Davis, 1989).

The TAM is the most common theory in e-Tourism research (Miranda & Briley, 2021; Pourfakhimi et al., 2020; Ukpabi & Karjaluoto, 2017). The TAM model predicts the behaviour of adopting new technologies by individuals based on two motivational factors: perceived usefulness (PU) and perceived ease of use (PEOU). PU refers to the subjective evaluation of the benefits obtained from using ICT. PEOU denotes the degree to which the prospective user expects the achievement of their goal to be simple and effortless (Davis, 1989). Both PU and PEOU have a direct and positive impact on consumer decision-making for purchasing online tourism services and can be determined by the characteristics of both the individual and the trip.

The logistic regression model has been the predominant method for modelling consumer behaviour in tourism studies (Boto-García et al., 2021; Inversini & Masiero, 2014; Jun et al., 2010; Mendieta-Aragón, 2022). Some reasons for using these statistical models are the binary nature of the dependent variables and the fact that logistic regression models allow the interpretation of the estimation coefficients. This last reason can be especially useful for identifying the determinants influencing consumer behaviour and making economic policy recommendations.

On the other hand, the improvement of ML algorithms and the continuous increase in the available data sources have attracted the attention of tourism researchers to apply these new techniques (Solano Sánchez et al., 2022). In particular, ML models have mainly been applied in tourism research to forecast tourism demand (Bigné et al., 2019; Hu & Song, 2020; Phillips et al., 2015; Wen et al., 2019; Xu et al., 2016), to reveal latent topics with textual data in bibliometric reviews (Mariani & Baggio, 2022; Rahmanov et al., 2021) or to identify tourist preferences (Alsayat, 2023; Arefieva et al., 2021).

ML models have potentially greater predictive power than conventional statistical models because, by construction, ML models use algorithms that learn from data and automatically fit relationships between variables without needing a formal theoretical framework. Researchers from many disciplines have recently compared the predictive performance of statistical and machine learning models in consumer behaviour studies and concluded that the non-parametric approach of ML models provides more accurate predictive results (Aluri et al., 2019; Chiang et al., 2006; Dang & Pham, 2021; Greene et al., 2017; Liébana-Cabanillas & Lara-Rubio, 2017). These experimental studies show that ML algorithms approximate the best possible predictions. Following Fudenberg et al. (2022), we evaluate the potential predictive power that could be obtained using machine learning techniques to explain consumer behaviour in e-Tourism and use it as a benchmark to compare it with the results obtained using traditional statistical methods.

III. DATABASE: AN OVERVIEW OF E-TOURISM IN SPAIN AND ITS RECENT EVOLUTION

In this study, we use a microdata set from the monthly information provided by the Residents Travel Survey (RTS) of the National Statistics Institute of Spain during the period 2016-2021.

The RTS is a continuous survey that aims to estimate the number of tourist travels made by residents in Spain and its main characteristics (destination, duration, purpose, accommodation, means of transport, expenditure and sociodemographic characteristics of the travellers, among others)[1]. This survey follows the methodology of the European Statistical Office (Eurostat). Therefore, it allows for comparisons between regions and

[1] The RTS defines tourist travel as all displacements to the main destination out of the regular residence area of the person that involve at least one overnight stay out of the mentioned setting, and with a length shorter than one year. In addition, the main reasons for it must be different from an employment in a company or residence in the place visited.

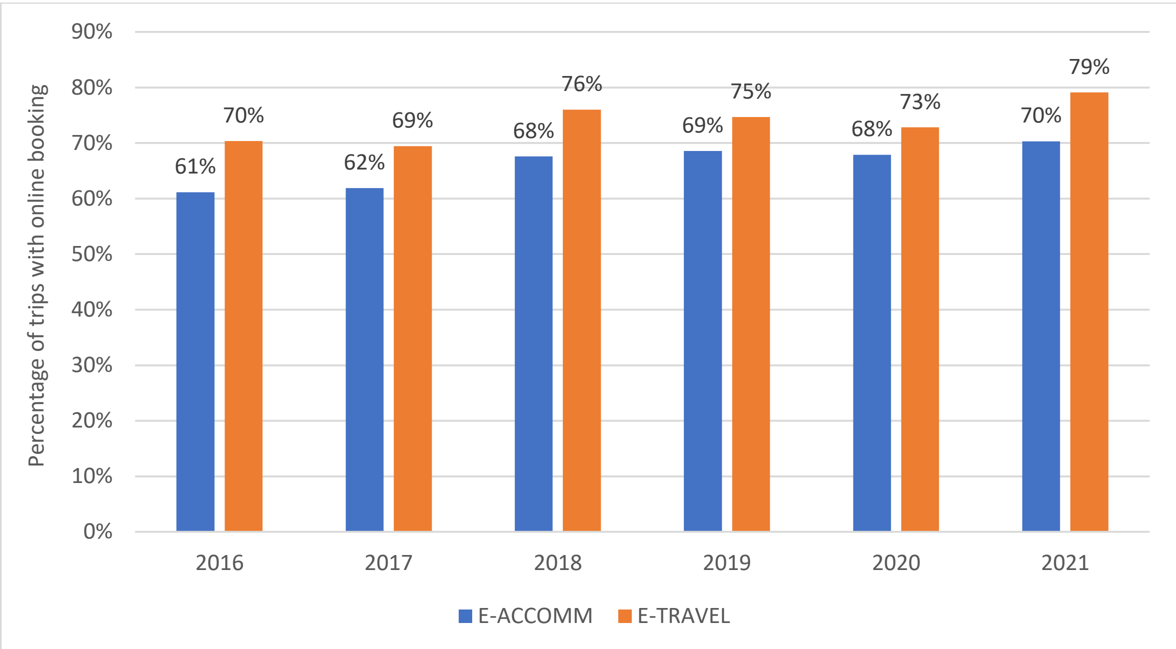
countries. Likewise, the richness of the data allows differentiation between tourist accommodation services (E-ACCOMM) and transportation services (E-TRAVEL). We define online purchases as bookings made via a website or an app. On the contrary, offline purchases are classified as those booked in person, by phone or mail (for more details, see Annex 1).

Although the RTS provides information on the different reasons for the trips, we have focused on leisure travellers who book tourism services. That is, in the case of accommodation, individuals staying in a second residence are not considered. Similarly, in the case of transport, travellers using their private vehicles are not considered.

The sample sizes were 69,752 and 23,779 observations for E-ACCOMM and E-TRAVEL, respectively. The data used in this study were weighted to obtain a representative sample at the population level. The relative weights were extracted from the RTS survey itself and allowed to get estimates that were not affected by a possible sample selection bias[2].

In order to better understand the sample, we proceed to carry out a brief descriptive analysis of the data. Figure 1 shows the evolution of the penetration rates of e-Commerce in the tourism sector based on the percentage of trips booked online by residents in Spain for each of the years 2016–2021. This figure is disaggregated according to tourist accommodation and transport services. The penetration rate of both services shows an upward trend over time, with a significant increase in 2021. Even though both services present penetration rates above 60 per cent, the transportation services have a slightly higher propensity to be purchased online than the accommodation services.

Figure 1. Evolution of penetration rates of e-Tourism in Spain (2016–2021)



Source: Self-elaborated. Data: Residents Travel Survey (INE)

[2] Specifically, we refer to the sample selection bias within the sample of actual travellers, as the weighting of the observations guarantees the representativeness of the actual trips made by residents in Spain.

Table 1 shows the penetration rates of e-Commerce in the tourism sector according to the different groups of travellers and the services purchased. This preliminary analysis suggests the existence of multidimensional digital inequalities.

From this descriptive analysis, we already obtain some preliminary findings relevant to our research. A significant digital divide can be highlighted by age in favour of younger groups. Nevertheless, this generational gap decreased during the period studied. For example, in the case of E-ACCOMM and E-TRAVEL, the digital divide decreased by 8.9% and 7.6%, respectively[3]. Similarly, a wide difference is observed between the group of travellers with higher education (University bachelor or more) and individuals with a primary education level. Even worse, the educational gap increased in the period analysed, especially in transportation services (9.1%).

[3] The digital divide is obtained as the difference between the highest and smallest penetration rate relative to the highest penetration rate for each factor. For example, in accommodation services, the generational gap in 2016 = (69-30)/69 = 56% and generational gap in 2021 = (78-38)/78 =51 %. The variation rate of generational gap for the period 2016–2020 = (51-56)/56 = -8.9%

Table 1. Penetration rate of e-Tourism services through different groups of trips

		E-ACCOMM		E-TRAVEL	
		2016	2021	2016	2021
GENDER	Female	61%	69%	70%	81%
	Male	61%	72%	70%	78%
AGE	[15-25)	64%	77%	63%	78%
	[25-35)	69%	78%	81%	90%
	[35-45)	61%	70%	82%	87%
	[45-55)	57%	69%	73%	82%
	[55-65)	60%	65%	64%	72%
	[65-75)	49%	54%	36%	44%
	+ 75	30%	38%	25%	33%
EDUCATION	Primary or less	43%	49%	37%	34%
	Secondary, first stage	52%	60%	55%	67%
	Secondary, second stage	63%	70%	60%	78%
	University bachelor or more	65%	75%	83%	86%
INCOME	[0–1000€)	59%	65%	45%	60%
	[1000–1500 €)	54%	66%	62%	63%
	[1500–2500 €)	59%	68%	68%	79%
	[2500–3500 €)	62%	73%	72%	84%
	[3500–4999 €)	72%	75%	85%	86%
	+ 5000€	67%	72%	89%	85%
AUTONOMOUS COMMUNITY OF RESIDENCE	Andalucía	54%	69%	62%	81%
	Aragón	63%	68%	74%	76%
	Asturias	56%	67%	66%	47%
	Islas Baleares	74%	72%	76%	90%
	Canarias	75%	79%	77%	89%
	Cantabria	67%	65%	82%	70%
	Castilla y León	55%	62%	54%	83%
	Castilla – La Mancha	57%	70%	52%	85%
	Cataluña	67%	75%	73%	70%
	Com. Valenciana	58%	67%	68%	82%
	Extremadura	56%	72%	67%	56%
	Galicia	57%	60%	61%	63%
	Madrid	64%	73%	80%	87%
	Murcia	55%	66%	66%	88%
	Navarra	57%	72%	65%	77%
	País Vasco	62%	69%	64%	78%
	La Rioja	59%	69%	49%	51%
DESTINATION	National	57%	69%	60%	77%
	International	83%	84%	88%	89%
LENGTH OF STAY	[1–5]	62%	71%	69%	76%
	[6–10]	63%	72%	76%	88%
	[11–15]	50%	63%	68%	79%
	+ 16	42%	43%	65%	69%
TYPE OF ACCOMMODATION	Hotel	67%	75%		
	Apartments	62%	75%		
	Rural	53%	65%		
	Others	35%	44%		
TYPE OF TRANSPORTATION	Air			89%	91%
	Sea			69%	88%
	Car			49%	65%
	Bus			28%	53%
	Train			53%	71%
GROUP SIZE	1	68%	76%	69%	76%
	2	63%	70%	67%	80%
	+ 3	57%	67%	78%	84%
TOTAL		**61%**	**70%**	**70%**	**79%**

Regarding consumer income, a higher penetration rate of e-Commerce is observed in the group of consumers with higher income levels. However, these inequalities have been reduced by 48.8% for the accommodation service and 39.5% for the transportation service.

The penetration rate is not homogeneous across Autonomous Communities and the tourist services acquired. Although there are wide differences depending on the region and the booked service, Madrid, the Balearic Islands, and the Canary Islands have the highest penetration rates for both accommodation and transport services. These results reveal the interregional differences in the adoption of e-Tourism in Spain, as well as the effect of the geographical characteristics of the insular regions, which force individuals to mainly use the transport services of airlines, which in some cases, they only offer their services through the electronic market.

The sample analysis also suggests that online reservations are much more intensive for international destinations than national destinations, but online reservations for national destinations have increased considerably. For example, the penetration rate for accommodation in national destinations increased from 57% in 2016 to 69% in 2021, and for transportation, this rate increased from 60% in 2016 to 77% in 2021.

The 2021 data also reveal differences depending on the type of service booked. Regarding accommodation, the highest penetration rate was for tourist apartments and hotels (75%), compared to rural housing (65%). With an average increase of 10.5 percentage points, this descriptive analysis demonstrates the relevant digitisation of the accommodation sector, which has experienced considerable growth in online reservations. Similarly, for the type of transportation, the online booking of airline tickets stands out (91%), compared to purchasing bus tickets (53%). The reason could be that airlines mostly use online sales channels, while bus transport companies mostly use offline sales channels. Therefore, these inequalities could come from the supply side rather than the demand side.

Summarising, the sample shows the unequal penetration rates of e-Commerce in the main tourist services and their recent evolution in the trips made by residents in Spain. According to these findings, it is timely and appropriate to model the digital behaviour of tourists considering the traveller's sociodemographic characteristics and the trip's characteristics and to analyse which factors interfere to a greater extent in the consumer decision process.

IV. METHODOLOGY

This study aims to identify the main determinants of tourists' use of online shopping channels. To do this, we use the statistical logistic regression model, widely used in the tourist behaviour literature. In addition, in order to explore the advantages of machine learning models in tourism research, this study evaluates the predictive capacity of the logistic regression with respect to the potential predictive capacity of ML methods. In particular, it compares the predictive capacity of the logistic regression model, and two machine learning models, one based on artificial neural networks (ANNs) and the other on decision trees (DTs).

Logistic regression models are mostly used for modelling, data analysis and inference tools. They allow us to understand the role of the input variables in explaining the output variable (Hastie et al., 2009). Logit models are usually fit by maximum likelihood and are parametric models because they assume that the relationship of variables takes a logistic functional form. On the contrary, ML models are mainly used for predictions. ML models are non-parametric, so they do not explicitly assume a parametric form for the function of variables. Instead, they computationally estimate the functional form that is as close to the given data points as possible.

The ANN classifier selected is a Multilayer Perceptron (MLP; Rosenblatt, 1961). MLP is the most common neural network architecture. It consists of a set of multiple hierarchical layers of processing elements (called nodes) that are fully connected from one layer to the next. Within the MLP method, the information flow is processed in a feed-forward manner, that is, the output of one layer is the input of the next layer. The training process is performed by backpropagating the errors and adjusting the network weights correspondingly to decrease the deviations of the outputs from the target values.

A DT is a hierarchical model composed of multiple decisions and leaf nodes. The decision rules are applied recursively to partition the attribute space of a dataset into pure, single-class subspaces or subtrees, where each leaf node represents a class or category of the explained variable (Myles et al., 2004). The DT algorithm selected for this analysis was random forest (RF). The RF is an ensemble learning method proposed by Breiman (2001). The RF constructs a wide range of decision trees with randomly selected observations from the training set. Then, every decision tree is used to make a prediction and output the class of each observation or instance.

The odds ratio (OR) of the explanatory variables using a logistic regression model is employed to measure the influence of each of the determinants of e-Tourism. The OR measures the magnitude of the association between two dichotomous variables. It is calculated as the quotient of the two odds (see Eq. 1.)

$$\text{Odds ratio (OR)}= \frac{\text{Odd}_{k,x_1}}{\text{Odd}_{k,x_0}} = \frac{\frac{P_{k,x_1}}{1-P_{k,x_1}}}{\frac{P_{k,x_0}}{1-P_{k,x_0}}} \qquad (1)$$

where Odd_{k,x_0} is the odds of the reference category. Although it is common to use the first category of each explanatory variable as the reference category to calculate the OR, the choice of the reference category is arbitrary in that, the resulting estimates are equivalent. Categories with estimated values greater than 1, indicate that the odds have increased, therefore, there is a higher probability of occurrence than the reference group. Conversely, categories with ORs below 1, indicate lower odds, hence, there is a lower probability of occurrence than the reference group.

The significance of the independent variables is used in statistical models to identify the variables that explain the dependent variable. In contrast, in ML models, the variable importance is used as a variable selection tool (Ellies-Oury et al., 2019; Zhang et al., 2018). Although they have different natures and methodologies, both measures have the same purpose, namely, to determine the factors that influence an event.

Therefore, we also compare the significance of the variables in the statistical models with the importance of the variables in ML models. We measure the variable importance in the logit model using the chi-square value of the Wald test of the joint significance of the categories of each variable and the variable importance in ML models by comparing the model's predictive accuracy.

This study also uses several evaluation metrics to identify the advantages and limitations of statistical and ML models. Specifically, we compute: Accuracy, which represents the proportion of correct model classifications; the area under the ROC curve (AUC), which indicates the ability of the model to distinguish between classes; Sensitivity and Specificity, which represent, in our case, the ability of the model to correctly classify the online and offline reservations, respectively. Specifically, these measures are defined as follows:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{False Positives} + \text{True Negatives} + \text{False Negatives}} \quad (2)$$

$$\text{AUC} = \int_0^1 \text{True Positive Rate}\left(\text{False Positive Rate}^{-1}(x)\right)dx \quad (3)$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (4)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}} \quad (5)$$

V. RESULTS AND DISCUSSION

5.1. Determining factors of e-Tourism

Based on the theoretical TAM model, the determinants of online purchasing of tourist services for holiday trips are identified. We present two alternative models for the use of the Internet by individuals for booking or purchasing tourist services. One is for accommodation services, and the other for transportation services.

The dependent variables are dichotomous, which take the value 1 if the individual has purchased or booked the considered tourist service online, and 0 if the purchase has been offline. Table 2 presents the independent or explanatory variables used in this study. These variables belong to two broad categories: the sociodemographic and economic characteristics of travellers and the characteristics of the trip. In the present analysis, all explanatory variables were converted into categorical values to calculate the OR estimates.

Table 2. Description of explanatory variables

Variable	Name	Description
Gender	GENDER	1 if male; 0 if female.
Age	AGE	Seven groups according to the age of the traveller.
Education	EDUC	Four groups of the educational level of the tourist.
Income	INCOME	Six groups measured by the monthly net income of households (in euros).
Autonomous Community of residence	CCAA_RES	Regional dummies to control the effects of the Autonomous Community of residence of the tourist.
Destination	DEST	Dummy variable equals to 1 if the final destination is abroad; 0 if it is a national destination.
Length of stay	L_STAY	Four categories measured by the number of overnight stays.
Type of accommodation	ACCOMM	Four categories according to the type of accommodation.
Type of transportation	TRANSP	Five categories according to the type of transportation.
Mode of booking for the accommodation service	RES_ACCOMM	Three groups depending on whether the tourist booked online, offline, or not booked the accommodation service.
Mode of booking for the transportation service	RES_TRANSP	Three groups depending on whether the tourist booked online, offline, or not booked the transportation service.
Group size	G_SIZE	Three categories according to the group size.
Year	YEAR	Temporal dummies to control the effect of time.

Multicollinearity may cause the estimated logistic regression coefficients to be inaccurate, consequently reducing the model's predictive power. Therefore, we use Cramér's V coefficients to evaluate the multicollinearity between explanatory variables (Hernández & Hernández, 2019; Mehmood, 2021). In our case, we removed the destination variable from the E-TRAVEL model because of its high association with the type of transport variable (V= 0.5465). Additionally, we conducted the Wald test and verified that the explanatory variables are jointly statistically significant to explain the dependent variables.

Table 3 shows the ORs estimated for the E-ACCOMM and the E-TRAVEL models using logistic regression in Stata v.17 statistical software. The interpretation of the estimation results is as follows:

Table 3. Estimated odds ratios

		E-ACCOMM		E-TRAVEL	
		Odds ratio		Odds ratio	
GENDER	Male	1.04		1.10	
Ref. category: Female					
AGE	[25–35)	0.89	*	1.00	
Ref. category: [15–25)	[35–45)	0.76	***	0.98	
	[45–55)	0.66	***	0.79	**
	[55–65)	0.56	***	0.60	***
	[65–75)	0.33	***	0.32	***
	+ 75	0.26	***	0.28	***
EDUC	Secondary, first stage	1.10		1.17	
Ref. category:	Secondary, second stage	1.49	***	1.39	*
Primary or less	University bachelor or more	1.58	***	2.05	***
INCOME	[1000–1500 €)	1.05		1.32	**
Ref. category: [0–1000€)	[1500–2500 €)	1.20	***	1.31	**
	[2500–3500 €)	1.24	***	1.55	***
	[3500–4999 €)	1.45	***	1.98	***
	+ 5000€	1.06		2.23	***
CCAA_RES	Aragón	1.24	***	1.93	***
Ref. category: Andalucía	Asturias	0.92		0.90	
	Islas Baleares	1.07		0.99	
	Canarias	1.47	***	0.99	
	Cantabria	1.20	***	0.78	
	Castilla y León	1.06		1.21	
	Castilla – La Mancha	1.22	***	1.09	
	Cataluña	1.41	***	0.82	*
	Com. Valenciana	1.10	*	1.05	
	Extremadura	1.01		1.25	
	Galicia	0.84	***	0.60	***
	Madrid	1.29	***	2.04	***
	Murcia	1.02		0.71	*
	Navarra	1.22	***	1.21	
	País Vasco	1.29	***	1.12	
	La Rioja	1.29	***	1.15	

		E-ACCOMM		E-TRAVEL	
		Odds ratio		Odds ratio	
DEST	International	2.48	***		
Ref. category: National					
L_STAY	[6–10]	0.92	***	1.15	**
Ref. category: [1–5]	[11–15]	0.56	***	1.06	
	+ 16	0.34	***	1.00	
ACCOMM	Apartments	1.00			
Ref. category: Hotel	Rural	0.69	***		
	Others	0.25	***		
TRANSP	Sea			0.51	***
Ref. category: Air	Car			0.22	***
	Bus			0.08	***
	Train			0.19	***
RES-ACCOMM	Online			12.20	***
Ref. category: Offline	No reservation			3.38	***
RES-TRANSP	Online	10.39	***		
Ref. category: Offline	No reservation	3.64	***		
G_SIZE	2	0.91	***	0.98	
Ref. category: 1	+ 3	0.79	***	0.75	***
YEAR	2017	1.03		1.05	
Ref. category: 2016	2018	1.33	***	1.31	***
	2019	1.37	***	1.14	
	2020	1.51	***	1.17	
	2021	1.66	***	1.76	***
constant		0.37	***	0.70	

Note: *** p-value < 0.01; ** p-value < 0.05; * p-value < 0.1

1) Gender: The results show that the gender variable is not statistically significant to explain the reservation mode of any of the considered tourist services (accommodation or travel).
2) Age: As expected, age has a negative impact on E-ACCOMM and E-TRAVEL. The age groups with a higher propensity to acquire tourism services online are individuals aged 15–25. Consumers aged 75 and above have a lower probability of adopting e-Tourism.
3) Education: The traveller's academic study level positively and significantly affects both models, especially the transport service.
4) Income: The income level also positively and significantly affects the use of the Internet as a reservation mode for tourist services. Consumers with a higher household income tend to engage in e-Tourism with a higher likelihood than those with lower income levels.
5) Destination: The destination of travel (national or international) is an important determining factor in the online booking of accommodation services, showing a significant and greater propensity for international travel.
6) Length of stay: Long-term trips show a lower probability of booking accommodation services through the Internet. However, the number of overnight stays is not a determining factor in the online booking of transport services.
7) Accommodation type: The results show that there are no differences between online reservations for hotel accommodation services and tourist apartments. However, there is a lower likelihood of booking online trips with accommodation in rural houses.
8) Type of transport: The type of transport is an important variable in E-TRAVEL and shows a high tendency to book airline tickets online. In contrast, bus tickets are less likely to be purchased online.
9) The results show a higher probability of booking the accommodation service online if the transport service is also booked online and vice versa.
10) Group size: As the travel group size increases, the level of Internet usage for purchasing tourist services in leisure trips decreases in both models.

The regional dummies reveal significant differences between tourists from different Autonomous Communities of Spain. We also include time dummy variables in the models, which reflect a growing trend in using the Internet as a shopping channel in tourism. Furthermore, these results confirm that the COVID-19 pandemic has accelerated a transformation in tourist behaviour and consumers' decision-making process towards their participation in digital markets. In the same sense, other works that reinforce the dynamic behaviour of the consumer in the tourism sector are Ghaedi (2022), Liu et al. (2023) and Sharma et al. (2020).

5.2. Prediction of e-tourist behaviour

In this section, the performance of the statistical and machine learning models is measured based on the predictive accuracy of the mode of booking accommodation and transportation services by tourists.

Table 4 shows the ranking of variable importance for both services using logit, RF, and MLP estimation methods. Owing to the multicollinearity problem, the destination factor is not included in the E-TRAVEL model. The ranking of the independent variables for RF is generally consistent with the logit models.

Table 4. Comparison of variable importance in statistical models and machine learning

	E-ACCOMM			E-TRAVEL		
	LOGIT	RF	MLP	LOGIT	RF	MLP
1	RES-TRANSP	RES_TRANSP	GENDER	TRANSP	TRANSP	GENDER
2	DEST	AGE	EDUC	RES-ACCOMM	RES_ACCOMM	AGE
3	AGE	L_STAY	AGE	CCAA_RES	CCAA_RES	EDUC
4	L_STAY	ACCOMM	INCOME	AGE	EDUC	INCOME
5	YEAR	EDUC	RES_TRANSP	EDUC	AGE	CCAA_RES
6	CCAA_RES	CCAA_RES	DEST	INCOME	INCOME	TRANSP
7	ACCOMM	GENDER	G_SIZE	YEAR	G_SIZE	RES_ACCOMM
8	INCOME	G_SIZE	ACCOMM	G_SIZE	L_STAY	G_SIZE
9	G_SIZE	INCOME	YEAR	GENDER	YEAR	L_STAY
10	EDUC	DEST	CCAA_RES	L_STAY	GENDER	YEAR
11	GENDER	YEAR	L_STAY			

Note: In logit models, ranking shows the order of attributes with the highest statistical significance. In ML models, ranking is the order in which the last attributes were removed, starting with all attributes (the variables removed in the first place are those that occupy the last positions). Attribute evaluator method: Wrapper subset evaluator (Kohavi & John, 1997). Search method: Greedy stepwise (backward).

The variables with a higher statistical significance in the logit models coincide with the most important variables in the RF models. The most important and significant variable for accommodation is the mode of booking the transport. Other important and highly significant variables are the age of the traveller and the number of overnight stays. The destination is a very significant variable for the logit model, but not important for the RF model. In the case of transportation, the variables: type of transport, the mode of booking the accommodation, and the Autonomous Community of residence, have the greatest significance for the logit model and the greatest importance for the RF model. The booking mode of the alternative tourism service (transport or accommodation, depending on the model) has a high significance and importance in predicting consumer behaviour. This variable reflects the consumer's digital skills and the habit of using online shopping channels. These findings are in line with Garín-Muñoz et al. (2020) and confirm that digital consumer skills are a determinant factor of e-Tourism adoption. On the other hand, airlines were the pioneers in introducing this type of online sales and have generated great confidence among consumers, favouring the adoption of electronic commerce. In the transport service, the type of transport has a high capacity to explain the individual's behaviour, possibly because, as anticipated by our descriptive analysis, air travel has a high propensity to be booked online.

Based on the previously identified consumer behaviour determinants, we evaluate these factors' predictive capacity to determine the booking mode of accommodation and transportation services using the evaluation metrics indicated in the methodology section. Table 5 presents the results of the procedure. We assess the predictive capacity of the models using 10-fold cross-validation with WEKA 3.8.4. software. K-fold cross-validation is one of the most popular methods for model evaluation that minimises overfitting issues (Vu et al., 2022). K-fold cross-validation divides the data set into *k* smaller subsets, called "folds", then trains the model on k-1 of these folds and evaluates the remaining fold. This process is repeated *k* times, using different combinations of training and test folds, and the evaluation scores are averaged to obtain a more accurate estimate of model performance.

Table 5. The predictive capacity of the logit, RF, and MLP models

	E-ACCOMM			E-TRAVEL		
	LOGIT	RF	MLP	LOGIT	RF	MLP
Accuracy	71.3%	**71.5%**	57.5%	82.5%	**84.5%**	59.4%
AUC	0.71	**0.71**	0.50	0.87	**0.89**	0.50
Sensitivity	**0.94**	0.85	0.70	0.92	**0.93**	0.70
Specificity	0.20	**0.42**	0.30	0.58	**0.64**	0.30

Note: The values in bold indicate the parameters with the best metric for each model.

The logit results show a higher accuracy for transport services (82.5%) than for accommodation services (71.3%). This may be because different tourist services have different levels of acceptance in the electronic market. The higher level of ICT adoption in booking the transport service, especially for some means of transport, such as the purchase of airline tickets, may have increased the predictive capacity of transport models.

Regarding comparing the three approaches, the results show that RF outperforms logit and MLP algorithms in terms of all evaluation metrics, except for the sensitivity test in E-ACCOMM. All the models have a high sensitivity rate, hence, they are suitable for predicting online booking cases. However, the specificity is significantly higher in RF algorithms. Therefore, managers concerned with predicting the person who books tourism services offline should mainly use RF. According to our results, MLP is not an adequate model to predict consumer behaviour in e-Tourism because it shows the worst measures in all cases.

Taking the predictive capacity of the Random Forest model as a reference, since it is the one with the highest potential predictive performance, the logistic regression model shows a very similar predictive capacity, indicating the good performance of the logit model, both for predicting tourist behaviour and for modelling determining factors and quantify the impact of each of them.

VI. CONCLUSIONS

This study contributes to the literature on electronic tourism by examining the factors that influence tourists to use the Internet as a purchase channel. The results show inequalities in their behaviours depending on the socioeconomic characteristics of the individual and the characteristics of the trip. This study also evaluates the benefits of using machine learning models in tourism behaviour studies for predictive purposes.

It is observed that the factors have varied in intensity and influence, demonstrating the dynamic nature of digital consumer behaviour and that the profiles of the e-tourist evolve over time, which justifies a work like this.

In line with Garín-Muñoz et al. (2020), our analysis supports the existence of a multidimensional digital divide according to age, educational level, and income level. In addition, we find differences between the regions of Spain regarding the use of e-Tourism. Our findings show the narrowing of the generational digital divide in both the booking of accommodation and transport services. However, the inequalities in the reservation mode of tourist services generated between individuals due to educational level have widened, especially in transport services. The analysis also shows different penetration rates depending on the characteristics of the trip. Therefore, we modelled the adoption of e-Tourism by residents in Spain based on the sociodemographic and economic characteristics of the individual and also considering the characteristics of the trip.

The modelling of the behaviour of the e-tourist reveals that age still has a negative effect on the adoption of the Internet as a reservation channel in tourism, and education and income have a significant and positive impact on the probability of booking both tourist services online, especially in transportation services. Regarding travel characteristics, the destination of the trip is an important factor in the online booking of accommodation services. The probability of booking accommodation online increases in international travel compared to national travel. The type of accommodation or transport (depending on the model) is also an important variable. In accommodation services, there are no significant differences in the mode of booking hotel rooms and the rental of tourist apartments. In recent years, there has been a boom in online peer-to-peer platforms such as Airbnb, which specialises in renting tourist apartments and Online Travel Agencies, such as Booking, which have increased online transactions for these types of accommodation. However, in transportation services, there is a clear propensity to purchase airline tickets online. This could be due to the emergence of low-cost airlines, some of which only operate through the web. A relevant finding is that when an individual contracts a tourist service online (accommodation or transport), there is a high probability that the other service will also be purchased online.

One of the main contributions to the existing literature is that this study explores the potential of using ML methods to investigate the complexity of tourist behaviour. The use of ML models to predict consumer behaviour has theoretical and practical implications. On the one hand, predictive models can help researchers and managers to identify patterns and trends in data that might not be detected with other methodological approaches. This can lead to new insights and findings that lead to the development of new theories and concepts about consumer behaviour. On the other hand, from a practical perspective, the capacity to accurately predict the behaviour of tourists can have important economic implications since it can help tourist companies and destinations to optimise their resources, reduce costs and increase revenues. Furthermore, an accurate prediction model can be used to develop targeted marketing strategies and personalised tourism products that meet the needs of specific tourism segments.

We agree with Liébana-Cabanillas and Lara-Rubio (2017) that the application of machine learning predictive models can be compatible with different sectors and markets, such as the tourism sector. Our results reveal the improvement of the predictive capacity of the RF model concerning the logistic regression model. However, in this particular case, this improvement is not significantly higher. Therefore, applying machine learning or statistical models would be appropriate depending on the objectives of academics or managers. In predictive terms, the improvement of machine learning predictive models could be differential for the decision-making of tourism companies. However, in addition to achieving a high predictive capacity, the logistic regression model allows managers and stakeholders to identify the determinants of tourism behaviour and decide on economic policy, acting on those factors whenever possible.

Despite its contributions, this study has some limitations that may open up opportunities for future research. Due to the fact that we use a secondary official data source capable of representing all the trips made by residents in Spain, our explanatory variables are restricted to the information provided by the RTS survey. Therefore, it would be interesting to analyse in future research to include a set of primary data with specific information about consumers, such as the perceived risk of using the Internet as a shopping channel. The predictive analysis has focused on two types of machine learning models based on neural network algorithms and decision trees. However, the application of different algorithms based on artificial intelligence is a clear avenue for new research.

In any case, we consider that our findings represent a little step forward in understanding the digital inequalities generated in e-Tourism, and in the use of machine learning models in studies of tourist behaviour. Our article shows the evolving nature of the digital tourist profiles and the usefulness of the application of new methods for tourism analysis.

BIBLIOGRAPHIC REFERENCES

Aeknarajindawat, N. (2019). The factors influencing tourists' online hotel reservations in Thailand: An empirical study. *International Journal of Innovation, Creativity and Change*, *10*(1), 121–136.

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. In *Action Control* (pp. 11–39). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69746-3_2

Ali, B. J. (2020). Impact of COVID-19 on Consumer Buying Behavior Toward Online Shopping in Iraq. *Economic Studies Journal*, *18*(42), 267–280. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3729323

Alsayat, A. (2023). Customer decision-making analysis based on big social data using machine learning: a case study of hotels in Mecca. *Neural Computing and Applications*, *35*(6), 4701–4722. https://doi.org/10.1007/S00521-022-07992-X

Aluri, A., Price, B. S., & McIntyre, N. H. (2019). Using Machine Learning To Cocreate Value Through Dynamic Customer Engagement In A Brand Loyalty Program. *Journal of Hospitality & Tourism Research*, *43*(1), 78–100. https://doi.org/10.1177/1096348017753521

Amaro, S., & Duarte, P. (2013). Online travel purchasing: A literature review. *Journal of Travel and Tourism Marketing*, *30*(8), 755–785. https://doi.org/10.1080/10548408.2013.835227

Amaro, S., & Duarte, P. (2015). An integrative model of consumers' intentions to purchase travel online. *Tourism Management*, *46*, 64–79. https://doi.org/10.1016/j.tourman.2014.06.006

Arefieva, V., Egger, R., & Yu, J. (2021). A machine learning approach to cluster destination image on Instagram. *Tourism Management*, *85*, 104318. https://doi.org/10.1016/j.tourman.2021.104318

Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on Twitter: A case study of short breaks in Spanish tourist cities. *Tourism Management*, *71*, 490–503. https://doi.org/10.1016/j.tourman.2018.10.013

Boto-García, D., Zapico, E., Escalonilla, M., & Baños-Pino, J. F. (2021). Tourists' preferences for hotel booking. *International Journal of Hospitality Management*, *92*, 102726. https://doi.org/10.1016/j.ijhm.2020.102726

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet-The state of eTourism research. *Tourism Management*, *29*(4), 609–623. https://doi.org/10.1016/j.tourman.2008.01.005

Buhalis, D., Leung, D., & Law, R. (2011). eTourism: Critical information and communication technologies for tourism destinations. *Destination Marketing and Management: Theories and Applications*, 205–224. https://doi.org/10.1079/9781845937621.0205

Chen, J. S., & Hsu, C. H. C. (1999). The Use of Logit Analysis to Enhance Market Segmentation Methodology. *Journal of Hospitality & Tourism Research*, *23*(3), 268–283. https://doi.org/10.1177/109634809902300303

Chen, J., Zhang, Y., Zhu, S., & Liu, L. (2021). Does covid-19 affect the behavior of buying fresh food? Evidence from Wuhan, China. *International Journal of Environmental Research and Public Health*, *18*(9), 4469. https://doi.org/10.3390/ijerph18094469

Chiang, W. Y. K., Zhang, D., & Zhou, L. (2006). Predicting and explaining patronage behavior toward web and traditional stores using neural networks: A comparative analysis with logistic regression. *Decision Support Systems*, *41*(2), 514–531. https://doi.org/10.1016/j.dss.2004.08.016

CNMC. (2021). *e-Commerce statistics in Spain*. http://data.cnmc.es/datagraph/

Coenders, G., Ferrer-Rosell, B., & Martínez-García, E. (2016). Trip Characteristics and Dimensions of Internet Use for Transportation, Accommodation, and Activities Undertaken at Destination. *Journal of Hospitality Marketing and Management*, *25*(4), 498–511. https://doi.org/10.1080/19368623.2015.1034827

Dang, H. D., & Pham, T. T. (2021). Predicting Contract Participation in the Mekong Delta, Vietnam: A Comparison between the Artificial Neural Network and the Multinomial Logit Model. *Journal of Agricultural and Food Industrial Organization*, *20*(2), 135–147. https://doi.org/10.1515/jafio-2020-0023

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*, *13*(3), 319–339. https://doi.org/10.2307/249008 del Rio-Chanona, M. R., Mealy, P., Pichler, A., Lafond, F., & Doyne Farmer, J. (2020). Supply and demand shocks in the COVID-19 pandemic: an industry and occupation perspective. *Oxford Review of Economic Policy*, *36*(Supplement_1), S94–S137. https://doi.org/10.1093/oxrep/graa033

Desplas, N., & Mao, M. (2014). Parallel analysis between e-tourism and e-government: evolution and trends. *Investigaciones Turísticas*, *7*, 1–22.

Din, A. U., Han, H., Ariza-Montes, A., Vega-Muñoz, A., Raposo, A., & Mohapatra, S. (2022). The Impact of COVID-19 on the Food Supply Chain and the Role of E-Commerce for Food Purchasing. *Sustainability*, *14*(5). https://doi.org/10.3390/SU14053074

Dsouza, D., & Sharma, D. (2021). Online food delivery portals during COVID-19 times: An analysis of changing consumer behavior and expectations. *International Journal of Innovation Science*, *13*(2), 218–232. https://doi.org/10.1108/IJIS-10-2020-0184

Ellies-Oury, M. P., Chavent, M., Conanec, A., Bonnet, M., Picard, B., & Saracco, J. (2019). Statistical model choice including variable selection based on variable importance: A relevant way for biomarkers selection to predict meat tenderness. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-46202-y

Femenia-Serra, F., Perles-Ribes, J. F., & Ivars-Baidal, J. A. (2019). Smart destinations and tech-savvy millennial tourists: hype versus reality. *Tourism Review*, *74*(1), 63–81. https://doi.org/10.1108/TR-02-2018-0018

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley.

Fudenberg, D., Kleinberg, J., Liang, A., & Mullainathan, S. (2022). Measuring the Completeness of Economic Models. *Journal of Political Economy*, *130*(4), 956–990. https://doi.org/10.1086/718371

Gardella, N. B., Díaz, A. P. O., & Hernández, J. C. R. (2021). Tourist preferences in hotels and destinations: An approach based on content analysis and decision trees. *Investigaciones Turísticas*, *22*, 121–147. https://doi.org/10.14198/INTURI2021.22.6

Garín-Muñoz, T., & Pérez-Amaral, T. (2011). Internet usage for travel and tourism. The case of Spain. *Tourism Economics*, *17*(5), 1071–1085. https://doi.org/10.5367/te.2011.0080

Garín-Muñoz, T., Pérez-Amaral, T., & López, R. (2020). Consumer engagement in e-Tourism: Micro-panel data models for the case of Spain. *Tourism Economics*, *26*(6), 853–872. https://doi.org/10.1177/1354816619852880

Ghaedi, M. (2022). Tourists' Booking Behavior: Online Social Media Perspectives. *Journal of Promotion Management*, *28*(6), 702–728. https://doi.org/10.1080/10496491.2021.2015509

Gratzer, M., Werthner, H., & Winiwarter, W. (2004). Electronic business in tourism. *International Journal of Electronic Business*, *2*(5), 450. https://doi.org/10.1504/IJEB.2004.005878

Gratzer, M., Winiwarter, W., & Werthner, H. (2002). State of the Art in eTourism. *3rd South Eastern European Conference on E-Commerce 2002*.

Greene, M. N., Morgan, P. H., & Foxall, G. R. (2017). Neural Networks and Consumer Behavior: Neural Models, Logistic Regression, and the Behavioral Perspective Model. *The Behavior Analyst*, *40*(2), 393. https://doi.org/10.1007/S40614-017-0105-x

Gretzel, U., Fuchs, M., Baggio, R., Hoepken, W., Law, R., Neidhardt, J., Pesonen, J., Zanker, M., & Xiang, Z. (2020). e-Tourism beyond COVID-19: a call for transformative research. *Information Technology and Tourism*, *22*(2), 187–203. https://doi.org/10.1007/s40558-020-00181-3

Guthrie, C., Fosso-Wamba, S., & Arnaud, J. B. (2021). Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown. *Journal of Retailing and Consumer Services*, *61*, 102570. https://doi.org/10.1016/j.jretconser.2021.102570

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hernández, A. M. M., & Hernández, C. F. (2019). Access of women to economic autonomy through rural tourism on the island of la Palma. *Investigaciones Turísticas*, *18*, 22–41. https://doi.org/10.14198/INTURI2019.18.02

Hu, M., & Song, H. (2020). Data source combination for tourism demand forecasting. *Tourism Economics*, *26*(7), 1248–1265. https://doi.org/10.1177/1354816619872592

Inversini, A., & Masiero, L. (2014). Selling rooms online: The use of social media and online travel agents. *International Journal of Contemporary Hospitality Management*, *26*(2), 272–292. https://doi.org/10.1108/IJCHM-03-2013-0140

Jun, S. H., Vogt, C. A., & Mackay, K. J. (2010). Online Information Search Strategies: A Focus On Flights and Accommodations. *Journal of Travel & Tourism Marketing*, *27*(6), 579–595. https://doi.org/10.1080/10548408.2010.507151

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Li, Z. F., Zhou, Q., Chen, M., & Liu, Q. (2021). The impact of COVID-19 on industry-related characteristics and risk contagion. *Finance Research Letters*, *39*, 101931. https://doi.org/10.1016/j.frl.2021.101931

Liébana-Cabanillas, F., & Lara-Rubio, J. (2017). Predictive and explanatory modeling regarding adoption of mobile payment systems. *Technological Forecasting and Social Change*, *120*, 32–40. https://doi.org/10.1016/j.techfore.2017.04.002

Liu, Y., Li, Q., Edu, T., & Negricea, I. C. (2023). Exploring the Continuance Usage Intention of Travel Applications In the Case of Chinese Tourists. *Journal of Hospitality & Tourism Research*, *47*(1), 6–32. https://doi.org/10.1177/1096348020962553

Lyu, S. O., & Hwang, J. (2021). A Discrete Choice Experimental Approach to Understand Sports Event Tourists' In-Stadium Beer Consumption Preferences. *Journal of Hospitality & Tourism Research*, *45*(7), 1324–1345. https://doi.org/10.1177/1096348021992099

Mahmood, H., Rehman, A. U., Sabir, I., Rauf, A., Afthanorhan, A., & Nawal, A. (2022). Restaurant Diners' Switching Behavior During the COVID-19 Pandemic: Protection Motivation Theory. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/FPSYG.2022.833627

Mariani, M. M., & Baggio, R. (2022). Big data and analytics in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management*, *34*(1), 231–278. https://doi.org/10.1108/ijchm-03-2021-0301

Martín-Baos, J. Á., García-Ródenas, R., & Rodriguez-Benitez, L. (2021). Revisiting kernel logistic regression under the random utility models perspective. An interpretable machine-learning approach. *Transportation Letters*, *13*(3), 151–162. https://doi.org/10.1080/19427867.2020.1861504

Mehmood, T. (2021). Regularized Feature Selection in Categorical PLS for Multicollinear Data. *Mathematical Problems in Engineering*, *2021*, 1–8. https://doi.org/10.1155/2021/5561752

Mendieta-Aragón, A. (2022). Cambios en el comportamiento turístico tras la COVID-19: hacia un nuevo perfil del turista y del viaje de ocio en España. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, *23*(1), 37–51. https://doi.org/10.24309/recta.2022.23.1.03

Mendieta-Aragón, A. (2023). Efectos de la COVID-19 en el sector turístico de la Unión Europea: Análisis económico para el caso de España. *Revista Universitaria Europea*, *38*(Enero-Junio), 41–70.

Miranda, D. D., & Briley, D. (2021). Digital tourist: Variables that define their purchasing behaviour. *Investigaciones Turísticas*, *21*, 1–21. https://doi.org/10.14198/INTURI2021.21.1

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics*, *18*(6), 275–285. https://doi.org/10.1002/cem.873

Navío-Marco, J., Ruiz-Gómez, L. M., & Sevilla-Sevilla, C. (2018). Progress in information technology and tourism management: 30 years on and 20 years after the internet - Revisiting Buhalis & Law's landmark study about eTourism. *Tourism Management*, *69*, 460–470. https://doi.org/10.1016/j.tourman.2018.06.002

Neidhardt, J., & Werthner, H. (2018). IT and tourism: still a hot topic, but do not forget IT. *Information Technology and Tourism*, *20*(1–4). https://doi.org/10.1007/s40558-018-0115-x

Nguyen, H. V., Tran, H. X., Van Huy, L., Nguyen, X. N., Do, M. T., & Nguyen, N. (2020). Online Book Shopping in Vietnam: The Impact of the COVID-19 Pandemic Situation. *Publishing Research Quarterly*, *36*(3), 437. https://doi.org/10.1007/S12109-020-09732-2

Ong, A. K. S., Cleofas, M. A., Prasetyo, Y. T., Chuenyindee, T., Young, M. N., Diaz, J. F. T., Nadlifatin, R., & Redi, A. A. N. P. (2021). Consumer Behavior in Clothing Industry and Its Relationship with Open Innovation Dynamics during the COVID-19 Pandemic. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(4), 211. https://doi.org/10.3390/joitmc7040211

Pejić-Bach, M. (2021). Editorial: Electronic Commerce in the Time of Covid-19 - Perspectives and Challenges. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(1), i–i. https://doi.org/10.4067/S0718-18762021000100101

Phillips, P., Zigan, K., Santos Silva, M. M., & Schegg, R. (2015). The interactive effects of online reviews on the determinants of Swiss hotel performance: A neural network analysis. *Tourism Management*, *50*, 130–141. https://doi.org/10.1016/j.tourman.2015.01.028

Pollak, F., Markovic, P., Vachal, J., & Vavrek, R. (2022). Analysis of E-Consumer Behavior During the COVID-19 Pandemic. In *Intelligent Processing Practices and Tools for E-Commerce Data, Information, and Knowledge* (pp. 95–114). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-78303-7_6

Pourfakhimi, S., Duncan, T., & Coetzee, W. (2019). A critique of the progress of eTourism technology acceptance research: time for a hike? *Journal of Hospitality and Tourism Technology*, *10*(4), 689–746. https://doi.org/10.1108/JHTT-08-2018-0077

Pourfakhimi, S., Duncan, T., Ould, L., Allan, K., & Coetzee, W. (2020). Acceptance and Adoption of eTourism Technologies. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-Tourism* (pp. 1–31). https://doi.org/10.1007/978-3-030-05324-6_58-1

Purwanto, H., Sidanti, H., & Kadi, D. C. A. (2021). Traditional Market Transformation Into Digital Market (Indonesian Traditional Market Research Library). *International Journal of Science, Technology & Management*, *2*(6), 1980–1988. https://doi.org/10.46729/ijstm.v2i6.384

Rahmanov, F., Mursalov, M., & Rosokhata, A. (2021). Consumer behavior in digital era: impact of COVID 19. *Marketing and Management of Innovations*, *5*(2), 243–251. https://doi.org/10.21272/mmi.2021.2-20

Reverte, F. G., & Luque, P. D. (2021). Digital Divide in E-Tourism. In Z. Xiang, M. Fuchs, U. Gretzel, & W. Höpken (Eds.), *Handbook of e-Tourism* (pp. 1–21). Springer, Cham. https://doi.org/10.1007/978-3-030-05324-6_109-1

Rosenblatt, F. (1961). *Principles of neurodynamics. Perceptrons and the theory of brain mechanisms*. https://apps.dtic.mil/sti/citations/AD0256582

Sharma, S., Singh, G., Pratt, S., & Narayan, J. (2020). Exploring consumer behavior to purchase travel online in Fiji and Solomon Islands? An extension of the UTAUT framework. *International Journal of Culture, Tourism, and Hospitality Research*, *15*(2), 227–247. https://doi.org/10.1108/IJCTHR-03-2020-0064

Sheth, J. (2020). Impact of Covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research*, *117*, 280–283. https://doi.org/10.1016/j.jbusres.2020.05.059

Solano Sánchez, M. Á., Nuñez Tabales, J. M., & Caridad y López del Río, L. (2022). Nuevos tipos de alojamiento: apartamentos turísticos y determinantes de valoración en el precio de la estancia. *Investigaciones Turísticas*, *23*, 94–111. https://doi.org/10.14198/INTURI2022.23.5

Timotius, E., & Octavius, G. S. (2021). Global Changing of Consumer Behavior to Retail Distribution due to Pandemic of COVID-19: A Systematic Review. *Journal of Distribution Science*, *19*(11), 69–80. https://doi.org/10.15722/JDS.19.11.202111.69

Ukpabi, D. C., & Karjaluoto, H. (2017). Consumers' acceptance of information and communications technology in tourism: A review. *Telematics and Informatics*, *34*(5), 618–644. https://doi.org/10.1016/j.tele.2016.12.002

van Nuenen, T., & Scarles, C. (2021). Advancements in technology and digital media in tourism. *Tourist Studies*, *21*(1), 119–132. https://doi.org/10.1177/1468797621990410

Vollero, A., Sardanelli, D., & Siano, A. (2021). Exploring the role of the Amazon effect on customer expectations: An analysis of user-generated content in consumer electronics retailing. *Journal of Consumer Behaviour*, 1–12. https://doi.org/10.1002/cb.1969

Vu, H. L., Ng, K. T. W., Richter, A., & An, C. (2022). Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation. *Journal of Environmental Management*, *311*, 114869. https://doi.org/10.1016/j.jenvman.2022.114869

Walker, K. W., & Jiang, Z. (2019). Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: A machine-learning approach. *The Journal of Academic Librarianship*, *45*(3), 203–212. https://doi.org/10.1016/j.acalib.2019.02.013

Wen, L., Liu, C., & Song, H. (2019). Forecasting tourism demand using search query data: A hybrid modelling approach. *Tourism Economics*, *25*(3), 309–329. https://doi.org/10.1177/1354816618768317

Xu, X., Law, R., Chen, W., & Tang, L. (2016). Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs. *CAAI Transactions on Intelligence Technology*, *1*(1), 30–42. https://doi.org/10.1016/j.trit.2016.03.004

Zhang, J., Cui, X., Cai, W., & Shao, X. (2018). A variable importance criterion for variable selection in near-infrared spectral analysis. *Science China Chemistry*, *62*(2), 271–279. https://doi.org/10.1007/S11426-018-9368-9

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society*, *20*, 22–35. https://doi.org/10.1016/j.tbs.2020.02.003

CONTRIBUTIONS OF THE AUTHORS

Author 1: Conceptualization and design of the article. He has written the original draft. He has also developed the section on methodology and the estimation and comparison of statistical and machine learning models.

Author 2: Conceptualization and design of the article. She has contributed to the writing, review, editing and supervision of the introduction, literature review, definition of the theoretical model, the analysis of the overview of e-Tourism in Spain and its recent evolution, the results on the determining factor of e-Tourism and conclusions.

ACKNOWLEDGEMENT

This work was supported by the National University of Distance Education (Spain) under Grant [BICI N.3, October 21, 2019].

The authors would like to thank Professor Mariano Matilla García (UNED), Francisco Femenia Serra and two anonymous reviewers for their helpful comments on earlier versions of this article, which have improved its quality.

ANNEX

Annex 1. Reservation mode of the tourist service according to the RTS questionnaires

Accommodation	Transportation
Online	**Online**
➢ Directly with the establishment through its web or App	➢ Directly with the company through its web or App
➢ In a travel agency or tour operator through its web or App	➢ In a travel agency or tour operator through its web or App
➢ Through a specialised web (for instance, AirBnb, Homeaway, Homelidays, Niumba, Rentalia, Housetrip, Wimdu, Interhome, Friendly Rentals)	➢ Through the web or App
➢ Directly with the owner through its web or App	
➢ In a travel agency or tour operator through its web or App	
Offline	**Offline**
➢ Directly with the establishment in person, by mail or by phone	➢ Directly with the company in person, by mail or by phone
➢ In a travel agency or a tour operator in person, by mail or by phone	➢ In a travel agency or a tour operator in person, by mail or by phone
➢ Directly with the owner in person, by mail or by phone	➢ Directly with the driver in person, by mail or by phone
➢ In a travel agency or a tour operator in person, by mail or by phone	