

Decoding hotel reviewers: Insights from a decision tree analysis

Research Article

Miguel Llorens-Marin^{1*}, Adolfo Hernandez², Maria Puelles-Gallo¹

¹ Department of Marketing, Complutense University of Madrid, Madrid, Spain

² Department of Financial and Actuarial Economics and Statistics, Faculty of Commerce and Tourism, Complutense University of Madrid, Madrid, Spain

Received 19 February 2025; Accepted 24 June 2025

Abstract: With the rise of user-generated content, understanding travelers' propensity to share hotel experiences online has become a critical research area in tourism and hospitality management. This study examines factors influencing travelers' likelihood of writing hotel reviews, with a particular focus on the role of different online platforms and sociodemographic characteristics. While previous research has explored general motivations for online review writing, little is known about how specific platforms influence review contributions. To address this gap, we conducted an online survey with 739 travelers and applied a machine learning (ML) technique – decision trees (DTs) – to classify customers based on their likelihood to write reviews. This approach allowed us to model non-linear relationships while ensuring interpretability for marketing practitioners. Our findings reveal that frequent use of Facebook for hotel searches, combined with being employed, is the strongest predictor of review-writing behavior, highlighting the role of social media engagement. However, limitations such as non-probability sampling and the omission of certain variables may impact generalizability. This study contributes to the literature by examining the link between platform usage and review-writing behavior, an area that has received limited attention. By employing DTs, a transparent ML approach, our findings offer actionable insights for hotel managers, helping them identify and engage potential reviewers more effectively. Understanding how platform engagement influences review behavior can enhance marketing strategies, encourage review generation, and ultimately shape consumer decision-making in the hospitality industry.

Keywords: Hotel reviews • eWOM • Electronic word-of-mouth • Digital marketing • Decision trees • Writing reviews

1. Introduction

This study aims to examine the factors that influence travelers' propensity of sharing their hotel experiences online, studying the use of different online platforms and sociodemographic traits of the customer as independent variables. Online platforms for gathering information about hotels are Online Travel Agencies (OTAs), such as OTA1 (BookingTM), OTA2 (ExpediaTM), OTA3 (HomeawayTM), OTA4 (TrivagoTM), OTA5 (KayakTM), OTA6 (AirBnBTM), OTA7 (HostelworldTM), and OTA8 (MomondoTM); Review Sites (RSs), such as RS1 (TripAdvisorTM) and RS2 (Google

reviewsTM); and Social Media platforms (SMPs), such as SMP1 (FacebookTM), SMP2 (InstagramTM), SMP3 (TwitterTM), and SMP4 (YouTubeTM). The list of the most commonly used OTAs and travel review sites was based on traffic ranking data from Statista (2019).

In the hotel industry, consumers typically engage with offerings initially through websites and the online environment. The lack of information about actual product features and the inability to physically assess them make electronic word of mouth (eWOM) one of the primary channels through which consumers make their booking decisions (Fan et al., 2018), because of its accessibility to a vast number of reviews (Buhalis & Law, 2008). Hennig-Thurau et al. (2004) defined eWOM as any favorable or unfavorable comment made by potential, existing, or past customers regarding a product or organization, accessible to a wide audience of people and institutions through

* E-mail: mllorens@ucm.es

ORCID: Miguel Llorens-Marin 0000-0002-9093-1700; Adolfo Hernandez 0000-0003-1078-2328; Maria Puelles-Gallo 0000-0001-7921-499X

the Internet. According to Fine et al. (2017), consumers' engagement in sharing their hospitality experience via eWOM varied based on their chosen review platform, revealing significant differences among consumers who identified their preferred platforms for reviews as SMP1, RS1, or RS2, and those who remained neutral without showing a preference for any particular option. Ghazi (2017) studied the reasons why travelers leave reviews on RS1; the main reasons for positive reviews are helping hotels and social benefits, while negative ones are expressing negative emotions, cautioning other consumers, and social benefits.

Given the significant impact of positive eWOM on consumer choices (and consequently on the economic outcomes in the hotel industry) and the relatively small percentage of users inclined to share their reviews (Serra-Cantalops et al., 2020), it seems very relevant to identify a customer profile prone to provide eWOM (Gössling et al., 2018).

Using data drawn from a survey of travelers about their behavior in relation to writing eWOM reviews, this study uses one of the white-box machine learning (ML) techniques, the decision tree (DT) method, to look for combinations among the variable values that describe the probability of writing reviews; in particular, they find the combinations that give the highest and lowest values of the intention to write reviews.

This study identifies the most important triggers for writing eWOM reviews and the combination of values that increases the probability of writing reviews. The use of SMP1 while looking for information about hotels appears as the most relevant trigger, and the intensive use of SMP1 combined with the "employed" work status provides the highest probability of writing hotel reviews.

The rest of this article is structured as follows. First, we introduce the state of the art of the provision of eWOM and outline the research questions; then, a section is devoted to the methodology used and data collection via a consumer survey. The next section presents the results of the classification technique and a comparison with an alternative logistic regression, considered as a reasonable benchmark. We present the discussion and conclusions of this work. The last two sections yield theoretical and practical implications and highlight the limitations and possible directions for future research.

2. Literature review

Extant research studies the impact of eWOM on hotels' customer-booking decisions (Fan et al., 2018; Tsao et al., 2015). Online reviews from travelers serve as an essential resource for information in the hospitality industry and

have been studied in terms of their different aspects, such as source credibility, perceived risk, and information usefulness (Gonzalez-Rodriguez et al., 2022). Online platforms have become important sources of information open to the general public, content is shared as online reviews, capturing travelers' experiences during their tours (Elliott, 2020). These reviews provide meaningful insights into travelers' levels of satisfaction or dissatisfaction, their engagement as customers, their sentiments, and the challenges they encounter with hospitality services (Roy et al., 2021; Zhao & Wang, 2019). In contemporary travel, online reviews have assumed a pivotal role in the evaluation of customers' accommodation experiences as well as their sentiment toward hospitality services. This evaluation process can result in behavioral outcomes, including the likelihood of revisiting a hotel, overall satisfaction levels, and future booking intentions (Park et al., 2020; Zhao & Wang, 2019). The current trend is to use sentiment analysis and natural language processing (NLP) to explore the relationship between different aspects of hotel amenities and services and various customer sentiments.

Some authors have studied the influence of positive reviews on consumer trust and attitudes (Gavilan et al., 2018; Song et al., 2022). Le and Ryu (2023) studied the impact of vloggers (video bloggers) negative reviews on booking intention in the hotel industry. Gregoriades et al. (2021) studied ways to measure tourist satisfaction (e.g., sentiment analysis and review scores) and how to optimize the message in the context of a Marketing Campaign. Significant academic attention has been directed toward devising strategies to alleviate the effect of negative online reviews within the hospitality sector. Such insights are instrumental for hospitality professionals in refining their strategic planning and marketing efforts (Chakraborty, 2019; Yousaf & Kim, 2023).

Shahid and Paul (2022) studied guests' experiences in luxury hotels, which were significantly shaped by hedonism, ambiance, escapism, personalization, and convenience. They also found that these enhanced experiences, in turn, were instrumental in encouraging positive eWOM behavior and strengthening consumers' intentions to revisit. Hotels can benefit greatly from positive eWOM, since positive eWOM can contribute to higher booking rates and enhanced customer satisfaction; therefore, it is relevant for hotels to monitor and leverage eWOM to improve their services and integrate it into their marketing strategies (Gössling et al., 2018). Ghazi (2016) discussed how hotel marketers encourage positive online reviews and manage positive and negative eWOM. His main insights were that marketing managers should focus on sponsoring opinion leaders, reminding/rewarding

reviewers, using eWOM campaigns, and using online communities as priority number one.

The personality of the reviewer has also been studied. Filieri *et al.* (2019) observed that highly unfavorable reviews can be perceived as more useful when they are detailed, easy to comprehend, and when the reviewer is either a specialist or discloses their identity. These findings provide actionable insights for hotel managers to better identify potentially valuable reviews. Research on eWOM provision is also wide, going from studying customer eWOM behavior (Alwash *et al.*, 2019; Ismagilova *et al.*, 2021) to studying the factors producing engagement at the time of writing reviews (Naumann *et al.*, 2020). Some social networks such as SMP1 have been studied in the context of eWOM. The study conducted by Chen *et al.* (2013) found that SMP1 users join Fan Pages to receive information about products and that their unique needs and personality traits significantly influence their willingness to engage in eWOM activities on these pages. Specifically, users' unique needs influence their use of the "Like" and "Share" buttons to express their motivations and readiness to share different kinds of content. Bastrygina *et al.* (2024) reviewed the literature on engagement conceptualization in marketing with a special focus on social media influencers (SMIs). Their study delves into the driving motivations for following SMIs and the subsequent outcomes from such behaviors, focusing on SMP2 influencers in hospitality and tourism and how brands can improve the efficiency of optimizing consumer engagement.

Some recent studies relate demographic factors, such as the age of tourists, to their propensity to provide feedback through SMP1 and the content of their comments. Li *et al.* (2023) state that the millennial generation is prone to using social networks to share information, and Zhong *et al.* (2023) delve into the aspects that most influence the perceptions and recommendations of senior audiences when traveling with their family members. Specifically, they found different impacts of service delivery, built environment, social environment, and hospitality amenities. Fine *et al.* (2017) tested the hypothesis that age has a negative relationship with eWOM review behavior engagement, which is supported by their data.

Regarding the influence of personality traits in engaging with eWOM behavior, Hu and Kim (2018) suggest that individuals high in openness are more likely to engage in eWOM as they are eager to share novel experiences and discoveries. They may use online platforms to express their creativity and appreciation for their unique products or services. Yen and Tang (2015) found that extroverted individuals are sociable, outgoing, and seek social interactions, showing a positive correlation between extroversion and

eWOM participation, while other motivations might be related to consumers' demographics, experiences, or pre-existing attitudes. It also reveals that choosing between RSIs and SMP1 is correlated with different sets of motivations. These findings suggest that motivations are not universally equal and that eWOM behaviors may be correlated with different motivations. Ismagilova *et al.* (2021) synthesized findings from existing studies on eWOM by employing meta-analysis, which helps reconcile conflicting findings on factors affecting consumers' intention to engage in eWOM communications. They divided the factors influencing eWOM provision behavior into four categories: personal, social, perceptual, and consumption-based.

This study focuses on the triggers of eWOM writing, not on its consequences. Some studies considered internal (Guo *et al.*, 2017) or external information regarding eWOM content to explore the factors that trigger eWOM behavior (e.g., Yen & Tang, 2019); internal factors refer to the actual eWOM content, while external factors refer to other features, such as demographic information of eWOM's producer (Gregoriades *et al.*, 2021). In the latter aspect, the authors find a research gap in assessing the connection between customers' propensity to use specific online platforms to gather information about hotels and the propensity to write a hotel review. Other external factors related to customer profile, such as age, gender, work status, household situation, education level, and income level, were also tested in the model as independent variables. A conceptual model framework of the variables used in this study is shown in Figure 1.

In order to sum up, while extensive literature has examined the impact of eWOM on consumer decision-making in the hospitality industry – covering topics such as review content, sentiment, reviewer credibility, and consumer response – less attention has been devoted to the antecedents of eWOM creation, particularly the role of specific online platforms in shaping users' willingness to contribute reviews. Existing studies tend to focus either on the consequences of eWOM or on generalized motivational factors, often neglecting how platform-specific behaviors interact with individual sociodemographic characteristics to influence review-writing. Few works address the asymmetrical and non-linear nature of these relationships. This study addresses this gap by examining the influence of platform engagement and user profiles on the propensity to write hotel reviews.

The set of online platforms (OTAs, review sites, social networks, etc.) has been selected using their traffic rank, and the use of propensity DT is proposed in this study, as they allow for the expression of non-linear relationships among variables, including the asymmetric relationship

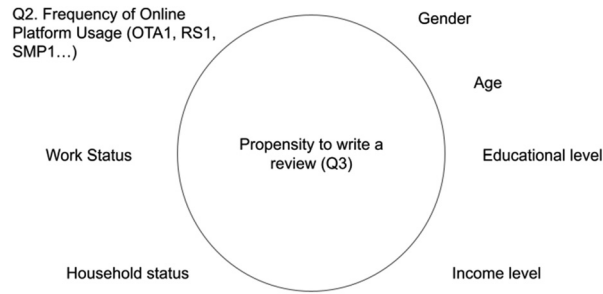


Figure 1. Conceptual model framework.

Source: Authors' contribution.

between the use of online platforms and sociodemographic factors and the propensity to write eWOM. After considering domain knowledge, the research questions (RQ) are as follows:

RQ1: What are the most important triggers for writing eWOM for hotels?

RQ2: What is the best combination of triggers for hotel eWOM writing?

gender, age, work status, household situation, educational level, and net monthly income.

3.2 Description of variables

The variables that were relevant to the research questions are shown in Table 1, indicating the questions and types of answers.

3. Methodology

3.1 Questionnaire design

Before creating the questionnaire, two focus groups were carried out, dividing consumers into two groups according to their age (18–34 and 35–60). The most relevant elements were identified during this qualitative part of the design. We also followed the work of Hyun and Park (2016) and Wang et al. (2012), adjusted to tourism studies, as in Ek-Styvén and Foster (2018), and the work of Wang (2018) on factors that influence travelers' use of eWOM and their ability to generate eWOM content in return.

The questionnaire and scales were validated through a pilot test, ran on 50 consumers varying in age and background, which led to the correction of technical expressions that were not clear for the general population.

Four sections were used to structure the questionnaire, with the first section consisting of an introduction and a question (Q1), to ensure that only consumers who search for tourist accommodations online can be selected. The second section focused on identifying the specific platforms where customers are likely to read reviews (Q2). The third section, the core of the study, examined how customers feel inclined to write reviews after their stay (Q3). Upon completion of the questionnaire, there were six sociodemographic questions that enabled the analysis:

3.3 Sample and data

During December 2022 and January 2023, a convenience sampling method (Malhotra & Birks, 2007) was used to conduct a self-administered survey online that is aimed at users of online tourism opinion platforms. This sampling method is suitable when the study population's limits are unknown or very wide (Goodman, 1961).

Regarding ethical approval, this research is low risk in nature and we have followed best practice as human participants were involved: (1) all participants received detailed written information about the study and its procedures; (2) no health-related data, either directly or indirectly, were collected, and therefore, the Declaration of Helsinki was not specifically referenced when informing participants; (3) data handling was restricted with data under custody by one of the researchers; (4) the anonymity of all collected data was strictly maintained throughout; (5) ethical approval from a board or committee was not obtained, as it was not required under the relevant institutional and national guidelines; (6) informed consent for the data to be used in research was explicitly requested as a preliminary question in the questionnaire before completing the survey.

In Madrid, Spain's capital, a first sample of 200 people ranging from 20 to 75 years old received the questionnaire and they were instructed to pass it on to their friends. The number of valid surveys was reduced to 739 after

Q1: Do you use or have you ever used the Internet (on computers, tablets, cell phones) to search for hotel information?

Yes/No (end of survey)

Q2: Please check the sites or platforms where you usually look for information about hotels, indicating the ones you use the most.

Q2 is subdivided into 14 variables, each one of them specifying a platform/site. Respondents need to evaluate all of them individually in a

4-point Likert scale: 1 never, 2 little, 3 quite a lot, 4 the most

Q2_OTA1; Q2_RS1; Q2_RS2; Q2_OTA3; Q2_OTA2; Q2_OTA4; Q2_OTA5;

Q2_OTA6; Q2_SMP1; Q2_SMP2; Q2_SMP4; Q2_SMP3

Q2_OTA7; Q2_OTA8

Q3: How likely are you to write an opinion after staying in a hotel (approximately how many times do you do it)?

5-point Likert scale: never, rarely, about half of the time, often, and always

Gender: male/female

Age: 18–21, 22–30, 31–45, 46–65, 66–80, over 80

Household status: I live:

Alone; with my partner; with friends; with my partner and children; with my family (parents, siblings, etc.)

Work status:

Not currently employed (studying, unemployed, retired); self-employed; work as a salaried employee in a small company; work as a salaried employee in a large company; manager in a small company; manager in a large company

Educational level:

School graduate; intermediate vocational training; higher vocational training; university degree/graduated; higher studies (Master's degree, doctorate, etc.)

Net monthly income (in euros)

Less than 1,000; 1,001–2,000; 2,001–3,000; 3,001–4,000; 4,001–5,000; 5,001–6,000; over 6,000

Table 1. Description of variables of questionnaire.

Source: Authors' own research.

obtaining a total number of 788 questionnaires with 49 missing responses. With this sample size, the margin of error for the 95% confidence interval was 3.6%.

Among the respondents, the number of males was 40.9%, while the percentage of females was 59.1%. Age was centered around 22–30 (34%) and 31–45 (25%). Both age brackets are over-represented but we consider that they are the more intensive users of social media. A total of 47% of the sample had a bachelor's degree. The most frequent household situation is to live “with my family (parents, siblings, etc.)” with 40% of the respondents.

Regarding the income level, respondents concentrate in the lower ranges (i.e., less than 1,000 and from 1,001 to 2,000), with 79% of respondents.

3.4 Statistical tools

The statistical techniques used in the questionnaire analysis are categorized as classification problems or supervised learning, an ML method. These techniques have been widely utilized since their appearance. More specifically, DTs were used. They were introduced following the original findings of Breiman *et al.* (1984).

When faced with a range of possible decisions, a DT assists in making more accurate decisions from a probability perspective (Hastie *et al.*, 2009), to analyze the results and visually determine how the model flows. In addition, another advantage of DT is that they are considered “white box” techniques (as opposed to “black box”) because the algorithm's logic is not obscured by them and hence provide an interpretable explanation, which increases trustworthiness (Gregoriades *et al.*, 2021). Recent applications of DT in the fields of economics and business include those of Liu and Yang (2022) and Rosado-Cubero *et al.* (2022). Using this technique, we were able to examine the effect of every variable on writing review intentions. DTs are constructed through the use of a recursive partitioning method. Using a feature value, data are separated into distinct groups at each node, resulting in subsets that are subtracted again into smaller subsets.

Let $U = \{X_1, \dots, X_p\}$ be a set of independent variables measured on a set Ω of objects. A DT is a directed acyclic-rooted tree. Both a single variable X_k of U and a subset of objects in Ω are recursively associated with each node k of DT as follows: all objects in Ω can be found in the root node. Let k be a node and S_k be the subset of Ω associated with k . For every different value v_k of attribute X_k , there is a child C_k of k , and the set of objects associated with C_k are the objects of S_k for which the value of attribute X_k is v_k . Merging the categories of each predictor was performed if they did not significantly differ from the dependent variable. A node is a leaf if either the set of objects associated with it contains objects of the same class, according to a specific dependent variable, or if a stopping criterion (i.e., number of levels of the tree) is met. We chose the division model chi-squared automatic interaction detection (CHAID), Kass (1980). CHAID selects the variable that has the most significant relationship with the dependent variable at each step and then uses the chi-square independence test to establish the splitting rule for every node. Given a dependent variable Y with J categories, for each

independent variable X_k with N_k categories, calculation of the Pearson chi-square statistic is made:

$$X^2 = \sum_{j=1}^J \sum_{i=1}^{N_k} \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}},$$

where n_{ij} is the observed cell frequency and m_{ij} is the expected cell frequency for cell ($x_k = i, y = j$) from the independence model. The p value was calculated as $p = p(\chi^2 > X^2)$, where χ^2 follows a chi-square distribution with $d = (J - 1)(N_k - 1)$ degrees of freedom.

The graphical representation provided by this method is a significant advantage as it enables us to identify the combination of predictive variables that produce the highest (and also the lowest) values of writing review intention. However, the main limitation is the possibility of overfitting the data.

We evaluated the quality of the predictions using the classification table, which presents the observed versus predicted (by the model) values, with the calculations of the percentages of correct classification overall (*accuracy*), and for each group (*sensitivity and specificity*).

IBM SPSS Statistics for Windows (v 29.0) was used. The significance level of splitting nodes and merging categories was controlled by setting the default significance level to 0.05. The likelihood ratio method was chosen for the calculation of the chi-square statistic. This method is more robust than Pearson's, although it takes a longer time to calculate. This is the preferred method for small samples, as in our case. The maximum number of iterations was fixed to the default value (100). The same decision was made with the minimum change in the expected cell frequencies (default 0.05). Due to the problem's nature, the authors introduced misclassification costs, which allowed the inclusion of information regarding the relative penalty associated with incorrect classification. In our case, the cost ratio was fixed 2 to 1 (e.g., the key objective is to detect those customers who are prone to write reviews). It is advisable to avoid overfitting of the model by pruning the tree: the tree is grown until stopping criteria are met, and then, it is trimmed automatically to the smallest subtree based on the specified maximum difference in risk. The risk value is expressed in standard errors. We have used the default value (1.0). The minimum leaf size has been set to 40.

The validation method used was split-sample validation: 70% training and 30% testing. The model was generated using a training sample and tested using a hold-out sample. The results are displayed for both training and testing samples.

Following the aforementioned discussion, a DT was built to predict the dependent variable Q3: Propensity to write a review: how often do you write a review after your stay in a hotel? The original 5-point Likert scale was recoded into a binary variable (never/rarely, half/often/always) to improve the interpretability of the results, avoiding branch atomization. This helps prevent the tree from generating too many branches that separate similar categories – such as “never” and “rare” – which can introduce noise and make it harder to identify the most relevant variables.

The 14 variables with a 4-point Likert scale corresponding to Q2 (e.g., OTA1, SMP1) were introduced as independent variables, as well as the 6 sociodemographic questions: gender, age, work status, educational level, and net monthly income. Again, a recoding process was performed for some of the variables. Work status was reduced to three categories: not working (student, unemployed, or retired), employed; self-employed, and manager. At the same time, the net monthly income variable was also recoded into three categories: 1,001–2,000, 2,001–3,000, and over 3,000. Figure 1 illustrates the model variables.

4. Results

The final DT is displayed in Figure 2, with depths of two, seven, and four terminal nodes. Only Q2_SMP1, Q2_OTA1, and work status are significant and therefore produce splits into nodes. Only three variables were selected because DTs prioritize variables based on their ability to split the data in ways that reduce classification error. Variables that do not lead to substantial improvements in classification are either placed in deeper, less influential branches or excluded altogether. The final value for the risk estimate is 0.438, with a standard error of 0.034.

It is easy to elaborate the rules for different terminal nodes. If we focus on more extreme situations (nodes 3 and 6), the rules are as follows:

Node 3: IF ((Q2_SMP1 ≤ 2) AND (Q2_OTA1 ≤ 3)) THEN Prediction is Never/rarely.

Node 6: IF ((Q2_SMP1 > 2) AND (Work status > 1)) THEN Prediction is Half/often/always.

That is, if no intensive use is reported on either SMP1 or OTA1, then it is highly likely that the traveler is not going to write a hotel review (probability 0.712 in node 3). On the other hand, the intensive use of SMP1 by employed travelers is the profile giving the highest probability (0.619 in node 6) of writing a review.

Table 2 lists the associated classification tables for both the training and test samples. Above all, it is necessary to

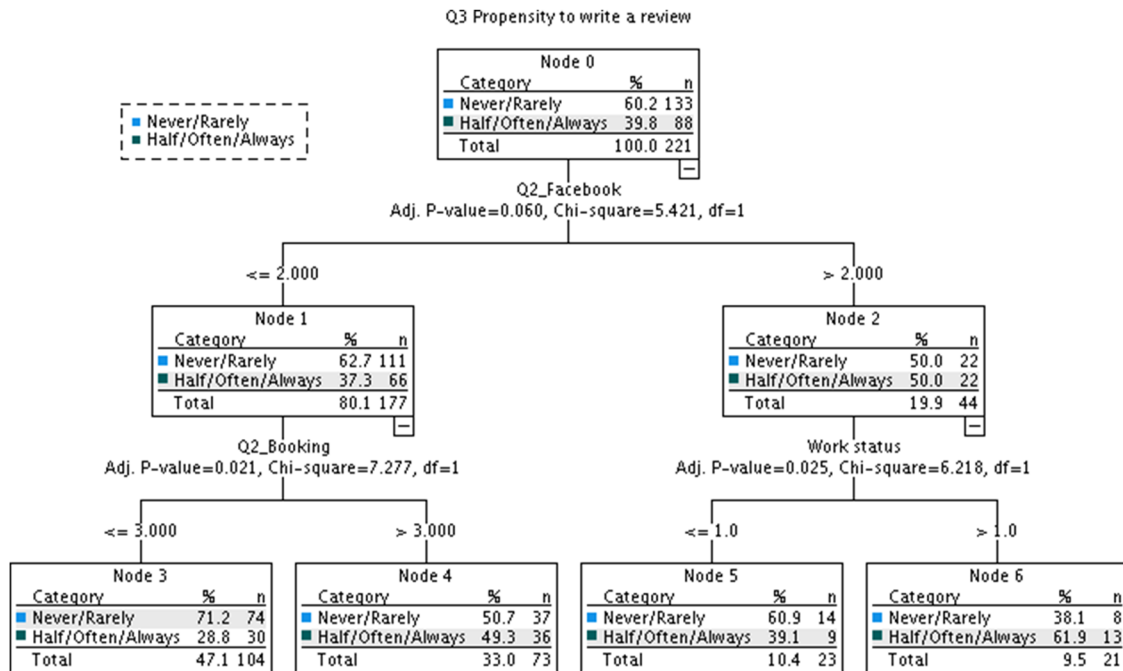


Figure 2. DT (test sample) of triggers for eWOM hotel reviews.

Source: Authors' contribution.

look at the test results because the model is overfitted and the estimates are likely to be biased (optimistic). The overall percentage of correct classifications (accuracy) in the chosen model was 59.7%, with a sensitivity of 65.9% and a specificity of 55.6%. There were 65.9% of

observations correctly classified into the group of travelers showing a propensity to write an online review, which is the group of interest. These results are good given the heterogeneity of individuals.

To validate these results, a comparison was performed against a logistic regression model with forward selection, both a reasonable baseline and a technique that is easy to implement. Table 3 shows the associated classification table, which allows a direct comparison with the proposed DT.

The overall percentage of correct classifications given by the logistic regression (62.9%) improved the corresponding DT percentage (57.4%), but the sensitivity was clearly worse: 27.9 vs 65.6%. Because the key objective of

Sample	Classification			
	Observed	Predicted		Percent
Training		Never/ rarely	Half/often/ always	Correct (%)
Test	Never/rarely	150	151	49.8
	Half/often/ always	75	142	65.4
	Overall			56.4
	percentage			
	Never/rarely	74	59	55.6
Test	Half/often/ always	30	58	65.9
	Overall			59.7
	percentage			

Growing Method: CHAID.

Table 2. Classification table of DT model.

Source: Data processed from IBM SPSS output.

Observed	Classification		
	Never/ rarely	Half/often/ always	Percent correct (%)
Never/rarely	380	54	87.6
Half/often/always	220	85	27.9
Overall percentage			62.9

Table 3. Classification table of logistic regression model.

Source: Data processed from IBM SPSS output.

the technique is to correctly detect eWOM providers), DT shows a better result.

There are other metrics that can be calculated from the classification table. In particular, *F1* score stands out in handling imbalanced datasets, as it is the case. It merges precision (the proportion of true positives among all instances that the model has identified as positive) and sensibility (also known as recall). More specifically, it is the harmonic mean of these two metrics. A higher *F1* score indicates a better balance between precision and recall. The *F1* score of the DT (56.6%) improves the result of the logistic regression (38.3%).

5. Discussion and conclusions

This work proposes a methodology for extracting patterns relating to travelers' propensity to write eWOM hotel reviews by analyzing potential triggers, such as their behavior on online platforms, as well as their sociodemographic characteristics. The authors developed a DT to identify and predict which hotel customers have the highest probability of writing eWOM.

The proposed model achieved reasonably good results (65.9%) when predicting customers' propensity to write online reviews. This allows the definition of a customer profile that is of most interest to hotel managers in the context of eWOM management, as they can actively activate some special treatment for those who are most likely to provide eWOM, and therefore, be of greater value for the hotel. The main findings are that customers who use SMP1 intensively to gather information about hotels and at the same time are employed, have the highest probability to write eWOM reviews. Employed status is recorded as opposed to those who are not working, are self-employed, or are managers. The proposed model also identifies travelers with a lower probability of writing hotel reviews as those with low-intensity use of SMP1 and OTA1 when looking for information about hotels. The interpretation of these results requires caution, since the sampling method is not probabilistic.

Our results differ from those of Fine et al. (2017), who support the influence of age on the probability of writing reviews, as age is not a relevant factor when classifying customers as more likely to write reviews.

Customer qualifications are needed to identify their potential for providing eWOM, which can be easily achieved with a short conversation in the check-in asking what their most used platforms to look up for hotel information.

The ML white-box technique used allows for the interpretation of results and argumentation, as there is a logic

of the algorithm used. To boost their confidence in the results, the authors tested them against a benchmark model: logistic regression. The proposed DT model improves the baseline performance of a logistic regression model in predicting customers belonging to the group of eWOM providers. A comparison with a reasonable benchmark represented by the logistic regression model needs interpretation because, although the logistic regression provides a higher overall percentage of correct classifications (accuracy), it fails to detect the percentage of correct classifications of the group of interest (eWOM providers), with a probability lower than 30%. A higher overall percentage is achieved because the logistic regression model allocates most users to the never/rarely category, which is the group of non-eWOM writers, with the majority in the sample and population.

The originality of the study resides in two features of this work: On the one hand, the use of an ML method, DTs, to solve the raised research questions. Although similar methods have been used in the past (Gregoriades et al., 2021; Vermeer et al., 2019), as far as we know, it is the first time it is used to identify potential review providers. On the other hand, the objective is to identify a combination of variables that determine the propensity of hotel customers to write reviews.

6. Theoretical and practical implications

This work has answered the two proposed research questions, identifying the main triggers for writing eWOM reviews and the best combination of values that increases the customer's propensity to write. This study demonstrates the relevance of both behavioral triggers (such as intensive use of SMP1 for hotel information) and sociodemographic characteristics (specifically employment status) in predicting customers' propensity to write eWOM hotel reviews. The use of a DT model is effective for identifying and predicting customers with the highest probability of writing eWOM reviews. This suggests the utility of advanced analytical techniques such as ML algorithms in uncovering complex patterns and relationships within consumer data for marketing purposes. This study emphasizes the interpretability and validation of results by comparing the DT model with a logistic regression benchmark. Although logistic regression outperforms overall accuracy, the DT model offers insights specifically relevant to the group of eWOM providers. This underscores the importance of selecting appropriate statistical

tools and benchmarking techniques to assess the performance of predictive models in marketing research.

Practical implications are that hotel managers can predict the probability that their customers will become eWOM providers, which is a key contribution to the hotel's online image and marketing strategy. There is a significant value attached to the knowledge that some specific customers are prone to write eWOM reviews and that the way to find out is within the scope of a brief conversation at the check-in stage. The treatment for those customers identified as having a higher propensity to write should be aimed at excellence, being aware that service satisfaction is not sufficient to drive them to write reviews, but is a necessary step. Although the authors do not distinguish between positive and negative reviews, it is clear that online hotels' reputation is at risk due to negative reviews, which could have serious economic consequences (Prayag et al., 2018), while positive reviews stimulate demand. Perez-Aranda et al. (2018) studied hotel managers' responses to eWOM to manage their respective reputations, developing a model with indications regarding applicable procedures. Some authors (Gössling et al., 2018) studied the growing competition faced by hotel managers due to ratings and rankings, while at the same time, guests are becoming more aware of the significance of reviews. Their results, based on a sample of hotel managers, suggest that managers strategically interfere with avoiding the impact of negative online reviews. There are different manipulation strategies used by hotel managers, such as encouraging staff, social media, and customers to write positive reviews; collaborating with platforms; writing reviews themselves; enlisting commercial raters to improve reputation; or posting negative reviews of rivals on open sites like RS1 (Gössling et al., 2018). In our study, we touch on the engagement of customers in the writing of reviews, although in order to solve the managers' ethical dilemma of interfering or not on the sign of the reviews, we do not suggest requesting reviews from customers who are prone to write positive reviews, but any type of reviews. The "manager's dilemma" has been described as the ethical choice that hotel managers face; considering that some competitors employ various manipulation strategies, this can create a scenario where "honest" managers feel compelled to adopt similar tactics to enhance ratings and rankings due to increasing market pressure (Gössling et al., 2019).

Our research provides rich insights to managers about how guests' profiles in their use of different online platforms and sociodemographic factors are relevant, attaching importance to various factors of their information-seeking behavior (use of SMP1) and working status.

Knowledge of the use of platforms provides an early warning signal to hoteliers to improve these aspects.

7. Limitations and future research

A limitation of this study is the use of non-probability sampling, as survey responses were collected through a snowball approach, where initial participants were encouraged to distribute the questionnaire within their own networks. The use of this approach makes results more difficult to extrapolate to the general population; therefore, results should be taken with caution. The authors also realized that some additional variables may have been useful to measure and include in the survey to improve the accuracy of the study. The technique has provided promising results, and further research can expand the number of variables by identifying the channel where the booking was done and the customers' frequency of traveling, as authors suspect that heavy users can have an impact on being more active in their eWOM provision.

Funding information

Faculty of Commerce and Tourism, Complutense University of Madrid.

Author contributions

Miguel Llorens-Marin: Conceptualization; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing – original draft; and Writing – review & editing. Adolfo Hernandez: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing – original draft; and Writing – review & editing. Maria Puelles-Gallo: Resources; Software; Supervision; Visualization; Writing – original draft; and Writing – review & editing.

Conflict of interest statement

Authors state no conflict of interest.

Data availability statement

Research data is available upon request to the corresponding author.

References

- Alwash, M., Savarimuthu, B. T. R., & Parackal, M. (2019). Shallow vs deep customer engagement – a study of brand value propositions in Twitter. *Proceedings of the 27th European Conference on Information Systems (ECIS)*.
- Bastrygina, T., Lim, W. M., Jopp, R., & Weissmann, M. A. (2024). Unraveling the power of social media influencers: Qualitative insights into the role of Instagram influencers in the hospitality and tourism industry. *Journal of Hospitality and Tourism Management*, 58, 214–243. doi: 10.1016/j.jhtm.2024.01.007.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
- 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. doi: 10.1016/j.tourman.2008.01.00.
- Chakraborty, U. (2019). Perceived credibility of online hotel reviews and its impact on hotel booking intentions. *International Journal of Contemporary Hospitality Management*, 31(9), 3465–3483. doi: 10.1108/IJCHM-11-2018-0928.
- Chen, C., Chen, T. H., Chen, Y. H., Chen, C. L., & Yu, S. E. (2013). The spatio-temporal distribution of different types of messages and personality traits affecting the eWOM of Facebook. *Natural Hazards*, 65, 2077–2103. doi: 10.1007/s11069-012-0468-y.
- Ek-Styvén, M., & Foster, T. (2018). Who am I if you can't see me? The “self” of young travelers as driver of eWOM in social media. *Journal of Tourism Futures*, 4(1), 80–92. doi: 10.1108/JTF-12-2017-0057.
- Elliott, C. (2020). *Should you use an online Travel Agency for your next trip?* <https://www.forbes.com/sites/christopherelliott/2020/02/16/should-you-use-an-online-travel-agency-for-your-next-trip/#26bbf8755f22>; (accessed 8 March 2024).
- Fan, A., Shen, H., Wu, L., Mattila, A. S., & Bilgihan, A. (2018). Whom do we trust?. Cultural differences in consumer responses to online recommendations. *International Journal of Contemporary Hospitality Management*, 30(3), 1508–1525. doi: 10.1108/IJCHM-01-2017-0050.
- Filieri, R., Raguseo, E., & Vitari, C. (2019). What moderates the influence of extremely negative ratings? The role of review and reviewer characteristics. *International Journal of Hospitality Management*, 77, 333–341. doi: 10.1016/j.ijhm.2018.07.013.
- Fine, M. B., Gironde, J., & Petrescu, M. (2017). Prosumer motivations for electronic word-of-mouth communication behaviors. *Journal of Hospitality and Tourism Technology*, 8(2), 280–295. doi: 10.1108/JHTT-09-2016-0048.
- Gavilan, D., Avello, M., & Martinez-Navarro, G. (2018). The influence of online ratings and reviews on hotel booking consideration. *Tourism Management*, 66, 53–61. doi: 10.1016/j.tourman.2017.10.018.
- Ghazi, K. (2016). Analyzing hotel practices applied to encourage and manage online guest reviews. *Journal of Association of Arab Universities for Tourism and Hospitality*, 13(3), 115–128. doi: 10.21608/JAAUTH.2016.53970.
- Ghazi, K. M. (2017). Guests' motives to write positive and negative hotel reviews on Trip Advisor. *Journal of Tourism and Hospitality*, 6(3), 1–9. doi: 10.4172/2167-0269.1000283.
- Gonzalez-Rodriguez, M. R., Díaz-Fernández, M. C., Bilgihan, A., Okumus, F., & Shi, F. (2022). The impact of eWOM source credibility on destination visit intention and online involvement: A case of Chinese tourists. *Journal of Hospitality and Tourism Technology*, 13(5), 855–874.
- Goodman, L. A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, 32, 148–170. doi: 10.1214/aoms/1177705148.
- Gössling, S., Hall, C. M., & Andersson, A. C. (2018). The manager's dilemma: A conceptualization of online review manipulation strategies. *Current Issues in Tourism*, 21(5), 484–503. doi: 10.1080/13683500.2015.1127337.
- Gössling, S., Zeiss, H., Hall, C. M., Martin-Rios, C., Ram, Y., & Grøtte, I. P. (2019). A cross-country comparison of accommodation manager perspectives on online review manipulation. *Current Issues in Tourism*, 22(14), 1744–1763. doi: 10.1080/13683500.2018.1455171.
- Gregoriades, A., Pampaka, M., Herodotou, H., & Christodoulou, E. (2021). Supporting digital content marketing and messaging through topic modelling and decision trees. *Expert Systems with Applications*, 184, 115546. doi: 10.1016/j.eswa.2021.115546.
- Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using Latent Dirichlet allocation. *Tourism Management*, 59, 467–483. doi: 10.1016/j.tourman.2016.09.009.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (2nd ed.). Springer.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-

- opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18, 38–52. doi: 10.1002/dir.10073.
- Hu, Yu., & Kim, H. J. (2018). Positive and negative eWOM motivations and hotel customers' eWOM behavior: Does personality matter? *International Journal of Hospitality Management*, 75, 27–37. doi: 10.1016/j.ijhm.2018.03.00.
- Hyun, S., & Park, S. H. (2016). The antecedents and consequences of travelers' need for uniqueness: An empirical study of restaurant experiences. *Asia Pacific Journal of Tourism Research*, 21(6), 596–623. doi: 10.1080/10941665.2015.1062404.
- Ismagilova, E., Rana, N. P., Slade, E. L., & Dwivedi, Y. K. (2021). A meta-analysis of the factors affecting eWOM providing behaviour. *European Journal of Marketing*, 55(4), 1067–1102. doi: 10.1108/EJM-07-2018-0472.
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119–127. doi: 10.2307/2986296.
- Le, H., & Ryu, S. (2023). The eWOM adoption model in the hospitality industry: The moderating effect of the vlogger's review. *Journal of Hospitality and Tourism Technology*, 14(2), 225–244. doi: 10.1108/jhtt-08-2021-0233.
- Li, Z., Chan, C., Chen, Y. F., Chan, W. W. H., & Im, U. L. (2023). Millennials' hotel restaurant visit intention: An analysis of key online opinion leaders' digital marketing content. *Journal of Quality Assurance in Hospitality & Tourism*, 25(6), 2074–2103. doi: 10.1080/1528008x.2023.2219467.
- Liu, Y., & Yang, S. (2022). Application of decision tree-based classification algorithm on content marketing. *Journal of Mathematics*, 2022, 6469054. doi: 10.1155/2022/6469054.
- Malhotra, N., & Birks, D. (2007). *Marketing research: An applied approach*. Prentice Hall.
- Naumann, K., Bowden, J., & Gabbott, M. (2020). Expanding customer engagement: The role of negative engagement, dual valences and contexts. *European Journal of Marketing*, 54(7), 1469–1499. doi: 10.1108/EJM-07-2017-0464.
- Park, E., Kang, J., Choi, D., & Han, J. (2020). Understanding customers' hotel revisiting behaviour: A sentiment analysis of online feedback reviews. *Current Issues in Tourism*, 23(5), 605–611. doi: 10.1080/13683500.2018.154902Y.
- Perez-Aranda, J., Arán, M. V., & Molinillo, S. (2018). A proposal for measuring hotels' managerial responses to User-Generated-Content Reviews. *Tourism & Management Studies*, 14, 7–16. doi: 10.18089/tms.2018.14SI101.
- Prayag, G., Hall, C. M., & Wood, H. (2018). I feel good! Perceptions and emotional responses of bed & breakfast providers in New Zealand towards trip advisor. *Journal of Hospitality Marketing & Management*, 27(1), 1–20. doi: 10.1080/19368623.2017.1318731.
- Rosado-Cubero, A., Freire-Rubio, T., & Hernandez, A. (2022). Entrepreneurship: What matters most. *Journal of Business Research*, 144, 250–263. doi: 10.1016/j.jbusres.2022.01.087.
- Roy, G., Datta, B., Mukherjee, S., & Basu, R. (2021). Effect of eWOM stimuli and eWOM response on perceived service quality and online recommendation. *Tourism Recreation Research*, 46(4), 457–472. doi: 10.1080/02508281.2020.1809822.
- Serra-Cantalops, A., Cardona, J. R., & Salvi, F. (2020). Antecedents of positive eWOM in hotels. Exploring the relative role of satisfaction, quality and positive emotional experiences. *International Journal of Contemporary Hospitality Management*, 32(11), 3457–3477. doi: 10.1108/IJCHM-02-2020-0113.
- Shahid, S., & Paul, J. (2022). Examining guests' experience in luxury hotels: Evidence from an emerging market. *Journal of Marketing Management*, 38(13–14), 1278–1306. doi: 10.1080/0267257X.2022.2085768.
- Song, M., Law, R., & Duan, Y. (2022). What are the obstacles in the way to “avoid landmines”? Influence of electronic word-of-mouth dispersion on order decision from the self-construal perspective. *International Journal of Hospitality Management*, 107, 103334. doi: 10.1016/j.ijhm.2022.103334.
- Statista. (2019). *Las plataformas de alojamientos turísticos preferidas en España/The preferred tourist accommodation platforms in Spain*. <https://es.statista.com/grafico/19459/paginas-apps-preferidas-para-reservar-alojamientos-en-espana/>; (accessed March 11, 2024).
- Tsao, W. C., Hsieh, M. T., Shih, L. W., & Lin, T. M. Y. (2015). Compliance with eWOM: The influence of hotel reviews on booking intention from the perspective of consumer conformity. *International Journal of Hospitality Management*, 46, 99–111. doi: 10.1016/j.ijhm.2015.01.008.
- Vermeer, S. A. M., Araujo, T., Bernritter, S. F. & van Noort, G. (2019). Seeing the wood for the trees: How machine learning can help firms in identifying relevant electronic word-of-mouth in social media. *International Journal of Research in Marketing*, 36(3), 492–508. doi: 10.1016/j.ijresmar.2019.01.010.
- Wang, P. (2018). *Understanding electronic word of mouth in tourism in the social media era*. University of Turku. www.utupub.fi/bitstream/handle/10024/145327/AnnalesE34Wang.Pdf. (accessed 11 March 2024).

- Wang, X., Yu, C., & Wei, Y. (2012). Social media peer communication and impacts on purchase intentions: A consumer socialization framework. *Journal of Interactive Marketing*, 26(4), 198–208. doi: 10.1016/j.intmar.2011.11.004.
- Yen, C. L., & Tang, C. H. (2015). Hotel attribute performance, eWOM motivations, and media choice. *International Journal of Hospitality Management*, 46, 79–88. doi: 10.1016/j.ijhm.2015.01.003.
- Yen, C. L., & Tang, C. H. (2019). The effects of hotel attribute performance on electronic word-of-mouth (eWOM) behaviors. *International Journal of Hospitality Management*, 76A, 9–18. doi: 10.1016/j.ijhm.2018.03.006.
- Yousaf, S., & Kim, J. M. (2023). Dark personalities and online reviews: A textual content analysis of review generation, consumption and distribution. *Tourism Management*, 98, 104771. doi: 10.1016/j.tourman.2023.104771.
- Zhao, X. X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111–121. doi: 10.1016/j.ijhm.2018.03.017.
- Zhong, Y. S., Zhang, T., & Ronzoni, G. (2023). Sentimental analysis of Facebook reviews: Does hospitality matter in senior living? *International Journal of Hospitality Management*, 112, 103384. doi: 10.1016/j.ijhm.2022.103384.