Article

Tourism and Hospitality Research
2025, Vol. 0(0) 1–19

DOI: 10.1177/14673584251379305
journals.sagepub.com/home/thr

A new hotel classification model combining guest reviews with official hotel classification systems: Bridging expert and consumer ratings

Ana Messias[1], Gilberto Rosa[2], Pedro Couto[1], Vítor Rodrigues[1], Luís M. Silva[1,3] and Jorge Marques[4,5]

Abstract

This paper explores the potential of combining online guest reviews with hotel classification systems, focusing on the feasibility of incorporating such reviews at the star-classification level. Leveraging machine learning techniques on a database of Portuguese hotels and their corresponding ratings on Booking.com, this study reveals a weak association between official hotel star categories and mean review scores of satisfaction items rated by users during the review process, suggesting a discrepancy between official star-classifications and consumer expectations and experiences. Based on the results, a new classification model is proposed, which integrates a classification system based on Booking.com reviews alongside traditional star categories, aiming to complement hotel star-classifications with a further quality dimension as perceived by customers through online reviews. This model provides travellers with more informative and reliable information, facilitating decision-making in the hotel selection process.

Keywords

Hotel classification systems, star-classification, online guest reviews, consumer ratings, review scores, hybrid classification model

Introduction

Hotel classification systems are widely used in the accommodation sector to provide measurable indicators for consumers and intermediaries (UNWTO, 2015). Using these indicators makes it easier to compare hotels' service levels and equipment standards. Therefore, any information that can help to better understand and compare the characteristics and expected quality of the accommodation experience can be critical (Arzaghi et al., 2023). For marketing purposes, classification systems are particularly useful in promoting the most varied types of tourist accommodation. However, establishing a classification system for tourist accommodation is a complex task due to the diversity of types of accommodation and the cultural, environmental, and economic contexts in which the systems are applied (UNWTO, 2015). Despite hotels' efforts to provide consumers with reliable, comparable, and relevant information, this industry continues to struggle with the problem of asymmetric information. Mainly because hotel classification systems may differ between countries and regions (Cser and Ohuchi, 2008; Rhee and Yang, 2015) and because they may be insufficient to inform about the quality of service and experience that

[1]Department of Mathematics (DMAT), University of Aveiro, Aveiro, Portugal
[2]Department of Medical Sciences (DCM), University of Aveiro, Aveiro, Portugal
[3]Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, Aveiro, Portugal
[4]Research on Economics, Management and Information Technologies (REMIT), Department of Tourism, Heritage and Culture, University Portucalense, Porto, Portugal
[5]Centre of Studies in Geography and Spatial Planning (CEGOT), University of Coimbra, Coimbra, Portugal

Corresponding author:
Jorge Marques, Department of Tourism, Heritage and Culture, University Portucalense, Rua Dr. António Bernardino de Almeida, 541, 4200-072, Porto, Portugal.
Email: jorgem@upt.pt

hotels offer, which is more subjective and does not always correspond to the expected level.

With the increasing prevalence of the Internet, information about guest reviews, ratings, and scores has become increasingly accessible to travellers. Electronic word of mouth (eWOM) has been reducing the asymmetry of information, and the scores provided by travel sites, such as Booking.com, Expedia, TripAdvisor or Google Travel, among others, have contributed to reduce the subjectivity of service quality and the relationship this has with the characteristics of hotels (Li et al., 2017). As travel-related online searches are rising, the hotel's official classifications and guest reviews play complementary roles. Traditionally, official classification systems focus on facilities and level of service, while guest reviews and ratings are based on expectations and quality of experience. Therefore, in this paper, the term "classification system" refers to formal hotel categorisation schemes (e.g., star-classifications assigned by official bodies based on technical criteria), whereas "rating system" is used to describe consumer-based evaluations such as review scores on online travel platforms.

Hensens (2015) anticipated the shift towards more dynamic, consumer-focused classification models, foreshadowing the integration of digital guest feedback with traditional classification systems. While formal classification systems already ensure that required facilities are present and meet predefined standards, guest reviews provide insight into how those facilities are perceived and experienced by customers – offering an additional, experiential layer of quality assessment. Therefore, the growing reliance on online travel-related content has reshaped how quality is perceived and communicated.

Before making an online hotel reservation, consumers visit on average almost 14 different travel-related websites, with about three visits per website, and perform nine travel-related searches on search engines (UNWTO, 2014). Official hotel classifications are often used by consumers as a filtering mechanism in the booking process, with guest reviews used to make a final selection from a narrower group of hotels. More recently, there has been interest in integrating classification processes into the digital and social era, with regions considering the use of online guest reviews in traditional hotel classification methods. According to the UNWTO (2014), there is a consensus among suppliers and consumers about the advantages of integrating guest reviews into hotel classification systems, provided that an appropriate methodology is developed to do so (UNWTO, 2014).

Despite the widespread use of both official hotel star-classifications and online guest reviews, few academic studies have explored how these systems can be meaningfully integrated into a unified classification framework. Most research either compares the two systems (Martin-Fuentes, 2016; Martin-Fuentes et al., 2018) or analyses the sentiment behind reviews (Krey et al., 2024) without proposing operational models for combining institutional and consumer perspectives. This study addresses that gap by developing a hybrid classification model that uses guest review data to complement and enrich formal hotel classification through machine learning techniques. It also examines the feasibility of incorporating guest reviews into large-scale hotel classification systems, specifically at the star-classification level. Thus, the proposed integrated approach aims to add a further consumer-based quality dimension to hotel classification, thereby refining the classification process by complementing the existing expert-led criteria with experiential guest perspectives.

The research was conducted using a database of Portuguese hotels and their corresponding ratings on Booking.com based on the satisfaction items rated by users during the review process after checking out. The satisfaction items refer to Overall, related to the global accommodation experience and overall satisfaction level, Value for money, Cleanliness, Location, Facilities, Comfort, Staff, and Wi-Fi. It is important to note that these satisfaction items are based on guests' subjective evaluations of their experience, as provided on the Booking.com platform. As such, they reflect perceived quality rather than a formal audit of infrastructure. Consequently, this study does not aim to reproduce the full technical criteria used in official star-classification systems (e.g., availability of lifts, pools, or specific surface areas), but rather to model how guests interpret and evaluate their stay.

From the guest's perspective, perceived quality is shaped by several specific factors that contribute to their overall satisfaction. Value for money refers to balancing perceived costs (primarily monetary) with perceived benefits. Customers seek accommodations offering the highest value at the lowest possible price (Gupta and Kim, 2009), irrespective of the accommodation type. Some authors found value for money to be the most critical factor in hotel selection, after the price criteria (Zaman et al., 2016). Cleanliness applies to rooms and other hotel areas, such as restrooms, entrances, parking areas, lobbies, and dining places. It is associated with safety and low health hazards and is one of the primary causes of dissatisfaction during a stay (Lockyer, 2002). Cleanliness and safety have become even more critical in the aftermath of the COVID-19 pandemic, with recent studies confirming that guests increasingly prioritise hygiene protocols and perceived health security when evaluating hotel quality (Pennington-Gray and Lee, 2024; Tiwari and Mishra, 2023; Tiwari and Omar, 2023). Hotel comfort is usually related to sleep quality, which includes factors like a cozy bed, noise level, adequate room temperature, lighting, and scent

(Zaman et al., 2016). Location is related to the proximity to points of interest, transportation convenience, and the surrounding environment. Location may be essential for tourists who want easy access to the sites they plan to visit and the events they plan to attend (Masiero et al., 2019; Yang et al., 2015). Hotel facilities, also frequently referred as amenities, typically refer to supplementary services (e.g., in-room coffee maker or kettle, safe, luggage storage, recreational equipment storage, complimentary parking, etc.). These facilities may be included in the accommodation price or require an additional fee, depending on the hotel and its standard. To stay competitive, hotels strive to provide an increasing number of facilities (Chu and Choi, 2000). Regarding the hotel staff, it is often a crucial aspect of customer satisfaction regarding hotel services (Chu and Choi, 2000; Kim et al., 2020, 2022). Finally, free and reliable Wi-Fi in hotels is now considered an essential part of modern hospitality and is often viewed by guests as the most important technology that should always be available. While most hotels offer complimentary access, issues such as limited coverage or slow connection speeds are common sources of dissatisfaction (Cain et al., 2024).

By leveraging machine learning techniques, this work proposes a hybrid classification model that analyses all items rated by users during the review process on Booking.com and assigns a refined star-classification that accurately reflects the quality and services provided by each hotel. This methodology ensures a comprehensive and objective assessment of the hotels, enhancing the reliability and usefulness of the complementary classifications.

By applying supervised and unsupervised learning techniques to Booking.com review data and comparing outcomes with official star-classifications, this study advances the conceptual understanding of classification as a multidimensional construct – both facility and experience-based (Koutoulas and Vagena, 2023). This aligns with current academic interest in reconciling objective and subjective indicators of quality in service contexts (Nilashi et al., 2022).

Literature review

Hotel classification systems

Classification systems categorise hotels and services by assigning them distinct grading levels. These systems provide comparative information about hotel facilities, such as view, room quality, room service, food, spa, and fitness services, and more recently, on the surrounding area's public services and facilities (Arzaghi et al., 2023). Classifications are attributed based primarily on the types of facilities and services offered, rather than on the subjective quality of service delivery. In systems such as the Automobile Association (AA) in the UK, classifications focus on the presence, scope, and consistency of services rather than their experiential quality. Some hospitality brands also have different classifications for their properties across geographies, sometimes under a different brand name, to target specific customer segments (Claver et al., 2006). This is because classifications indicate not only the facilities provided by the hotel but also the price levels (Nilashi et al., 2022). Potential guests with different needs consider various criteria to make stay-related decisions, and the classification systems may serve as a credible and trustworthy signal of the hotel's services to make that decision easier (Masiero et al., 2015).

Classification systems can also be divided into those that only evaluate objective criteria and those that evaluate both objective and subjective criteria and can be either statutory (or official) or voluntary (UNWTO, 2015). Most statutory or official systems are government or state-owned classification systems and focus mainly on physical attributes and services, relying more on quantitative and technical aspects than service quality. However, the combination of private and public systems is more intended towards guests, their needs, and expectations (Minazzi, 2010). Hence, public authorities must be more guest-oriented and interested in regulating properties to increase international competitiveness (Khan et al., 2022). These distinctions are reflected in the diversity of classification approaches adopted around the world. Although there is no internationally centralised hotel classification system, several prominent national and regional systems have emerged. These are applied across different parts of the world and use symbols such as diamonds, stars, crowns, suns, coffee pots, letters, and even feathers to categorize hotels (Vallen and Vallen, 2017).

In 1900, Michelin Tyres introduced pictorial symbols to point out the facilities of French establishments (Khan et al., 2022), giving rise to what is now one of the most famous travel guides, the Michelin Guide. In 1912, the AA launched the hotel star-classification in the UK, and today, it is the most used grading system in the country, rating and awarding stars to hotels based on quality, facilities, and services (Blomberg-Nygard and Anderson, 2016). AA has worked closely with VisitBritain, VisitEngland, VisitScotland and Wales Tourist Board to implement Common Quality Standards for hotel inspections, ensuring consistent ratings across the UK (AA Hotel and Hospitality Services, 2024). In addition to standard "Black Star" classifications, the AA awards Silver Stars (for hotels exceeding quality expectations) and distinguished Red Stars, which recognise properties that deliver exceptional hospitality and service levels across all star

categories, thereby providing an additional layer of recognition above the traditional star classification (AA Hotel and Hospitality Services, 2024).

In 1958, the oil and gas company Mobil, through their magazine Mobil's Travel Guide (known today as Forbes Travel Guide), rated hotels using a 1-to-5-star system (Arzaghi et al., 2023). In 1962, the International Union of Official Travel Organizations (UNWTO) developed a consensus on using 5-categories of hotel classification (Vine, 1981). In 1976, the American Automobile Association (AAA) started rating hotels and restaurants using the Diamond Grading system, being considered, nowadays, the most extensive classification system as it grades more lodging properties than any other system in the world based on facilities and services offered (Nalley et al., 2019).

More recently, in 2009, a joint initiative led to the founding of the Hotelstars Union, under the patronage of HOTREC Hospitality Europe (The Confederation of National Associations of Hotels, Restaurants, Cafés and Similar Establishments in the European Union and European Economic Area). This platform aimed to harmonise European hotel classification based on a standard criteria catalogue. Although this initiative did not get the adhesion of all HOTREC members, more than 22,000 hotels are classified within the Hotelstars Union. This system has 247 harmonised criteria (mandatory plus optional criteria), uses a 1 to 5-star grading system and demands revision of criteria every 5 to 6 years. It aims to improve transparency for guests and hoteliers, as well as quality control and fair competition (https://www.hotelstars.eu/).

Within the 1-to-5-star grading system, variants have also emerged. The European Hotelstars Union has a higher "Superior" mark to account for some extra features in each star category. Another example is the Australian classification system, which has half-star increments for their hotels, making it possible to find 1.5-star hotels (Arzaghi et al., 2023). According to Vallen and Vallen (2017), some other differences and similarities can be pointed out. In Sweden, Germany, Switzerland and France, the "Hotel Garni" means no restaurant but includes continental breakfast. Besides the 1-to-5 classifications in Switzerland, a luxury class, "Gran Tourism" or "Gran Especial," has been added. The same happens in Italy, India, and Spain with an extra classification of 5-star "Deluxe" (UNWTO, 2015). The Irish Tourist Board takes a different approach, listing the facilities available (e.g., elevator, air conditioning, laundry) rather than grading them. Directories of the European Community follow a different approach and classify by location: seaside/countryside, small town/large city. European auto clubs go further by distinguishing privately owned from government-run accommodations (Vallen and Vallen (2017). In this context, Spain has standardised its Paradores' rating system, consisting of a government-operated chain of charming hotels in historic buildings. Portugal also has its Pousadas, which can be compared to the Spanish Paradores. In Japan, the traditional inns, the Ryokans, are rated according to their rooms, baths and gardens.

Although the most frequent and worldwide recognisable is a 1-to-5-star classification (Tiwari and Omar, 2023), it is still necessary to work on a universal, more credible and more customer-oriented system so that international travellers can have a more accurate picture of what hotels are offering (Núñez-Serrano et al., 2014). Classification systems serve hotels, hotel guests, and the travel trade, such as tour operators and travel agencies (Narangajavana and Hu, 2008; Nunkoo et al., 2020). In some cases, online travel agencies (OTA) show the official star-classifications side by side with their guest rating scores of the hotels displayed on their online platforms (Koutoulas and Vagena, 2023). The main limitation in using star-classifications for comparing hotels is the fragmentation of hotel classification systems, as each country, and sometimes each region, uses its system with a distinct set of criteria, thus creating confusion to hotel guests about what level of quality and comfort to expect (Núñez-Serrano et al., 2014).

In addition to being based primarily on facilities and services, traditional classification systems also face criticism for other limitations. These include the reliance on scheduled inspections, which may not reflect the hotel's continuous performance. Moreover, consumers are frequently unaware of the criteria underlying star-classification, leading to misunderstandings or mistrust (UNWTO, 2015).

More recently, new classification and rating initiatives have emerged. In 2024, Michelin introduced the "Michelin Keys" for hotels, aiming to recognise outstanding establishments worldwide based on consistent excellence and guest experience (Guide, 2024). At the same time, several hotels – particularly in the Middle East – have adopted unofficial "6-star" or "7-star" labels as part of branding strategies, despite the absence of formal global standards. Among the most notable examples are the Burj Al Arab and the Jumeirah Marsa Al Arab, both in Dubai, which are often marketed as "7-star hotels" (Forbes, 2023; Jumeirah Group, 2024). These differences among classification systems reflect the respective countries' cultural, economic, or national traditions (Maravić, 2017).

Online guest reviews and integrated approaches

With the continued growth of social media and online reservation platforms which allow and encourage guest

feedback, the playing field for hotel classification is changing rapidly. The information on hotels' characteristics and attributes and the customers' experiences, reviews, and scores have become increasingly available directly to travellers (Arzaghi et al., 2023). A recent systematic review by Pestana et al. (2024) mapped the growing body of literature on online hotel reviews, emphasizing their rising influence on service quality assessment and classification methods. Consumers are giving more importance to ratings given by other consumers, and less importance to official classifications. Recent studies have demonstrated that sentiment analysis applied to guest reviews can effectively forecast hotel performance, providing valuable predictive insights complementary to traditional star-classifications (Krey et al., 2024). Therefore, eWOM can significantly impact the reputation of a hotel and booking rates. Positive reviews attract potential guests, while negative feedback deters them from booking (Hensens, 2015). Most of the online reviews focus on service quality. At the same time, conventional classification systems tend to focus primarily on objective, tangible criteria such as the availability and size of facilities and services, occasionally on subjective tangible criteria such as cleanliness and state of maintenance, and rarely on service quality (Hensens et al., 2010).

The customers' view of hotel quality is largely subjective and depends on their perceptions of its characteristics, facilities, services, location, and even the price. For instance, Kim et al. (2022) found systematic differences in online reviews between distinct traveller segments, highlighting the importance of incorporating varied consumer perspectives into classification frameworks. eWOM is a staple feature of online customer-to-customer communication, reducing information asymmetry of lesser-known hotels more than higher-quality hotels (Yang et al., 2018). Specialised sites, such as Tripadvisor.com, and customer reviews and scores provided by travel sites, like Booking.com and Expedia.com, have significantly contributed to resolving the quality information problem in the travel and hotel industries (Li et al., 2017) while also providing review scores that simplify comparisons.

Nowadays, many online platforms generate a substantial number of reviews and user-generated content, including hotel reviews and ratings. This amount of new data may play a crucial role in decision-making by providing additional information and ultimately influencing the traveller. Besides ratings and textual comments, most review platforms allow users to upload photos, offering visual evidence that enhances the credibility and richness of guest feedback. Recent studies show that the consistency between visual and verbal content can significantly impact consumer perception and hotel ratings (Liu et al., 2024). Moreover, hotel managers can publicly respond to reviews, a practice that has been shown to positively influence booking behaviour when responses are timely and customer-focused (Krey et al., 2024; Lopes et al., 2024). These systems can offer an independent and trusted reference on the standard and quality of hotel service and facilities, thereby facilitating consumers in choosing their accommodation. They also provide a framework for accommodation providers to market, position themselves appropriately, and leverage their investments in the quality of their product-service offers (UNWTO, 2014).

However, one of the preconditions for this is sharing accurate information and sometimes it can be a problem if customers give biased and superficial reviews or inadequate observation (Hensens, 2015). The presence of fake or manipulated reviews further undermines the credibility of user-generated content. As Tuomi (2021) points out, the emergence of deepfake consumer reviews in tourism makes it increasingly difficult for other users to assess the authenticity and reliability of the feedback they read.

Additionally, the large number of reviews makes it time-consuming for customers to read and draw conclusions. Since these issues can make it difficult for customers to make decisions confidently, authenticating such reviews and scores may constitute a basis for future demand, as the experience of past customers is a key criterion for choosing a hotel (Arzaghi et al., 2023). Travellers can become overwhelmed by the sheer volume of reviews and struggle to extract relevant and valuable information for their selection process. This issue of information overload can make decision-making more difficult and time-consuming for potential guests. As a result, there is a need for the compilation and summarisation of this data to aid travellers in overcoming the discrepancies between the star-classification system and guest satisfaction. Therefore, conventional classification systems and online travel platforms, such as Booking.com or TripAdvisor, may complement each other through integrated classification models. Several countries are moving towards integrated models, which can be grouped into two types: full integration and comparative performance.

Full integration implies that the hotel can adjust its star level up or down, depending on its perceived quality, as measured by guest reviews, compared to other hotels. In a comparative performance model, the aggregated guest review rating is displayed separately from the hotel classification. However, integrating consumer reviews into hotel classification is not new; some travel sites have been doing so for the past few years, such as Hotwire.com and Priceline.com, which primarily operate in the United States. These sites sell rooms not in specific hotels but in classes of hotels in

general areas, such as a 4-star hotel in Times Square, New York City, for example. The accuracy of the star information is, therefore, critical to the success of these sites. Consumers may not revisit the travel site if they purchase a 4-star hotel but feel it is a 3-star hotel due to the quality of service or facilities. Norway and Switzerland have established models for integrating guest reviews into hotel classification, and regions such as the United Arab Emirates, Germany, and Australia are also developing integrated platforms. The model in Norway, developed by QualityMark Norway and yet to be implemented due to resistance from major hotel chains, is an example of full-scale integration. On the other hand, the system currently being used in Switzerland, which uses Hotelstars Union criteria for its official classification, involves instead a parallel presentation of aggregated guest review information alongside traditional hotel classifications (UNWTO, 2014; UNWTO, 2015).

In this context, the integration of guest reviews from online platforms into traditional hotel classification systems focuses on the feasibility of incorporating such reviews at the star-classification level. On the one hand, this approach respects the traditional characteristics and classification models of each country or region (in this case, star-classification) and, on the other, it integrates a classification obtained through machine learning techniques based on Booking.com reviews. These recent contributions underscore the need for a model that unites institutional classifications with user perception data – an integration still underexplored in empirical research, and which this study aims to address. Therefore, this model is considered innovative in that it can be adapted to the specific contexts of any region or country.

Methods

Data

For this study, a database combining hotel star-classifications from the Portuguese National Tourism Board (Turismo de Portugal) with online guest review ratings from Booking.com was compiled. All data were collected in October 2021, with star-classifications sourced from the official database and review scores gathered manually from Booking.com. To build our sample, we began by consulting the list of 1,426 hotels registered with Turismo de Portugal at the time. This registry gathers all information regarding official registration number, star classification, number of rooms and beds, and other available facilities. This search revealed that 226 hotels did not have all the information available and were therefore excluded due to their inconsistency. Thus, our final sample consisted of 1,200 hotels.

The Booking.com guest review scores are obtained after a guest has checked out of a property that had a reservation made through the platform. The platform emails the guest one questionnaire containing one mandatory question on the overall score of the property, 6 specific questions relating to cleanliness, comfort, value, facilities, location and staff that are optional and a few more optional ratings on breakfast and Wi-Fi facilities. Guests are invited to rate the property by attributing scores from 1 to 10. The platform also encourages guests to provide feedback in the form of an open question, even though it is also optional. After that, the average values are recalculated and, together with the characteristics, prices and photographs of the hotel, the ratings for each of the 8 items previously mentioned are presented (Overall, Value for money, Cleanliness, Location, Facilities, Comfort, Staff, Wi-Fi). The score metrics only considers reviews of the previous 36 months and is in constant update.

Software and libraries

The RStudio program 2023.0301 with R-4.3.0 was used to analyse the data. To fulfil the objectives of this work several R libraries were used, including: *MASS* for support functions; *dplyr* for data manipulation; *e1071* for support vector machines training; *psych* and *gtsummary for* summary statistics; *caret* for classification and regression training; *cluster* for cluster analysis; *randomForest* for Random Forests classification and regression; *factoextra* to extract and visualise the results of multivariate data.

Support vector machine

The Support Vector Machine (SVM) is a popular machine learning algorithm (Bishop, 2006; Cervantes et al., 2020) originally derived for binary classification problems. In its simplest form, known as *hard margin* SVM, the model seeks the optimal (linear) decision boundary $y(\mathbf{x}) = w_1 x_1 + \cdots + w_D x_D + w_0$, as the one that perfectly separates, with the maximum margin, the data points of two linearly separable classes in the feature space. The margin is defined as the distance between the decision boundary and the closest data points from each class, the support vectors. The solution is obtained by solving, with quadratic programming techniques, a convex optimization problem. When the classes are not linearly separable or even overlapped, the model assumes a *soft margin* version that allows data points inside the margin (some of them, misclassified) by introducing some relaxation variables (slack variables) ε_n for each data point. The optimization problem is then expressed as

$$\min \left[\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{n=1}^{N}\varepsilon_n\right] \quad \textit{subject to}$$

$$t_n(\mathbf{w}\cdot\mathbf{x}_n + w_0) \geq 1 - \varepsilon_n \;\; \textit{and} \;\; \varepsilon_n \geq 0; \quad n = 1, \ldots, N$$

where n represents the nth observation index from a dataset with N observations, $\mathbf{w} = (w_1, \ldots, w_D)$, $\mathbf{x}_n = (x_{1n}, \ldots, x_{Dn})$ is the nth data point and $t_n \in \{-1, 1\}$ is the corresponding target value (class). The hyperparameter C controls the trade-off between minimizing misclassifications (high value of C) and maximizing the margin (low value of C). The SVM model can be further extended to provide nonlinear decision boundaries. The idea is to map the input features to a higher dimensional space using a nonlinear mapping where the linear model can solve the problem more easily (with high probability). By mapping back such solution to the original space, a nonlinear decision boundary is achieved. This process is possible by using the well-known kernel trick (Bishop, 2006; Cervantes et al., 2020), where a kernel function is used to compute the dot product in the higher dimensional space. Moreover, the SVM model can be applied to a multi-class setting by considering a *one* vs *all* strategy, building an SVM for each class against the others and combining the results. In this work it was considered the linear kernel, corresponding to perform no mapping to higher dimensionality (the original SVM formulation) and the radial basis function (RBF) kernel, to allow nonlinear solutions.

Gradient Boosting Machines

Gradient Boosting Machines (GBM) are a class of ensemble methods whose rationale is based on the idea that combining several weak models (eventually slightly better than random guessing) can produce a single stronger model (Bishop, 2006; Mienye and Sun, 2022; Natekin and Knoll, 2013). Usually using low depth decision trees as weak models, GBM's training is performed in sequence in a way that the following weak model is trained to correct the errors of the previous ones. Formally, GBM is an additive model

$$Y_M(\mathbf{x}) = \sum_{m=1}^{M}\alpha_m y_m(\mathbf{x})$$

where each y_m is a (weak) decision tree which is sequentially added to the ensemble, guaranteeing that the previous $m-1$ trees stay fixed. For that, gradient descent is used to minimise, at step m, a cost function of the form:

$$\sum_{n=1}^{N} E(t_n, Y_{m-1} + \alpha_m y_m)$$

where Y_{m-1} (the current model) is fixed and $E(\cdot)$ is an appropriate measure of the discrepancy between the targets and the model's output. The number M of weak models is usually in the order of dozens or hundreds and α_m weights each weak model in the sum according to its performance (high performing models get higher weight in the ensemble). GBMs are known to have good performance and are quite flexible regarding the type of data used, as they can easily handle both numerical and categorical data. Tuning a GBM model includes estimating the best set of hyperparameters that include: the number of trees, the learning rate of the gradient descent optimization, the (interaction) depth of the trees and the minimum number of observations required in a node. In this work, the hyperparameters are tuned using a cross-validation scheme.

K-means clustering

Clustering algorithms leverage the underlying structure of a data distribution by partitioning the dataset into clusters based on specified criteria without prior knowledge of the dataset. Each cluster contains similar data instances, distinct from those in other clusters, with dissimilarity measured according to the algorithm's objective and the data characteristics. Clustering is crucial in many data-driven applications and is extensively studied in fields like optimization, bioinformatics, computational geometry, statistics, pattern recognition, and image processing (Bishop, 2006; Ikotun et al., 2023). In this work, k-means clustering is used to discover cluster structures within Portuguese hotels based on the Booking.com's online scores.

k-means is a popular partitioning clustering algorithm based on the distances between data points and cluster centroids. The algorithm starts by initializing k centroids (representatives of the k clusters), either randomly or through advanced techniques like density-based initialization. Each data point is then assigned to the nearest centroid, and the centroids are recalculated. This process is repeated until the centroids stabilize (convergence), reaching a local minimum of the objective function (Bishop, 2006; Ikotun et al., 2023; Steinley and Brusco, 2007).

Choosing the optimal number of clusters k to use is a fundamental problem for k-means. Incorporating domain knowledge about the data can provide valuable insights into a reasonable range for k. In this work, for example, a sensible k would be 5, corresponding to the number of hotel stars. Other strategies include the elbow method or the Gap statistic as ways of estimating such value. In the former, the total within sum of squares errors (SSE), measuring how tightly the data points in a cluster are grouped around the cluster centroid, is computed for several values of

Table 1. Median scores (interquartile range) for Booking.com's 8 review items, by hotel star category. *p*-value obtained for the Kruskal-Wallis test for significant differences between categories.

Review item	* (*N* = 54)	** (*N* = 259)	*** (*N* = 435)	**** (*N* = 532)	***** (*N* = 146)	*p*-value
Overall	8.00 (7.60 – 8.40)	8.10 (7.80 – 8.60)	8.30 (7.90 – 8.60)	8.60 (8.20 – 8.90)	8.90 (8.60 – 9.20)	<0.001
Cleanliness	8.40 (8.20 – 8.90)	8.60 (8.20 – 9.00)	8.70 (8.30 – 9.10)	9.00 (8.60 – 9.30)	9.30 (9.00 – 9.50)	<0.001
Location	8.85 (8.40 – 9.28)	8.80 (8.20 – 9.25)	8.80 (8.30 – 9.20)	8.90 (8.50 – 9.40)	9.20 (8.80 – 9.50)	<0.001
Staff	8.80 (8.50 – 9.20)	8.90 (8.50 – 9.20)	8.90 (8.50 – 9.20)	9.00 (8.70 – 9.30)	9.20 (8.93 – 9.50)	<0.001
Comfort	8.00 (7.50 – 8.40)	8.10 (7.70 – 8.60)	8.40 (7.90 – 8.80)	8.80 (8.40 – 9.20)	9.30 (9.00 – 9.50)	<0.001
Value for money	8.30 (7.80 – 8.60)	8.30 (7.80 – 8.70)	8.20 (7.80 – 8.60)	8.30 (7.90 – 8.60)	8.40 (8.10 – 8.60)	0.13
Facilities	7.70 (7.30 – 8.07)	7.90 (7.40 – 8.40)	8.10 (7.70 – 8.60)	8.60 (8.10 – 8.93)	9.00 (8.70 – 9.30)	<0.001
Wi-Fi	8.10 (7.50 – 8.60)	8.30 (7.70 – 8.70)	8.30 (7.80 – 8.70)	8.50 (8.10 – 8.90)	8.80 (8.33 – 9.20)	<0.001

k and the point where the rate of decrease in SSE sharply slows down (the "elbow point") is chosen. This point represents a balance between the compactness of the clusters (low SSE) and the simplicity of the model (fewer clusters). The Gap statistic (Tibshirani et al., 2001) compares the total within-cluster variation for different values of k with their expected values under a null reference distribution of the data. The goal is to identify the number of clusters that significantly improves clustering performance over random noise, which corresponds to a higher Gap value.

Random forests and variable importance

Random forest is another class of ensemble models that builds many trees and combines their predictions into a single one (Bishop, 2006; Mienye and Sun, 2022). Differently from boosting models, in random forests many bootstrap samples (sampling with replacement) are obtained from the original set, and each sample is trained with a full tree (training is performed in parallel). Full trees here give more model variance but lower bias. By combining predictions, variance is also reduced, and a more robust model is obtained. In the construction of each tree, only a random subset of the available input features is allowed to compete for each node which fosters the variability among trees (reducing the effect of stronger variables that consistently win the first nodes of every tree). Random forests allow us to track and measure the importance of each feature in the construction of the model. Two approaches to measure importance are usually provided: the mean decrease in accuracy (MDA) and the mean decrease in impurity (MDI) as measured by the Gini index. MDA is generally preferred as it directly measures how much permuting a variable reduces prediction accuracy, reflecting its true contribution to cluster discrimination, while MDI can be biased toward variables with more categories or continuous scales (Louppe et al., 2013; Sikdar et al., 2025).

In this work, random forests are used to measure feature importance in predicting cluster membership for the clustering solutions obtained with k-means. This allows to identify which Booking.com's scores are more important to define the clusters and therefore characterise the corresponding hotels.

Results

Description of the hotel sample

Table 1 presents some descriptive statistics by star category for the scores of Booking.com's 8 review items, along with the p-value for the non-parametric Kruskal-Wallis test for significant differences between categories.

Based on the results, higher star-classifications generally correspond to higher review ratings. However, this trend is not uniform across all review items, as (Figure 1(a)) illustrates. Specifically, the relationship between star-classification and review score is less straightforward for location and value for money. In terms of location, 1-star hotels receive higher ratings compared to 2- and 3-star hotels. As mentioned by Masiero et al. (2019), location is related to the proximity to points of interest, transportation convenience, and the surrounding environment. In this context, 1-star hotels are generally associated with smaller hotels, sometimes located in pre-existing buildings in historic centres and, therefore, close to transport infrastructures such as metro and train stations or bus stops. Also, in value for money, 1-star hotels have the same rating as 2-star hotels and higher than 3-star hotels, for example, while no significant differences are found (p = 0.13) between the five categories. As Gupta and Kim (2009) refer, value for money refers to seeking accommodations offering the highest value at the lowest possible price. In this context, 1-star hotels – typically positioned in the budget segment – tend to have lower prices than the rest and, focusing on the quality of service, regardless of the existing level of facilities, they may find a

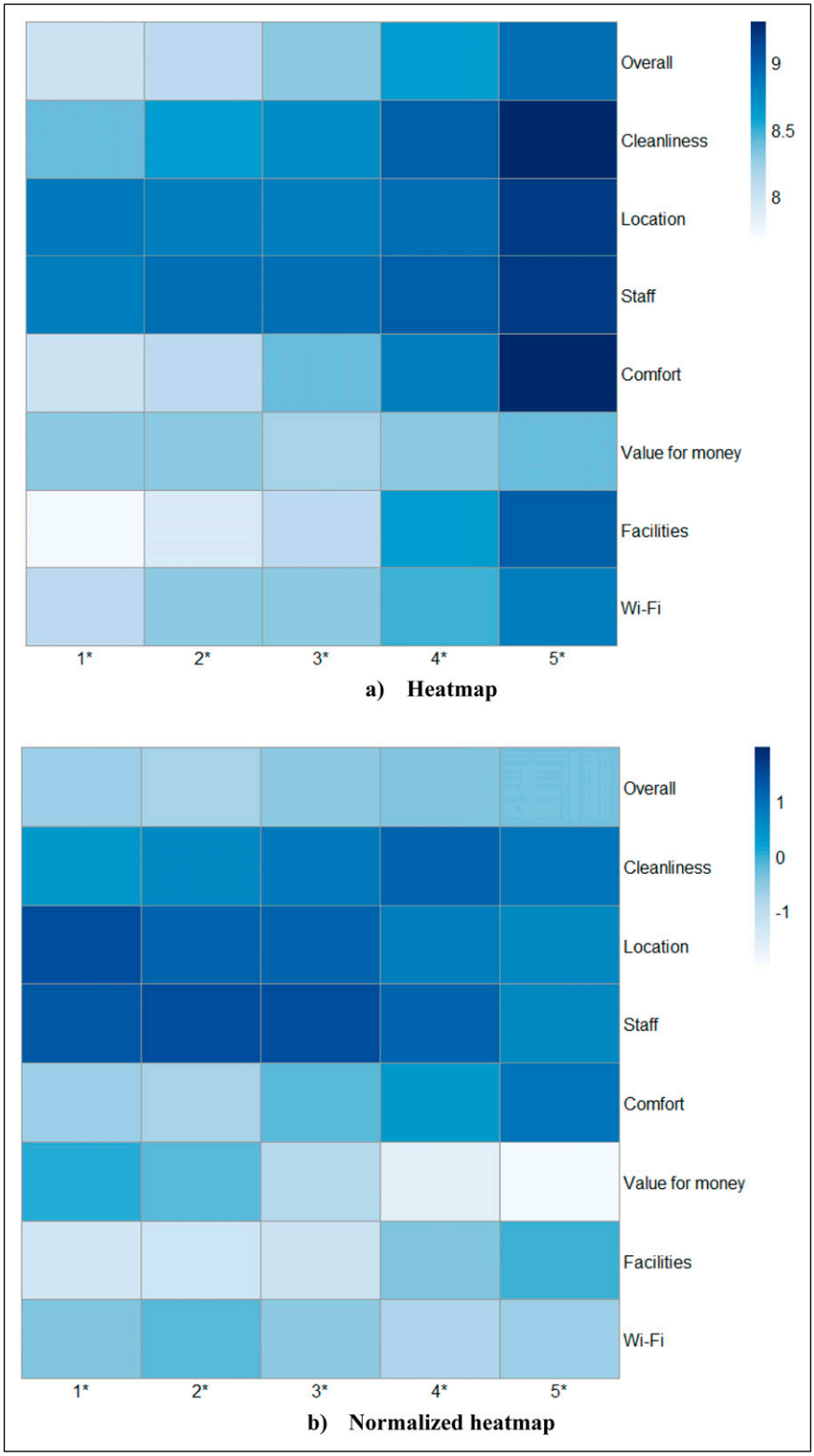

Figure 1. (a) Heatmap of Booking.com's median values for the 5-star hotel ranking. (b) Normalized heatmap (by column) of Booking.com's median values for the 5-star hotel ranking.

Table 2. Accuracy of the SVM model for hotel star prediction based on guest reviews using linear and RBF kernel functions and different values of C.

	Linear kernel		RBF kernel	
	$C = 5$	$C = 100$	$C = 5$	$C = 100$
Accuracy train	0.61	0.61	0.67	0.81
Accuracy test	0.59	0.59	0.57	0.54

strategic advantage here compared to other higher star-classification hotels. Within this 5-star classification system, it is also interesting to note that, regardless of the star category, Cleanliness, Location and Staff are rated with higher values than the other items. This is clear from the normalized heatmap of (Figure 1(b)).

Predicting star-classifications based on guest review scores

We applied two predictive models to see if Booking.com's guest scores constitute a good set of predictors of the official hotel star-classification, and therefore, verify if the customer perceptions of the quality of a stay align with the parameters that accredit a given star-classification.

SVM models were trained using both linear and RBF kernel functions to predict the hotel category based on the scores collected from Booking.com. The dataset was split with 80% of data for training the model with two values for the C hyperparameter (5 and 100), and 20% for testing. The results are summarized in Table 2 and indicate that changes in the kernel function from linear to nonlinear do not improve generalization ability (in fact, the test set accuracy decreases). Moreover, using a high-cost parameter with the RBF kernel greatly increases training accuracy but not the accuracy in the test set, showing that the model is overfitting in the training phase. Overall, there is only 54%–59% accuracy in the prediction of hotel star-classification.

In addition to the SVM models, a GBM was also trained using a repeated cross-validation approach with 10-fold cross-validation repeated 10 times. The hyperparameters of the GBM model were fine-tuned using an exhaustive grid search. The grid search explored the interaction depth in the range of 1 to 3, the number of trees ranging from 100 to 1,000 in steps of 50, a learning rate of 0.1, and a minimum number of observations required in a node set to 5. The trained GBM model achieved the best performance with the following hyperparameters: 100 trees, an interaction depth of 2 and a minimum of 5 observations per node. Table 3 presents the test set confusion matrix of this best GBM model, where we can easily compute a 0.53 accuracy (95% CI: 0.48 – 0.58) metric. Again, a weak overall accuracy is observed, reinforcing that customer perceptions are not directly related to the star categories.

Table 3. Confusion matrix of the registered hotel star categories and the predicted categories using GBM.

		Hotel star category				
		*	**	***	****	*****
Predicted category	*	1	2	1	0	0
	**	5	17	19	6	0
	***	4	34	54	23	1
	****	2	8	36	102	20
	*****	0	0	0	6	16

Table 4. Performance measures of the GBM model on the use of user-generated content from Booking.com on the prediction of hotel categories. PPV – Positive predicted value; NPV – Negative Predicted Value; F1-measure = 2 * Precision * Recall / (Precision + Recall).

	*	**	***	****	*****
Sensitivity (Recall)	0.08	0.28	0.49	0.74	0.43
Specificity	0.99	0.90	0.75	0.70	0.98
PPV (Precision)	0.25	0.36	0.47	0.61	0.73
NPV	0.97	0.86	0.77	0.81	0.94
Prevalence	0.03	0.17	0.31	0.38	0.10
Detection prevalence	0.01	0.13	0.32	0.47	0.06
Balanced accuracy	0.54	0.59	0.62	0.72	0.71
F1-measure	0.12	0.32	0.48	0.67	0.54

Table 4 presents a more complete set of the model's performance measures for each star category, giving a deeper understanding of the results. Poor recall per category (with the exception of 4-star hotels) show that the model has difficulties in correctly detecting the star of the hotel, although it quite well predicts which star is not, given the very high specificity values, specifically in the determination of the extreme hotel categories, that is 1- (0.99) and 5-star hotels (0.98). Precision values are also low, mainly for 1- to 3-star hotels, showing low confidence when predicting in such categories. Finally, the F1-measure, which gives us an overall (balanced) view between Precision and Recall, shows that the model has clear difficulties in using the Booking.com's scores to discriminate between hotel star categories.

Hotel segmentation based on Booking.com average review scores

The previous results have shown that the perception of the quality of a stay by Booking.com customers does not reflect the hotel segmentation that the star categorization currently provides. Therefore, we used k-means clustering to investigate the existence of a different

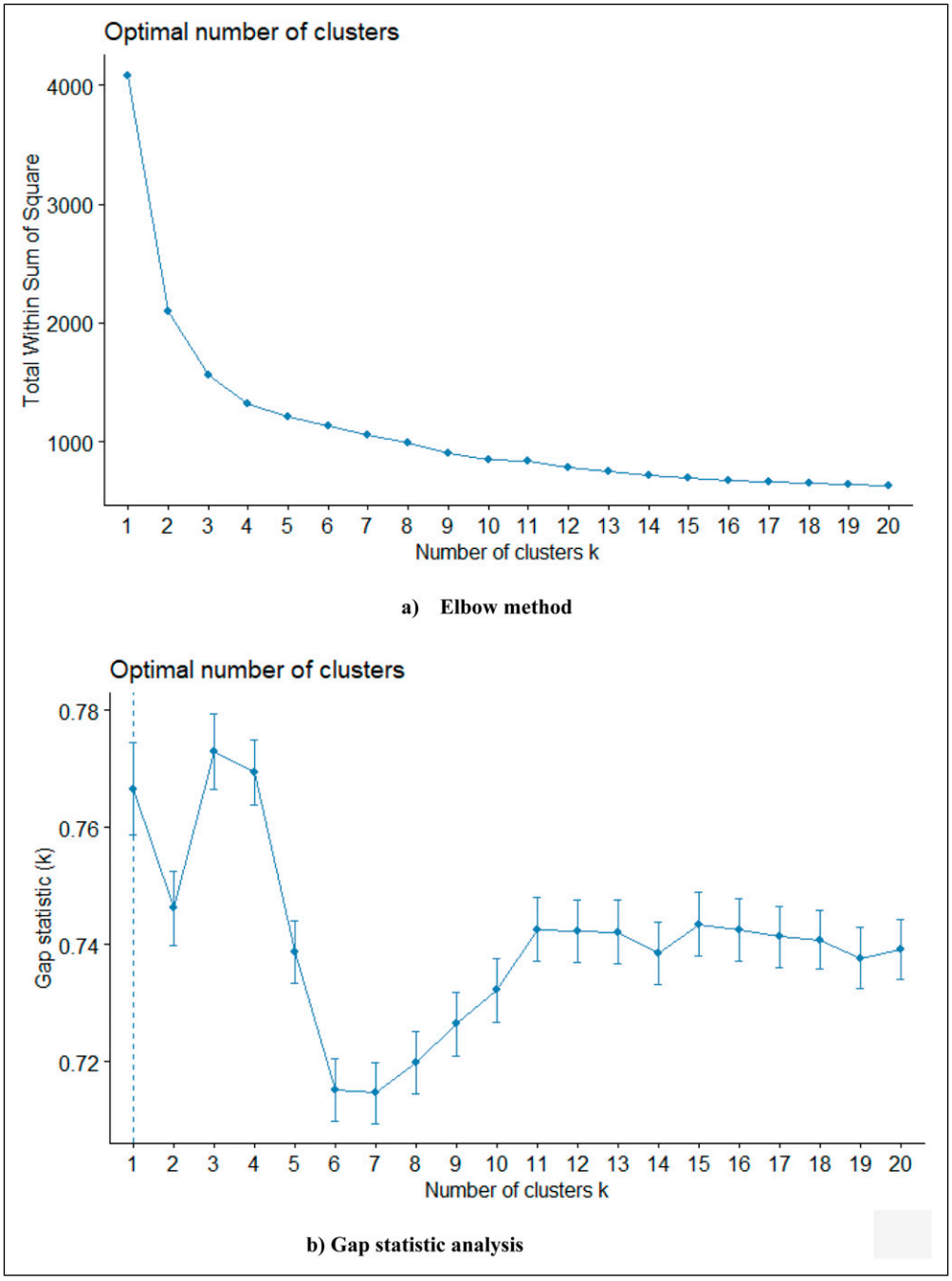

Figure 2. (a) Graphical representation of the elbow method, where a notorious curve inflexion is observed at 3 to 4 clusters. The addition of more clusters does not generate a significant reduction in the total within-clusters sum of squares; (b) Graphical representation of the Gap statistic analysis, also showing an optimal number of clusters of 3 or 4.

segmentation structure that could better reflect such perceptions. We started to estimate the optimal number of clusters for this data using both the elbow method and the Gap statistic. As Figure 2 shows, both approaches estimate a hotel segmentation solution with 3 to 4 clusters as the best options. Interestingly, there seems to be no advantage in using 5 clusters as the star-classification system suggests.

k-means was then applied to our data, both using k = 3 and k = 4. Given each of the hotel segmentation solutions, the mean value of the 8 Booking.com item scores and the corresponding within sum of squares were computed for each cluster and are detailed in Table 5. It is possible to see that the solution with 4 clusters is slightly better that the 3-cluster solution, providing more compact clusters (higher between-

Table 5. Mean scores of Booking.com reviews obtained for each cluster in the 3- and 4-cluster solutions. Between / Within SS ratio indicates the ratio of between clusters sum of squares and within clusters sum of squares, providing a measure of cluster compactness.

Cluster solution		Number of hotels	Overall score	Cleanliness	Location	Staff	Comfort	Value for money	Facilities	Wi-Fi
3 clusters	1	191	7.40	7.80	8.32	8.31	7.38	7.40	7.11	7.24
	2	497	8.22	8.66	8.72	8.81	8.39	8.12	8.10	8.16
	3	512	8.91	9.32	9.14	9.32	9.18	8.68	8.94	8.85
	Between / Within SS ratio		61.7%							
4 clusters	1	335	9.02	9.43	9.24	9.41	8.77	8.77	9.07	9.01
	2	430	8.51	8.93	8.86	9.01	8.36	8.36	8.45	8.40
	3	324	7.94	8.38	8.54	8.64	7.89	7.98	7.77	7.89
	4	111	7.18	7.55	8.31	8.16	7.17	7.17	6.84	6.99
	Between / Within SS ratio		67.6%							

within sum of squares ratio). In both solutions, all clusters are quite balanced in terms of the number of hotels except for the cluster with lower scores (cluster 1 in the 3-cluster solution and cluster 4 in the 4-cluster solution). It is also interesting to notice that both solutions provide an ordered ranking of all review scores.

Figure 3 plots each of the cluster solutions along with the star category for each hotel in the principal components space. Although in general it is possible to observe that a higher star category tends to have higher review scores, each cluster is composed of hotels having different star categories showing that this classification is not completely associated with the customer perception of quality. Taking the 4-cluster solution as an example, we see that cluster 4 (the cluster rated with lower review score values) contains hotels from 1 to 4 stars, while cluster 1, which is essentially populated with 5- and 4-star hotels, also contains 1-, 2- and 3- star hotels. A similar behaviour is observed in the solution with 3 clusters.

Considering the segmentation structures obtained from *k*-means, it was further investigated the impact of each Booking.com item to their characterization. For that, a random forest model was built for each of the cluster solutions using the cluster assignments as the target variable, and variable importance was measured. Figure 4 presents the variable importance plots for both cases.

For the 3-cluster solution, the Overall, Cleanliness, Facilities and Comfort scores present the highest mean decrease in the Gini impurity index (191.70, 130.54, 123.33, and 112.67, respectively), indicating that these are the items that contribute the most for the purity of nodes in the random forest and therefore are strongly present in the trees of the random forest. Additionally, Wi-Fi is the score that provides the highest mean decrease in accuracy (57.21) in the model, indicating its importance in predicting cluster membership (Figure 4(a)).

The results are similar in the 4-cluster solution, with the same variables considered by the same order of importance. Thus, the Overall, Cleanliness, Facilities and Comfort scores present 202.11, 157.68, 149.73, and 117.33, respectively mean decrease in the Gini impurity index, and Wi-Fi also is identified as the score with the highest mean decrease accuracy (62.15), as shown in Figure 4(b).

It is noticed that Value for money, Location and Staff are consistently in the bottom rank of importance in all situations, thus not contributing significantly to the reduction of impurity, cluster distinction or cluster membership prediction.

Discussion

This study analysed the agreement between hotel star categories in Portugal and the corresponding mean review scores of Booking.com on 8 items: Overall satisfaction, Cleanliness, Comfort, Facilities, Staff, Value for money, Location and Wi-Fi. The results, regardless of the methodology of machine learning employed (SVM or ensemble GBM), indicate a weak association between the hotel star category and the mean review scores. In fact, the maximum accuracy obtained with an SVM was 59%, whereas for GBM was 53%. Other authors (Soifer et al., 2020) also reported such discrepancy between hotel attributes and facilities and online user ratings which has been found to be particularly relevant in 5-star hotels in Lisbon (Rita et al., 2022) and reflects the disconformity between consumer expectations and consumer experience (Li et al., 2020).

One of the reasons for the discrepancy between the hotel star-classification and the guest review scores

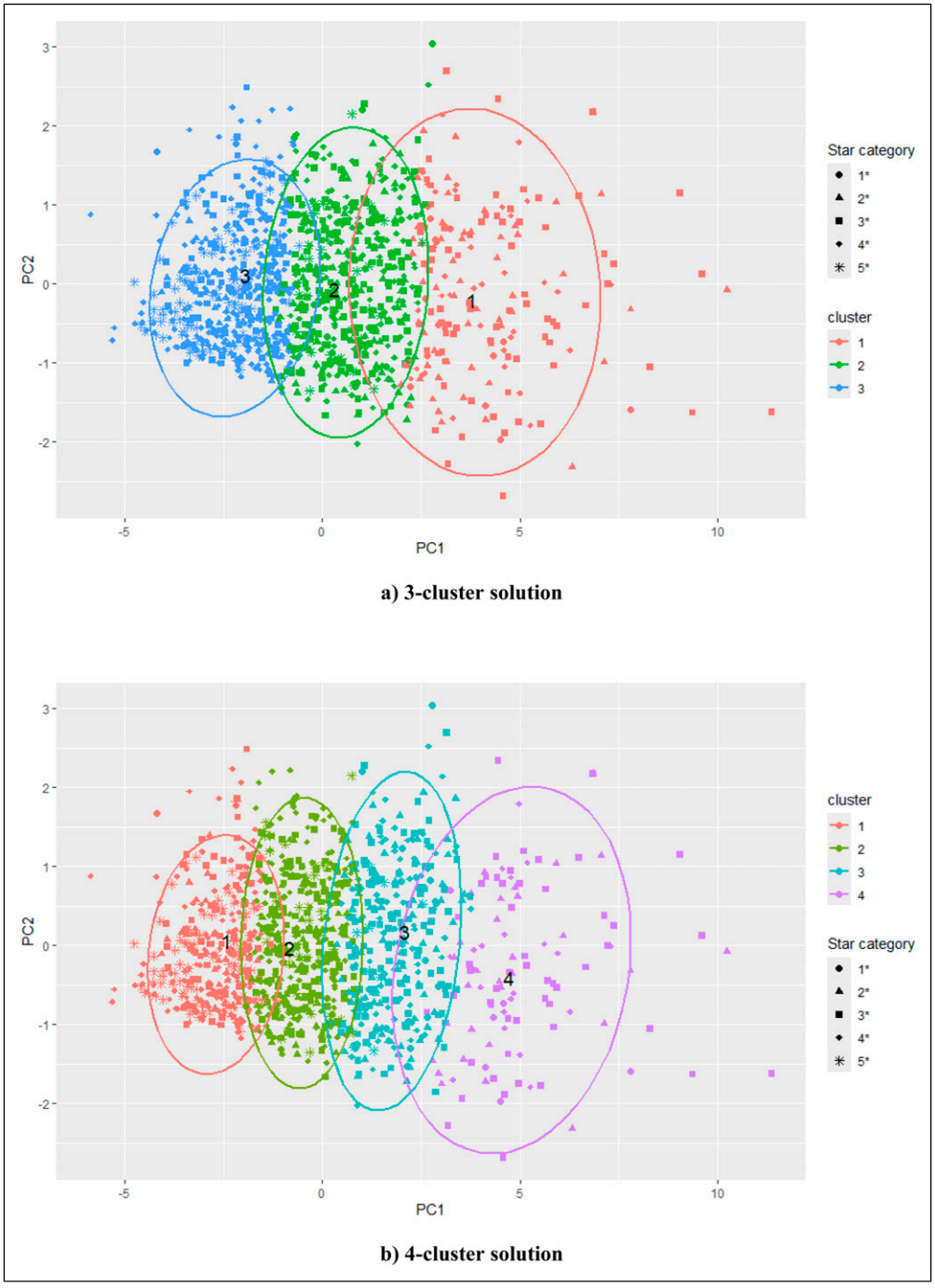

Figure 3. (a) Graphical representation of the cluster solutions with 3 clusters; (b) Graphical representation of the cluster solutions with 4 clusters. Each data point (hotel) is represented by a specific symbol associated with the star category. Clusters are represented with different colors, and a bivariate gaussian 95% confidence ellipse is plotted to approximate the shape and spread of each cluster in the principal components space.

might be related to the fact that the hotel star-classification is reviewed only once every 5 years whereas the Booking.com scores retain the reviews of the previous 36-month period and keep updating the metric whenever new reviews are added. Another reason is related to the fact that star-classification systems rely more on facilities and level of service, while guest reviews are based on expectations and quality of experience (UNWTO, 2014), and, therefore, not always the star-classification level matches the guest reviews appreciation.

A previous study on 1,500 reviews of 50 small and medium hotels in the region of Lisbon identified that

guests pay more attention to the room conditions (including cleanliness and comfort to rest) when writing a review and attributing a score (Chaves et al., 2012), which is in line with the results here presented for the 3- and 4-cluster solutions, for which Cleanliness is the second most important variable for the categorization of hotel review scores.

As some other countries are working to implement integrated systems (UNWTO, 2014; UNWTO, 2015), the 3- or 4-cluster solution that derives from the Booking.com scores could be considered an option to complement the current hotel star system, gathering the best of the two worlds. On one hand, the hotel star category would provide detailed information regarding the type of facilities and services guests should expect. On the other hand, the integration of a categorization system based on reviews provided by other guests would generate trustful, easy to interpret information, directed mainly to service quality.

According to the predicted cluster membership, each hotel could receive a quality rating of Bronze, Silver or Gold categorization, in the case of adoption of the 3-cluster solution or Bronze, Silver, Gold or Platinum in the case of 4-cluster. This would be very helpful for the traveller searching for a hotel in the context of the multicriteria decision process. For instance, if the traveller is looking for a 3-star hotel, providing information regarding the experience of previous guests such as a very good experience (3-star Gold) or a not-so-remarkable experience (3-star Bronze) clearly facilitates the process, yet not preventing the traveller from choosing based on other features namely price or location.

In fact, online platforms such as TripAdvisor or Booking.com have employed a similar quality rating system that serves as a guide for a variety of alternative accommodations such as apartments or villas. This rating system includes information on both the facilities and the average review score as well as anonymized and aggregated historical data, corroborating the importance of adding summarized information of the review scores to the hotel star-classification. Such system is welcomed by the hotel management sector (Koutoulas and Vagena, 2023; Vagena and Manoussakis, 2021) and would overcome the limitation of the non-universality of the star-classification system. Additionally, it would include up-to-date and reliable information to the star-classification system.

This proposal also responds to current demands for integrating real-time, consumer-centred insights into regulatory systems in tourism, reflecting broader shifts toward digital trust and user empowerment.

Therefore, this research moves beyond technical modelling by proposing an applied framework that can bridge the current disconnection between institutional

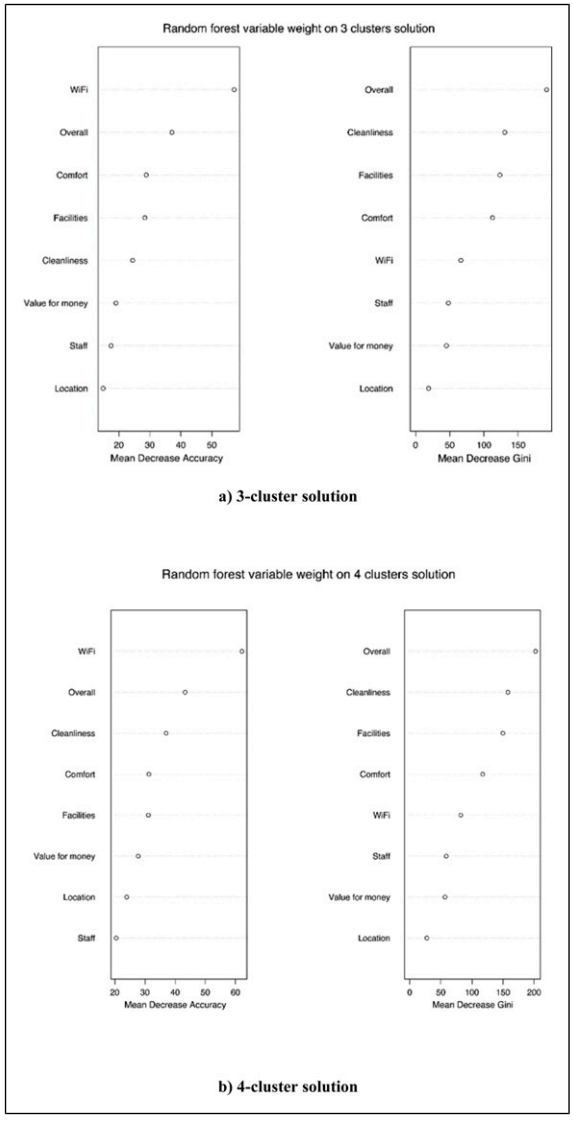

Figure 4. (a) Variable importance plots with 3-cluster solution; (b) Variable importance plots with 4-cluster solution.

classification systems and real guest experience. Whereas professional inspectors assess hotels based on standardised criteria related to physical infrastructure and service provision, guest reviews offer insights into how those services are perceived and experienced. Integrating both perspectives enables a more comprehensive and accurate reflection of hotel quality in today's digital environment. This conceptual innovation expands the theoretical understanding of how quality is communicated and perceived in digital hospitality environments. On a practical level, the model introduces a dual-layer classification which offers clearer signals for travellers, supports data-driven decision-making for platforms and policymakers, and encourages hotels to prioritise experiential quality,

not just infrastructure compliance. For formal classification organizations, this model provides a path to modernise and enhance the credibility of national classification systems without undermining their traditional structure.

Conclusion

This study contributes to the ongoing debate about combining online guest reviews with traditional hotel classification systems. By leveraging data from Portuguese hotels and analysing guest satisfaction ratings on Booking.com, this paper highlights the misalignment between the existing star-classification system and the actual experiences of hotel guests. The weak association between hotel star-classifications and guest satisfaction scores emphasizes the need for a more dynamic and comprehensive approach to hotel classification. One of the most striking findings of this research is that hotel star-classifications, which primarily reflect facilities and services, do not consistently align with guest satisfaction scores related to factors such as cleanliness, comfort, and value for money, among others. This discrepancy indicates that the traditional star-classification system may not fully capture the subjective, experience-based aspects of hotel quality that modern travellers prioritize. As the analysis shows, while 5-star hotels generally score higher on most satisfaction items, lower-tier hotels, mainly 1- and 2-star establishments, can sometimes outperform higher-tier hotels on specific dimensions like location and value for money. This observation suggests that travellers are not solely motivated by the amenities a hotel provides but also by the quality of the service and the overall experience. Integrating guest reviews into the star-classification system offers a potential solution to this gap, providing a dual framework that enhances the objectivity of traditional classification with the subjectivity of guest reviews. Moreover, the introduction of machine learning techniques, such as Support Vector Machines (SVM) and Gradient Boosting Machines (GBM), in this context demonstrates that predictive models can be used to assess the service quality offered by hotels more accurately. Although the prediction models in this study showed moderate accuracy (around 53%–59%), they present an important first step in developing a refined classification system that could better serve both travellers and hoteliers.

This paper proposes an innovative hybrid classification model incorporating guest reviews alongside traditional star-classification. The clustering analysis results suggest the potential for 3- or 4-cluster solutions to complement existing star-classification and point to the feasibility of categorizing hotels by facilities and the quality of service as perceived by guests. For example, a hotel currently classified as 3-star could receive additional "quality" distinctions such as Bronze, Silver, or Gold based on guest review scores. This would result in a dual classification – e.g., 3-Star Bronze – that clearly communicates both the technical compliance of the hotel (in terms of infrastructure and services) and the experiential quality as perceived by guests. Such a label would be simple for consumers to interpret and could serve as an intuitive decision-making aid, particularly when comparing hotels within the same star category.

Theoretical implications

From a theoretical standpoint, this study contributes to reconceptualising hotel classification as a dynamic system that combines experiential guest data with formal institutional criteria. It addresses a known limitation in the literature, which often treats these sources as separate or incompatible. By proposing a hybrid model that integrates both perspectives, this research presents a new conceptual framework for understanding hotel quality. The use of machine learning reinforces this contribution, demonstrating how data-driven techniques can support more nuanced and multidimensional classification systems in the hospitality industry. Furthermore, the model introduces a clustering-based distinction—Bronze, Silver, or Gold (and Platinum in a four-cluster solution)—that complements existing star ratings and reflects guest-perceived service quality in a structured and interpretable format.

Managerial implications

At a practical level, the proposed model enables consumers to differentiate more effectively between hotels within the same star category, facilitating informed booking decisions that are based not only on technical classification but also on perceived service quality. This dual system provides additional transparency and clarity in a highly competitive digital marketplace. For hotel managers, the framework offers an opportunity to monitor and enhance performance based on guest perceptions, thereby improving satisfaction and competitive positioning. Public rating agencies and online travel intermediaries (OTAs) may also benefit, as this model presents a scalable and adaptable solution to modernise traditional classification systems without compromising their institutional structure. By integrating guest feedback into official frameworks, rating bodies can enhance public trust and provide a more accurate, up-to-date reflection of hotel quality – particularly relevant in a post-pandemic context where travellers increasingly prioritize hygiene, comfort, and staff engagement.

Limitations and future research

While this study provides a solid foundation for combining guest reviews with hotel classification systems, several limitations suggest new paths for future research. First, the data collected for this analysis refer to October 2021, during the recovery phase following the peak of the COVID-19 pandemic. Traveller expectations and behaviours may still have been influenced by pandemic-related concerns – particularly regarding hygiene and safety – which could have shaped review patterns. Future research should consider how guest priorities evolve in fully post-pandemic contexts, and whether the discrepancies between star-classifications and guest reviews persist. In addition, future studies could benefit from incorporating data from other review platforms, such as TripAdvisor or Google Reviews, to validate results and explore platform-specific variations in user behaviour and satisfaction assessment.

Additionally, future studies could benefit from incorporating larger datasets and using more advanced machine-learning techniques to improve the accuracy of predictive models. It's also crucial to explore how the combination of guest reviews with classification systems could be standardized across different regions and hotel types, ensuring the development of a universally applicable system.

Finally, a deeper examination of other factors influencing guest satisfaction, such as cultural differences, price sensitivity, and the role of loyalty programs, would further refine the proposed model and its effectiveness in capturing the full spectrum of hotel quality.

ORCID iDs

Pedro Couto https://orcid.org/0009-0000-2337-6455
Luís M. Silva https://orcid.org/0000-0001-9677-4315
Jorge Marques https://orcid.org/0000-0001-5392-5128

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the support UID/05105: REMIT – Investigação em Economia, Gestão e Tecnologias da Informação, and by CIDMA under the Portuguese Foundation for Science and Technology (FCT, https://ror.org/00snfqn58) Multi-Annual Financing Program for R&D Units..

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

AA Hotel and Hospitality Services (2024) Hotels quality standards. https://www.aahospitalityservices.co.uk/aa-quality-standards/hotels/

Arzaghi M, Genc IH and Naik S (2023) Rating vs. reviews: does official rating capture what is important to customers. *Heliyon* 9: e16337.

Bishop C (2006) *Pattern Recognition and Machine Learning*. Springer.

Blomberg-Nygard A and Anderson CK (2016) United nations world tourism organization study on online guest reviews and hotel classification systems: an integrated approach. *Service Science* 8(2): 139–151.

Cain LN, Mistry TG, Douglas S, et al. (2024) Luxury hotel technology trends: a multigenerational analysis. *International Journal of Contemporary Hospitality Management* 36(12): 4031–4054.

Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, et al. (2020) A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* 408: 189–215.

Chaves MS, Gomes R and Pedron C (2012) Analysing reviews in the web 2.0: small and medium hotels in Portugal. *Tourism Management* 33(5): 1286–1287.

Chu RKS and Choi T (2000) An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: a comparison of business and leisure travellers. *Tourism Management* 21(4): 363–377.

Claver E, Tari JJ and Pereira J (2006) Does quality impact on hotel performance? *International Journal of Contemporary Hospitality Management* 18(4): 350–358.

Cser K and Ohuchi A (2008) World practices of hotel classification systems. *Asia Pacific Journal of Tourism Research* 13(4): 379–398.

Forbes (2023) Which jumeirah hotel to choose in Dubai? Retrieved from. https://www.forbes.com/sites/michelerobson/2023/01/20/which-jumeirah-hotel-to-choose-in-dubai/

Guide M (2024) Everything you need to know about the MICHELIN key for hotels. Retrieved from. https://guide.michelin.com/us/en/article/travel/everything-to-know-about-the-michelin-keys-hotels-announcement

Gupta S and Kim HW (2009) Value-driven internet shopping: the mental accounting theory perspective. *Psychology and Marketing* 27(1): 13–35.

Hensens W (2015) The future of hotel rating. *Journal of Tourism Futures* 1(1): 69–73.

Hensens W, Struwing M and Dayan O (2010) Guest-review criteria on TripAdvisor compared to conventional hotel-rating systems to assess hotel quality. In: *Eurochrie Annual Conference*. https://dspace.nmmu.ac.za:8080/jspui/bitstream/10948/1631/4/ArticleEuroChrie2010.pdf

Ikotun AM, Ezugwu AE, Abualigah L, et al. (2023) K-means clustering algorithms: a comprehensive review, variants

analysis, and advances in the era of big data. *Information Sciences* 622: 178–210.

Jumeirah Group (2024) Jumeirah marsa Al Arab: redefining ultra-luxury in Dubai. Retrieved from. https://www.jumeirah.com/en/stay/dubai/jumeirah-marsa-al-arab

Khan MH, Hussain A and Khan MA (2022) A review on comparison of hotels star rating systems. *Journal of Managerial Sciences* 16(2): 144–155. https://journals.qurtuba.edu.pk/ojs/index.php/jms/article/view/628/219

Kim D, Hong S, Park BJ, et al. (2020) Understanding heterogeneous preferences of hotel choice attributes: do customer segments matter? *Journal of Hospitality and Tourism Management* 45: 330–337.

Kim JM, Ma H and Park S (2022) Systematic differences in online reviews of hotel services between business and leisure travelers. *Journal of Vacation Marketing* 29(2): 189–205.

Koutoulas D and Vagena A (2023) The present and future of hotel star ratings through the eyes of star rating operators. *Journal of Tourism Futures*, ahead-of-print(ahead-of-print).

Krey N, Wu S, Hsiao S-H, et al. (2024) The joint effect of online reviews and manager responses in driving company ratings. *Tourism Management Perspectives* 50: 101215.

Li C, Cui G and Peng L (2017) The signalling effect of management response in engaging customers: a study of the hotel industry. *Tourism Management* 62: 42–53.

Li H, Liu Y, Tan CW, et al. (2020) Comprehending customer satisfaction with hotels: data analysis of consumer-generated reviews. *International Journal of Contemporary Hospitality Management* 32(5): 1713–1735.

Liu X, Zhang Z, Law R, et al. (2024) Words meet photos: how visual content impact rating. *International Journal of Hospitality Management* 123: 103945.

Lockyer T (2002) Business guests' accommodation selection: the view from both sides. *International Journal of Contemporary Hospitality Management* 14(6): 294–300.

Lopes AI, Malthouse EC, Dens N, et al. (2024) Is webcare good for business? A study of the effect of managerial response strategies to online reviews on hotel bookings. *Journal of Service Management* 35(6): 22–41.

Louppe G, Wehenkel L, Sutera A, et al. (2013) Understanding variable importances in forests of randomized trees. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'13)*. Red Hook, NY, USA: Curran Associates Inc, Vol. 1, 431–439.

Maravić MU (2017) *Accomodation Classification System in. Slovenia* 23(2): 235–249. Available at: https://doi.org/10.20867/thm.23.2.1

Martin-Fuentes E (2016) Are guests of the same opinion as the hotel star-rate classification system? *Journal of Hospitality and Tourism Management* 29: 126–134.

Martin-Fuentes E, Fernandez C, Mateu C, et al. (2018) Modelling a grading scheme for peer-to-peer accommodation: stars for Airbnb. *International Journal of Hospitality Management* 69: 75–83.

Masiero L, Heo CY and Pan B (2015) Determining guests' willingness to pay for hotel room attributes with a discrete choice model. *International Journal of Hospitality Management* 49: 117–124.

Masiero L, Yang Y and Qiu RTR (2019) Understanding hotel location preference of customers: comparing random utility and random regret decision rules. *Tourism Management* 73: 83–93.

Mienye ID and Sun Y (2022) A survey of ensemble learning: concepts, algorithms, applications, and prospects. *IEEE Access* 10: 99129–99149.

Minazzi R (2010) Hotel classification systems: a comparison of international case studies. *Acta Universitatis Danubius OEconomica* 6(4): 65–88. https://journals.univ-danubius.ro/index.php/oeconomica/article/view/761/693

Nalley ME, Park J-Y and Bufquin D (2019) An investigation of AAA diamond rating changes on hotel performance. *International Journal of Hospitality Management* 77: 365–374.

Narangajavana Y and Hu B (2008) The relationship between the hotel rating system, service quality improvement, and hotel performance changes: a canonical analysis of hotels in Thailand. *Journal of Quality Assurance in Hospitality & Tourism* 9(1): 34–56.

Natekin A and Knoll A (2013) Gradient boosting machines, a tutorial. *Front. Neurorobot* 7(21): 21.

Nilashi M, Abumalloh RA, Alrizq M, et al. (2022) What is the impact of eWOM in social network sites on travel decision-making during the COVID-19 outbreak? A two-stage methodology. *Telematics and Informatics* 69: 101795.

Núñez-Serrano JA, Turrión J and Velázquez FJ (2014) Are stars a good indicator of hotel quality? Assymetric information and regulatory heterogeneity in Spain. *Tourism Management* 42: 77–87.

Nunkoo R, Teeroovengadum V, Ringle CM, et al. (2020) Service quality and customer satisfaction: the moderating effects of hotel star rating. *International Journal of Hospitality Management* 91: 102414.

Pennington-Gray L and Lee S (2024) The demand for safety measures in the hospitality industry: changes over three phases of a pandemic. *International Journal of Hospitality Management* 122: 103873.

Pestana MH, Gageiro M, Santos JAC, et al. (2024) Network structure of online customer reviews and online hotel reviews: a systematic literature review. *Information* 15(6): 334.

Rhee HT and Yang SB (2015) Does hotel attribute importance differ by hotel? Focusing on hotel star-classifications and customers' overall ratings. *Computers in Human Behavior* 50: 576–587.

Rita P, Ramos R, Borges-Tiago MT, et al. (2022) Impact of the rating system on sentiment and tone of voice: a booking.com and TripAdvisor comparison study. *International Journal of Hospitality Management* 104: 103245.

Sikdar S, Hooker G and Kadiyali V (2025) Variable importance measures for multivariate random forests. *Journal of Data Science* 23(1): 243–263.

Soifer I, Choi EK and Lee E (2020) Do hotel attributes and amenities affect online user ratings differently across hotel star ratings? *Journal of Quality Assurance in Hospitality & Tourism* 22(5): 539–560.

Steinley D and Brusco MJ (2007) Initializing K-means batch clustering: a critical evaluation of several techniques. *Journal of Classification* 24(1): 99–121.

Tibshirani R, Walther G and Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society - Series B: Statistical Methodology* 63(2): 411–423.

Tiwari V and Mishra A (2023) The effect of a hotel's star-rating-based expectations of safety from the pandemic on during-stay experiences. *Journal of Retailing and Consumer Services* 71: 103223.

Tiwari V and Omar A (2023) The impact of the hotel star rating system on tourists' health safety and risk perceptions: study based on tourists' vacation experiences. *Journal of Vacation Marketing* 31(1): 157–173.

Tuomi A (2021) Deepfake consumer reviews in tourism: preliminary findings. *Annals of Tourism Research Empirical Insights* 2(2): 100027.

UNWTO (2014) *Online Guest Reviews and Hotel Classification Systems – an Integrated Approach*. Madrid: UNWTO.

UNWTO (2015) *Hotel Classification Systems: Recurrence of Criteria in 4 and 5 Stars Hotels*. Madrid: UNWTO.

Vagena A and Manoussakis G (2021) Group analysis of official hotel classification systems: a recent study. In: *Current Approaches in Science and Technology Research 4*, 1–17.

Vallen GK and Vallen JJ (2017) *Check-In Check-Out. Managing Hotel Operations*. 10th edition. Pearson.

Vine P (1981) Hotel classification – art or science? *International Journal of Tourism Management* 2(1): 18–29.

Yang Y, Tang J, Luo H, et al. (2015) Hotel location evaluation: a combination of machine learning tools and web GIS. *International Journal of Hospitality Management* 47: 14–24.

Yang Y, Park S and Hu X (2018) Electronic word of mouth and hotel performance: a meta-analysis. *Tourism Management* 67: 248–260.

Zaman M, Botti L and Thanh TV (2016) Weight of criteria in hotel selection: an empirical illustration based on TripAdvisor criteria. *European Journal of Tourism Research* 13: 132–138.

Author biographies

Ana Lúcia de Pereira Neves Messias is currently an Assistant Professor at the Faculty of Medicine of the University of Coimbra and an Integrated Member of the Center for Mechanical Engineering, Materials, and Processes. She completed her Bachelor's degree (pre-Bologna) in Dentistry and the process for equivalence to an Integrated Master's degree in Dentistry from the Faculty of Medicine of the University of Coimbra in 2008 and 2010, respectively. She obtained her PhD in Health Sciences - Dental Medicine in 2019 from the Faculty of Medicine of the University of Coimbra. She has published more than 50 articles in journals indexed in the Web of Science Core Collection, resulting from national and international collaboration with more than 125 authors. Approximately 40% of her published works are among the 25% most cited documents, demonstrating a strong impact on the scientific community. She also has 5 chapters and 1 registered patent(s). She regularly participates in scientific events, having presented over 90 papers and/or posters, resulting in approximately 70 abstracts published in indexed journals. She has edited one book and served as an editorial staff member for five indexed scientific journals. She also serves as a referee for over 20 indexed journals. She has supervised or co-supervised 19 master's dissertations in Dentistry, Biomedical Engineering, and Mechanical Engineering, and two doctoral projects in Health Sciences and Data Science. She has received four international awards and eight awards at scientific events. She has received two pedagogical distinctions. She is or has participated as a doctoral fellow in one project and as a researcher in 11 projects. Works in the area(s) of Medical and Health Sciences with an emphasis on Clinical Medicine (Dentistry) and Engineering Sciences and Technologies with an emphasis on Mechanical Engineering (Biomechanics) and Data Science (Statistics).

Gilberto Rosa, Master's student in Clinical Bioinformatics at the University of Aveiro.

Pedro Couto holds a master's degree in Medical Statistics from Aveiro University, with an interest in scale development and validation and public health.

Vítor Rodrigues holds a Bachelor's degree in Biomedical Sciences and a Master's degree in Medical Statistics from the University of Aveiro. His work and interests include clinical trials, biostatistics, clinical decision support, and the application of machine learning in healthcare.

Luís M. Silva holds a bachelor's degree in Mathematics from the Faculty of Sciences of the University of Porto, Portugal, a master's degree in Statistics from the same faculty, and a PhD in Engineering Sciences from the Faculty of Engineering of the University of Porto, specializing in machine learning. He is currently an Assistant Professor in the Department of Mathematics and a researcher in the Probability and Statistics Group

at CIDMA (Center for Research and Development in Mathematics and Applications of the University of Aveiro), also collaborating on the Biomathematics (BioMath) thematic line at the same center. His research interests focus on probability and statistics, particularly machine learning, with a recent emphasis on applications in areas such as health sciences, life sciences, human resources, and tourism.

Jorge Marques is an Associate Professor in the Department of Tourism, Heritage and Culture of the Universidade Portucalense, in Porto, Portugal. He is also a researcher at REMIT-Research in Economics, Management and Information Technologies (University Portucalense) and CEGOT-Centre of Studies on Geography and Spatial Planning (University of Coimbra). He holds a bachelor 's degree in Hotel Management and PhD in Tourism, Leisure and Culture, with specialization in Tourism and Development. His current research interests cover several areas, such as tourism development policies, destination management and planning, tourism and hospitality management, business tourism, and hotel management.