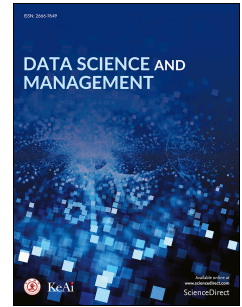


# Journal Pre-proof

The relationship between attribute performance and customer satisfaction: An interpretable machine learning approach

Jie Wang, Jing Wu, Shaolong Sun, Shouyang Wang



PII: S2666-7649(24)00003-1

DOI: <https://doi.org/10.1016/j.dsm.2024.01.003>

Reference: DSM 88

To appear in: *Data Science and Management*

Received Date: 23 September 2023

Revised Date: 5 January 2024

Accepted Date: 10 January 2024

Please cite this article as: Wang, J., Wu, J., Sun, S., Wang, S., The relationship between attribute performance and customer satisfaction: An interpretable machine learning approach, *Data Science and Management* (2024), doi: <https://doi.org/10.1016/j.dsm.2024.01.003>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Xi'an Jiaotong University. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd.

# The relationship between attribute performance and customer satisfaction: an interpretable machine learning approach

Jie Wang<sup>a</sup>, Jing Wu<sup>a</sup>, Shaolong Sun<sup>a\*</sup>, Shouyang Wang<sup>b,c,d</sup>

<sup>a</sup>School of Management, Xi'an Jiaotong University, Xi'an 710049, China

<sup>b</sup>Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

<sup>c</sup>School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

<sup>d</sup>Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China

\*Corresponding author. School of Management, Xi'an Jiaotong University, Xi'an, 710049, China. Tel: +86 15911059725; Fax: +86 29 82668382.

E-mail address: [s1120825@stu.xjtu.edu.cn](mailto:s1120825@stu.xjtu.edu.cn) (J. Wang)

[wujingsom@stu.xjtu.edu.cn](mailto:wujingsom@stu.xjtu.edu.cn) (J. Wu)

[sunshaolong@xjtu.edu.cn](mailto:sunshaolong@xjtu.edu.cn) (S. L. Sun\*)

[sywang@amss.ac.cn](mailto:sywang@amss.ac.cn) (S. Y. Wang)

## Acknowledgments

This research was partially supported by the National Key R&D Program of China (Grant No.: 2022YFF0903000) and the National Natural Science Foundation of China (Grant Nos. 72101197 and 71988101).

# The relationship between attribute performance and customer satisfaction: an interpretable machine learning approach

**Abstract:** Understanding the relationship between attribute performance (AP) and customer satisfaction (CS) is crucial for the hospitality industry. However, accurately modeling this relationship remains challenging. To address this issue, we propose an interpretable machine learning-based dynamic asymmetric analysis (IML-DAA) approach that leverages interpretable machine learning (IML) to improve traditional relationship analysis methods. The IML-DAA employs extreme gradient boosting (XGBoost) and SHapley Additive exPlanations (SHAP) to construct relationships and explain the significance of each attribute. Following this, an improved version of penalty-reward contrast analysis (PRCA) is used to classify attributes, whereas asymmetric impact-performance analysis (AIPA) is employed to determine the attribute improvement priority order. A total of 29,724 user ratings in New York City collected from TripAdvisor were investigated. The results suggest that IML-DAA can effectively capture non-linear relationships and that there is a dynamic asymmetric effect between AP and CS, as identified by the dynamic AIPA (DAIPA) model. This study enhances our understanding of the relationship between AP and CS and contributes to the literature on the hotel service industry.

**Keywords:** Hotel service; AP-CS relationship; Interpretable machine learning; Dynamic asymmetric analysis; XGBoost

## 1. Introduction

Customer satisfaction (CS) is crucial to the success of all companies and organizations in the service industry, and analyzing the relationship between CS and attribute performance (AP) is important for enhancing competitiveness. Consequently, this topic has received extensive attention recently (Berezina et al., 2016; Bi et al., 2020; Chen, 2014; Davras and Caber, 2019). Most studies have employed a multi-attribute approach to evaluate the relationship between AP and CS, meaning that different attributes contribute differently to CS (Ji et al., 2023). Based on the multi-attribute approach, scholars have proposed methods and theories, such as the Kano model, importance-performance analysis (IPA) (Chen, 2014), penalty-reward contrast analysis (PRCA) (Albayrak and Caber, 2013a; Bi et al., 2019a), and asymmetric impact-performance analysis (AIPA) (Bi et al., 2020; Li et al., 2020), to help managers improve service attributes and enhance CS. Initially, the relationship between AP and CS was conceptualized as linear or symmetrical in most CS studies; that is, equal changes in positive and negative AP result in equal changes in CS (Chen et al., 2015; Liu et al., 2017). However, some studies have found that the equal changes in positive and negative AP lead to different amounts of change in CS. The relationship between AP and CS may be non-linear or asymmetrical. Thus, research on the AP-CS relationship has gradually evolved from symmetric to asymmetric perspectives (Albayrak and Caber, 2013b; Caber et al., 2013).

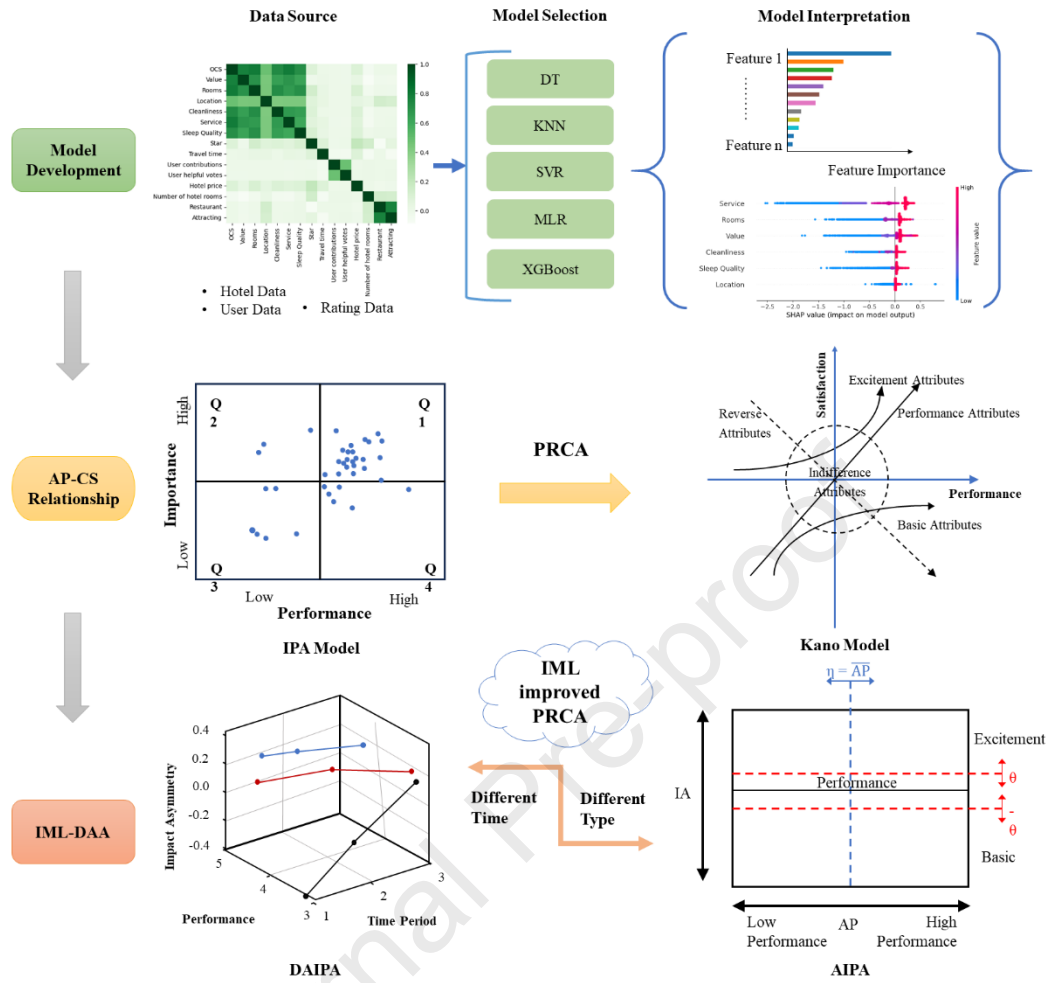
Three-factor theory (Conklin et al., 2004) divides attributes with distinct asymmetric associations into three categories: excitement, basic, and performance. Studying the asymmetric relationship between AP and CS and identifying the categories of each attribute are important for prioritizing hotel attributes for improvement (Joung and Kim, 2022). Various studies have used a static viewpoint to explain the asymmetrical relationship between AP and CS. However, this relationship is dynamic; that is, customers' expectations and requirements change over time as their knowledge increases and when certain events occur (Rita et al., 2022; Voss et al., 1998). CS depends on the variance between customers' expectations and perceived performance, as described in the expectation-disconfirmation paradigm (McKinney et al., 2002). Thus, as technology evolves, word of mouth spreads, and alternative experiences emerge, an attribute will become the market standard (File et al., 1994), and some attributes will no longer show unlimited benefits and even negatively affect CS when their performance levels fall below customer expectations. The symmetry effect evolves into a negative asymmetry effect in accordance with the Kano model, in which an attribute changes from a performance item to a basic one. Thus, this study aims to examine the dynamic asymmetric effect of the AP-CS relationship.

With the rapid growth of social media and online presence of online travel agencies (OTAs), an increasing number of users are returning to OTA platforms to post comprehensive reviews and ratings of their experiences, a form of user-generated data (UGD). UGD has characteristics that traditional questionnaire data does not, such as objectivity, immensity, no sample bias, and real-time updates (Schuckert et al., 2015). As customers accumulate and the number of reviews increases, a large amount of UGD is generated on online platforms, creating a large, fast-moving dataset that appears to be a promising source for studying hotel marketing (Ji et al., 2023b; Yi and Oh, 2022). With UGD, researchers can use new technologies to effectively examine consumer feedback from a wide range of people within a short period. Therefore, our study uses user-generated ratings from the TripAdvisor website to conduct the research, where overall customer satisfaction (OCS) and CS ratings for six related service attributes ("location," "cleanliness," "room," "service," "sleep quality" and "value") are available. With the development of artificial intelligence (AI) and computer technology, many machine learning models, such as k-nearest neighbors (KNN) (Cunningham and Delany, 2021), decision trees (DT) (Quinlan, 1986), neural networks (Jain et al., 1996), and support vector regression (SVR) (Dong et al., 2015), have been used in various fields (Li et al., 2023; Liu et al., 2020). However, research on the asymmetric relationship between AP and CS using machine learning remains limited, especially in the hospitality industry.

The widely used PRCA method is mainly employed to classify attributes using multiple regressions (Matzler and Sauerwein, 2002; Radojevic et al., 2018). However, studying the asymmetric effects of AP and CS using multiple regression has two limitations: (1) the relationship between the independent and dependent variables is linear (i.e., equal changes in the performance of the attributes will cause equal changes in consumer satisfaction). Although some later studies have split the performance of the attributes into two categories, this cannot overcome the disadvantage that regression models can only construct linear relationships; and (2) there is no information interaction between the attributes (i.e., the effect of "service" attribute on CS is the same, regardless of whether the customer gives a five-star rating or a one-star rating for "cleanliness."

To remedy the shortcomings of multiple regression and explore the dynamic asymmetry between the AP and CS relationships, we constructed the interpretable machine learning dynamic

asymmetric analysis (IML-DAA) approach. First, we introduced four machine learning models, namely, DT, KNN, SVR, and XGBoost, and compared their predictions of hotel CS with those of multiple linear regression (MLR), ultimately finding that XGBoost performs the best. XGBoost is a relatively new machine learning approach that is less computationally expensive and less complex while maintaining high accuracy and speed (T. Chen and Guestrin, 2016). Few studies have used the XGBoost model to examine CS in the hospitality industry. Machine learning models such as XGBoost are often referred to as “black boxes” because we only know what the inputs and outputs of the model are but have no way of understanding why the outputs are what they are. In other words, there is a tradeoff between the predictability and interpretability of the model. Based on this, we decided to use IML with SHapley Additive exPlanations (SHAP) to model the interpretation and measure the marginal contribution of each feature value to the final prediction (Ribeiro et al., 2016; Strumbelj and Kononenko, 2014). Originally proposed by Shapley in 1953, SHAP is based on game theory (Shapley, 1953) and provides a powerful and insightful measure of the importance of features in a model (Lundberg and Lee, 2017). Second, based on the results of the SHAP framework for interpreting the constructed model, we used the IML-improved PRCA model to calculate the impact asymmetry (IA) values. Next, because different attributes behave differently at different stages, we used dynamic asymmetric analysis (DAA) to analyze the AP-CS relationship. We propose a dynamic asymmetric impact-performance analysis (DAIPA) model to monitor the dynamic asymmetry of the AP-CS relationship and use the AIPA model for attribute categorization, focusing specifically on attribute quadrant changes during COVID-19.



**Fig. 1.** Framework of this study design and model development.

Note: Decision Tree (DT), k-Nearest Neighbour (KNN), Multiple linear regression (MLR), Support Vector Regression (SVR), XGBoost, eXtreme Gradient Boosting (XGBoost), Attribute Performance (AP), Customer satisfaction (CS), Impact Asymmetry (IA), Penalty-Reward Comparison Analysis (PRCA), Asymmetric Impact Performance Analysis (AIPA), Dynamic Asymmetric Impact Performance Analysis (DAIPA), SHapley Additive exPlanations (SHAP).

In summary, this study aimed to investigate the relationship between AP and CS using the IML approach. OCS was predicted using the XGBoost model to compare the predictive efficacy of different machine learning models. SHAP analysis was employed to detect the importance of individual service features and construct a consumer satisfaction function. Improved PRCA and AIPA were utilized to identify attribute categories and obtain improved orders. The data for this study were sourced from the TripAdvisor website using user rating data to construct the dataset. As time evolves, customer awareness increases, or major public health events such as the coronavirus disease (COVID-19) pandemic, customer requirements, and expectations for different service attributes change accordingly. DAIPA was used to identify the dynamic asymmetry of the AP-CS relationship between the periods. The research framework for this study, based on the theoretical analysis and empirical data gathered, is presented in **Fig. 1**.

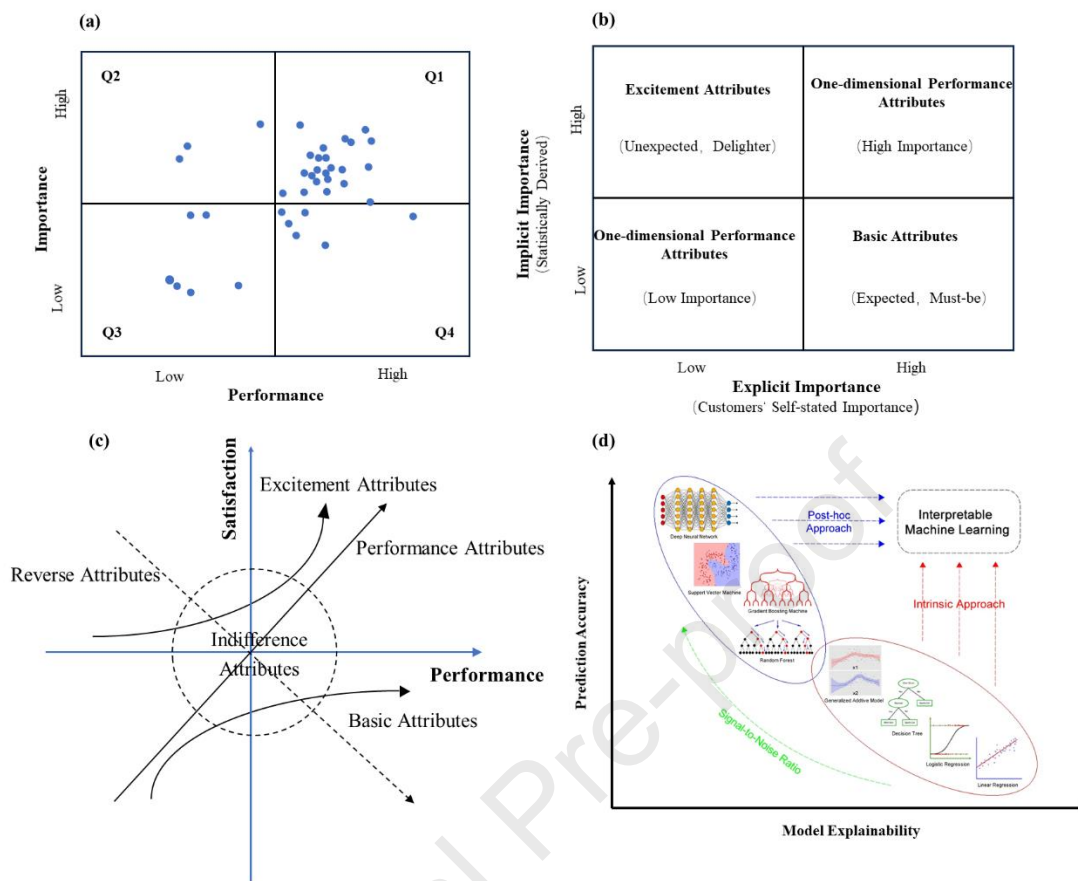
Different attributes contribute differently to OCS and are influenced by perceived performance,

customer expectations, and asymmetric relationships. Using data from 29,724 reviews posted by users on TripAdvisor covering 423 hotels in New York City, we explored the dynamic asymmetric effects between hotel AP and CS using the IML-DAA approach. This involved distinguishing between different types of attributes and capturing their changes at different times of consumption, which allowed us to gain valuable knowledge that can assist hotel managers in their decision-making and evaluation processes. This study's contributions are as follows. First, a hotel's improvement priorities for different attributes can be understood by distinguishing between them. For example, some attributes should be allocated more resources if they perform poorly, and users care more about them. Second, we can capture attribute type changes due to different consumption times (e.g., more attention should be paid to managing hygiene aspects of hotels during epidemics), which can help hotel managers make future decisions and evaluations when facing major events. Finally, based on the constructed AP-CS relationship, we clarify what measures can be taken to improve customer satisfaction, enhance hotel competitiveness, and increase hotel revenue.

The remainder of this paper is organized as follows. Section 2 summarizes the related work. Section 3 describes our approach, and Section 4 presents and discusses the results. Finally, Section 5 analyzes the implications and limitations of this study.

## 2. Related work

Section 2 presents the related work. Section 2.1 summarizes the study about the relationship between AP and CS, Section 2.2 describes studies on interpretable machine learning (IML), and **Fig. 2** shows related studies and methods.



**Fig. 2.** Models of AP-CS relationship and IML method.

Note: (a) IPA diagram. Quadrant 1 (Q1) is labeled “Keep up the good work.” Quadrant 2 (Q2) is known as “concentrated here,” the features described in Q2 can be deemed the product/service’s fundamental flaws. Q3 is referred to as the “low priority” quadrant. Q3 qualities have lower performance and importance. Quadrant 4 (Q4) is designated as the “Possible Overkill” quadrant as its listed attributes exhibit high performance but are of low importance. Consequently, the quality of Q4 may waste scarce resources. (b) Source: Vavra (1997). The importance grid model places each attribute in a matrix with its explicit and implicit importance values on the x and y axes, respectively. Q4: Must-be factors; Q2: Delighter elements; Q1: Crucial performance drivers. Q3: Unimportant performance variables. (c) The Kano model categorizes product attributes into five types: basic, performance, excitement, indifference, and reverse. (d) Copyright: The University of Hong Kong Zhang’s Group. The horizontal axis represents model interpretability, with a more positive direction representing higher model interpretability. The vertical axis represents model accuracy, and a more positive direction represents higher model accuracy. This figure shows the relationship between models’ accuracy and explainability. The IML method has high accuracy and interpretability.

## 2.1 Correlational study of the relationship between attribute performance (AP) and customer satisfaction (CS)

The relationship between AP and CS has been extensively investigated in the marketing



literature. **Table 1** summarizes the main models used to investigate the relationship, including IPA, Vavra's importance grid model (Matzler and Sauerwein, 2002), the Kano model, impact asymmetry analysis (IAA) (Mikulić and Prebežac, 2012), AIPA (Albayrak and Caber, 2013a) and impact range performance analysis (IRPA) (Lee and Min, 2013). Many studies use one or more of the six dominant models to explore the relationship between service attributes and CS in tourism literature.

Expectation-disconfirmation (Oliver et al., 1997) assumes that the relationship between AP and CS is either linear or symmetric. IPA, also known as “action grid analysis” (Martilla and James, 1977), is a common symmetry technique. IPA divides the attributes of a product or service into four quadrants or categories by splitting each of the two dimensions (AP and importance) into two levels. An example IPA plot is shown in **Fig. 2 (a)**. The importance grid model displays the explicit and implicit importance of the service attributes in a two-dimensional grid. Vavra (1997) argued that the importance of a product/service attribute can vary considerably, depending on whether its measurement is explicit (e.g., customer self-reported importance values obtained through questionnaires) or implicit (e.g., values obtained through regression analysis, partial correlation, or AI methods). Deng et al. (2008) presented a backpropagation neural network (BPNN)-based IPA model that determines the importance of attributes by training the BPNN with natural logarithmic AP and OCS as the input and output variables, respectively. Importance grid models are widely used in the tourism, marketing, and e-commerce literature (Albayrak et al., 2016; Mathe-Soulek et al., 2015). As studies about AP and CS progressed, other extended variants of IPA were proposed. Albayra and Caber (2015) created an important performance competitor analysis (IPCA) model taking rivals' performance into account.

Nevertheless, the above researchers conceptualized the relationship between AP and CS as either linear or symmetrical; that is, equal changes in positive and negative APs will result in equal changes in CS. However, the existing research suggests that AP may influence overall satisfaction asymmetrically. Specifically, the same degree of variation in a given attribute's positive and negative performance may affect the overall satisfaction differently (Caber et al., 2013). Additionally, product/service attributes can be categorized into excitement, base, and performance. Kano et al. (1984) observed that consumers have varied attitudes toward different product qualities and proposed a model that categorized product attributes into different types (see **Fig. 2 (c)**). However, the original Kano model is qualitative and does not adequately measure CS. Several scholars have quantitatively extended the Kano model (Brandt, 1988; Matzler et al., 1996). According to three-factor theory (Li et al., 2020), researchers often focus on three attribute types: excitement, performance, and basic.

**Table 1.** Study on the relationship between attribute performance and consumer satisfaction.

Models	Variables			Implications			References	Data source	Context
	attribute importance	Attribute performance	Attribute impact - asymmetry	Attribute priority	Linear relationship	Asymmetric relationship			
IPA	√	√	—	√	√	—	Jang, et al. (2009)	Questionnaire	Restaurant
Kano	—	—	√	—	—	√	Bi, et al. (2019a)	Online reviews	Hotel
							Davras and Caber (2019)	Questionnaire	Hotel
							Zhang, et al. (2023)	Online reviews	Retailing
							Back (2012)	Questionnaire	Restaurant
IRPA	√	√	—	√	√	—	Ju, et al. (2019)	Online reviews	Platform
IAA	√	—	√	√	—	√	Lee, et al. (2017)	Questionnaire	Travel
							Wong and Lai (2018)	Questionnaire	Exhibition
							Mikulic, et al. (2016)	Questionnaire	Travel
Vavra	√	—	—	√	—	√	Albayrak, et al. (2016)	Questionnaire	Hotel
							Matzler and Sauerwein (2002)	Questionnaire	Computer
							Mathe-Soulek, et al. (2015)	Questionnaire	Restaurant
AIPA	—	√	√	—	—	√	Albayrak and Caber (2016)	Questionnaire	Destination
							Albayrak (2019)	Online reviews	Hotel
							Bi, et al. (2020)	Online reviews	Hotel

**(1) Excitement attribute:** This category of attributes leads to CS when expectations are exceeded, but does not result in dissatisfaction when expectations are not met. Therefore, its positive effect on CS is greater than its negative effect.

**(2) Performance attribute:** This type of attribute is closely linked to CS because its performance affects both satisfaction and dissatisfaction. Specifically, dissatisfaction and satisfaction occur when customer expectations are not met and fulfilled, respectively.

**(3) Basic attribute:** This attribute works in direct contrast to the excitement attribute but is often taken for granted. Customers remain neutral when their expectations are met but become highly dissatisfied when their expectations are not fulfilled; thus, this attribute's unsatisfactory performance has a greater impact on CS than its satisfactory performance.

Based on the above observations, Mikulić and Prebežac (2008) devised the IRPA and IAA models. The authors recommended that attributes with a greater range of impacts on CS should be prioritized to enhance CS. Following that, Caber et al. (2013) proposed the AIPA model to investigate the asymmetric impact of limited company resources (e.g., time, money, and human resources) on OCS. The above improved the IPA model and enhanced our understanding of the relationship between AP and CS. Hu et al. (2020) used AIPA to optimize hotel service offerings based on UGD and examined the asymmetric impact of service attributes on CS in the context of three hotel chains. Many studies have analyzed the non-linear asymmetric relationship between AP and CS (Bi et al., 2019b; Radojevic et al., 2018).

However, the asymmetric relationship between the AP and CS is dynamic. As technology evolves, word of mouth spreads, and alternative experiences emerge, the customer's expectations (reference points) and requirements will change over time. The types of attributes will also change. Consequently, this study seeks to examine the dynamic asymmetric effect of the AP-CS relationship. Bi et al. (2019b) developed the dynamic IPA model to examine the performance and importance trends of product/service attributes over time. However, the asymmetry between AP and CS has not yet been explored.

An increasing number of researchers have used UGD to conduct relevant research on the hospitality industry. For example, Nie et al. (2023) investigated the impact of segmentation and temporal dynamics caused by the COVID-19 epidemic on the classification of service quality attributes by analyzing 67,623, 76,730, and 17,507 reviews in three segments: id-scale, budget, and economy, respectively, to develop improvement strategies to meet customer needs and address threats. Zhang et al. (2021) employed an improved PRCA to quantify the emotional tendency and intensity of 3,777 four-season hotel online reviews, identify the types of service attributes, and prioritize hotel resource allocation, along with managers' subjective opinions. Previous research indicates that user-generated ratings are a valuable source of data for researchers seeking to understand consumer preferences and satisfaction (Chatterjee, 2019; Zhang et al., 2020). The abundance of user-generated ratings collected via the Internet is remarkable. In the hotel domain, different hotel and traveler types and travel periods can be covered. Thus, user-generated ratings

present a potential avenue for investigating the dynamic asymmetric effect of AP on hotel CS.

It is important to note that in the hotel industry, the asymmetric effect of AP on CS has been studied mainly from a static perspective for specific markets at specific times. As there is a notable lack of research on the asymmetric effect of AP on CS from a dynamic perspective, a thematic study on the asymmetric effect of AP on CS in hotels is needed.

## 2.2 Interpretable machine learning (IML)

The challenge of studying the AP-CS relationship lies in (1) identifying and constructing non-linear relationships between AP and OCS and (2) quantifying and explaining the extent to which different APs affect OCS. Although machine learning can capture these non-linear relationships, it cannot explain how satisfaction with product attributes affects OCS owing to its aforementioned black-box nature. Therefore, most previous studies have used multiple linear regressions (MLR) to identify attribute importance (Albayrak and Caber, 2013a; Bi et al., 2020), but MLR cannot capture non-linear relationships and do not consider the interactions between attributes. The method developed in this study uses IML to address these issues, as described in the following subsections.

IML is the process of extracting significant knowledge about relationships learned by machine learning models (Murdoch et al., 2019). Recent machine learning models, such as artificial neural networks (ANN), random forests (RF), light gradient boosters (LGBM), and XGBoost models, have demonstrated exceptional predictive performance in various data analyses. Gou et al. (2022) used XGBoost to predict CS using multiple data sources to determine the intrinsic relationship between CS and product experience. Gou's experiments demonstrated that the trained XGBoost model outperformed conventional linear regression models. Liu et al. (2020) found that the non-linear XGBoost model can better utilize user characteristics and achieve better prediction results. However, the prediction results lack interpretability. In other words, there is a tradeoff between the predictability and interpretability of the model, as shown in **Fig. 2 (d)**.

IML provides an easy-to-understand explanation. It not only provides the prediction value of the model but also gives the reason for obtaining the prediction value, thus achieving safety, transparency, and fairness. IML is divided into intrinsically explainable models and post-hoc explanatory approaches. Post-hoc interpretation methods are used to interpret machine learning models after they have been constructed and are advantageous in that they allow researchers the freedom to use multiple machine learning models (Covert et al., 2020). This study used the SHAP algorithm as a post-hoc interpretation approach. Based on game theory and local interpretation (Ribeiro et al., 2016), SHAP provides a way to estimate the contribution of each feature, and SHAP values have been proposed as a unified measure of feature importance. It can be applied to each review to analyze the non-linear relationship between AP and OCS (Lundberg and Lee, 2017). Moreover, it can be used to visualize the importance of features and determine the elements that have the greatest impact on the final model. After understanding SHAP's benefits, increasingly more researchers began to use this technique. Conklin et al. (2004) used cooperative game theory and risk

analysis tools to address the key drivers of CS associated with the Kano model. Liu (2023) constructed a profit-driven hotel order cancellation predictor based on grid search and XGBoost to solve the hotel order cancellation problem. The prediction results of the proposed XGBoost were interpreted using SHAP values, and the key factors affecting hotel order cancellations were analyzed. However, hotel CS prediction and analysis based on SHAP have not received much attention from researchers.

### 3. Methodology and practice

#### 3.1 Data collection

This study collected UGD from TripAdvisor (<https://www.tripadvisor.com/>). Many previous studies have also used data obtained from this website (Chang et al., 2022; Rita et al., 2022). We used a crawler to extract information and UGD (e.g., online reviews, overall ratings, multi-attribute ratings, dates of stay) for New York City hotels using the TripAdvisor. As of February 2023, we retrieved a total of 940,283 user reviews for 928 New York City hotels. The initial sample of online reviews was filtered in two ways. (1) We removed hotels with fewer than 100 user reviews, and (2) customer attribute ratings are optional; therefore, we excluded data with missing multi-attribute ratings. Finally, we obtained 297,244 valid data points for 423 hotels.

Rating data, including OCS and aspect-level attribute ratings (value, rooms, location, cleanliness, service, and sleep quality) were used in this study. OCS and CS on six pertinent criteria were evaluated using a 5-point Likert scale. Ratings ranged from 1 to 5, indicating “terrible,” “poor,” “average,” “very good,” and “excellent.” In addition, travel times were recorded. **Table 2** shows the summary statistics for the rating dataset as a whole and the travel time. As shown in **Table 2**, “location” scored the highest rating of all the criteria, indicating that location is a benefit for hotels in New York City. The average overall satisfaction of users also notably drops in 2021 and later steadily rebounds in 2023, which can be plausibly linked to the COVID-19 pandemic and its effect on tourism.

**Table 2.** Descriptive statistics of the data.

Year	Count	OCS		Value		Rooms		Location		Cleanliness		Service		Sleep quality	
		Mean	std.	Mean	std.	Mean	std.	Mean	std.	Mean	std.	Mean	std.	Mean	std.
2007	69	3.87	1.07	3.57	1.27	3.65	1.07	4.71	0.60	4.04	1.09	3.68	1.27	3.86	1.14
2008	124	4.16	0.83	3.84	1.00	3.98	0.95	4.69	0.58	4.20	0.91	3.99	0.98	4.10	0.91
2009	1,184	<b>4.12</b>	0.95	3.89	1.07	3.91	1.07	4.62	0.73	4.31	0.94	3.91	1.10	4.12	1.05
2010	16,999	4.20	1.00	4.04	1.09	4.08	1.05	4.60	0.72	4.45	0.89	4.16	1.12	4.19	1.07
2011	22,162	4.20	0.97	4.02	1.04	4.09	1.01	4.62	0.68	4.41	0.89	4.20	1.05	4.22	1.02
2012	44,799	4.21	0.94	4.04	1.00	4.12	0.97	4.65	0.64	4.41	0.86	4.26	0.99	4.27	0.95
2013	58,654	4.23	0.93	4.07	0.99	4.12	0.96	4.67	0.62	4.42	0.85	4.30	0.97	4.29	0.95
2014	32,946	4.24	0.96	4.11	1.01	4.15	0.98	4.67	0.62	4.44	0.86	4.35	0.97	4.30	0.96
2015	13,109	4.31	0.99	4.14	1.02	4.22	1.00	4.69	0.62	4.49	0.84	4.45	0.92	4.33	0.98
2016	20,891	4.29	1.02	4.12	1.07	4.22	1.03	4.70	0.62	4.48	0.88	4.42	0.97	4.34	1.00
2017	24,989	4.33	1.02	4.15	1.06	4.29	1.01	4.73	0.60	4.51	0.87	4.44	0.98	4.39	0.98
2018	17,276	4.28	1.08	4.12	1.12	4.24	1.06	4.72	0.62	4.49	0.92	4.41	1.04	4.36	1.02
2019	21,556	4.20	1.16	4.06	1.18	4.17	1.13	4.71	0.64	4.43	0.98	4.36	1.10	4.31	1.07
2020	3,628	4.35	1.09	4.33	1.07	4.30	1.07	4.73	0.65	4.53	0.93	4.47	1.02	4.39	1.01
2021	<b>5,594</b>	<b>3.97</b>	<b>1.40</b>	3.93	1.37	3.99	1.36	4.55	0.87	4.20	1.28	4.11	1.36	4.17	1.24
2022	10,917	4.08	1.31	3.94	1.31	4.07	1.26	4.66	0.73	4.34	1.13	4.22	1.25	4.25	1.16
2023	2,347	<b>4.26</b>	1.20	4.26	1.15	4.25	1.13	4.73	0.67	4.50	0.99	4.35	1.19	4.35	1.10
ALL	297,244	4.23	1.02	4.08	1.07	4.15	1.03	<b>4.67</b>	<b>0.65</b>	4.44	0.90	4.32	1.03	4.29	1.00

Note: std. means standard deviation.

## 3.2 Research method

### 3.2.1 Model selection and interpretation

Previous studies (Matzler et al., 1996; Caber et al., 2013) have demonstrated that the relationship between service AP and OCS is non-linear and counterintuitive. Therefore, in this study, this relationship was established using supervised machine learning methods, and OCS was predicted using a machine learning model. We investigated five regression models: SVR, DT, XGBoost, KNN, and MLR.

For each method, a regression model was implemented using the Scikit-Learn machine learning package in Python (Pedregosa et al., 2011). We performed the following steps to ensure reliability of the results.

- (1) We used 80% and 20% of the samples for training and testing, respectively.
- (2) We examined the optimized calculation of the hyperparameters defined for each machine learning approach (e.g., computation solver, learning rate, and tree depth) through a grid search with k-fold cross-validation (k=5). The parameter combinations are available in the data file provided in the **Appendix 2**.
- (3) We selected the parameter combination that yielded the most favorable outcome for the regression forecast for each machine learning method. The Appendix 2 material data file displays both the parameter combinations and their optimal configurations.
- (4) We used the explained variance (Evar), mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), root mean square error (RMSE), and R-squared as the evaluation criteria.

The specific formulas and meanings of each evaluation indicator are provided in the **Appendix 1**. These metrics are widely used in regression tasks to evaluate the performance and predictive ability of models, helping select the most appropriate model or optimize the model parameters. According to the modeling results in **Table 3**, XGBoost exhibits the best performance, which is consistent with the findings of many earlier studies (Liu et al., 2019, 2023). Consequently, we employed the XGBoost model for further analyses. However, there are other reasons for selecting this model for our study.

- (1) Highly interpretable: The XGBoost model ranks the importance of the features, allowing us to clearly understand which features contribute the most to the results.
- (2) Parameters are easy to adjust: The XGBoost model provides various parameter options to suit the specific characteristics and requirements of the data.
- (3) Wide range of applications: XGBoost can handle various types of data, including structured and unstructured data, and is more robust to outliers and noisy data.
- (4) Immunity from multicollinearity: Two variables can be retained even if they capture the same phenomenon in the system, which is particularly desirable because we perform important

characterizations through SHAP.

**Table 3.** Modeling results.

Model	Evar	MAE	MAPE	MSE	RMSE	$R^2$
DT	0.814	0.290	0.091	0.193	0.105	0.814
KNN	0.800	0.296	0.094	0.207	0.109	0.800
MLR	0.803	0.302	0.097	0.205	0.111	0.803
SVR	0.803	0.288	0.088	0.204	0.107	0.803
XGBoost	<b>0.824</b>	<b>0.286</b>	<b>0.088</b>	<b>0.182</b>	<b>0.101</b>	<b>0.824</b>

Note: Explained Variance (Evar), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Root Mean Square Error (RMSE), R-squared ( $R^2$ ).

XGBoost is a gradient-boosting tree (GBT) machine-learning algorithm that optimizes distributed gradient boosting by continuously adding trees (Chen and Guestrin, 2016; Friedman, 2001). Each DT predicts the error of the previous DT and improves the prediction of the previous step, thereby reducing the overall error (Gou et al., 2022).

Next, we trained an explanatory model to illustrate the impact of AP on CS. All trained data were used in XGBoost as inputs to train the SHAP explanation model. SHAP provides the contribution of each feature to the model output, which enables us to better understand the prediction process of the model. SHAP has several advantages, as follows:

- (1) High stability: Compared with traditional feature importance ranking methods, SHAP does not suffer from bias, owing to factors (e.g., data distribution or feature relevance) and thus has high stability and reliability.
- (2) Wide applicability: SHAP is applicable not only to classification problems but also to regression problems and deep learning models.
- (3) Good visualization: SHAP provides a variety of visualization tools that can intuitively illustrate the importance and contribution of features, thus helping us to better understand the prediction results of the model.
- (4) Satisfying cooperativity: The total SHAP value of the features is equal to the difference between the model predictions and the benchmark value. The benchmark value is the average prediction result of the model's overall features.

### 3.2.2 Asymmetric effect of attribute performance on customer satisfaction

Brandt's (1988) PRCA is a popular method for investigating the asymmetric effects of AP on CS, and its results can group the characteristics into distinct classes. In previous studies, this was typically accomplished by creating two dummy variables, denoted as  $d_{lp}^i$  and  $d_{hp}^i$ , where the first dummy variable is used to estimate the effect of low AP on CS (Mikulić and Prebežac, 2008), and the second is used to estimate the effect of high AP on CS;  $n = 6$  (i.e., the six attributes). Multiple regression analysis was performed using the two dummy variables obtained to assess the effect of



each feature on CS at very low and very high-performance levels:

$$OCS = \beta_0 + \sum_{i=1}^n (\beta_{\text{penalty}}^i a_{\text{low}}^i + \beta_{\text{reward}}^i a_{\text{high}}^i) + \varepsilon \quad (1)$$

Additionally, neither the interaction between characteristics nor the non-linear relationship between AP and OCS are modeled by the MLR model. Therefore, this study improves PRCA based on the interpretation results of high-precision machine learning models that can model non-linear relationships. In this improved PRCA model. We used Shapley values, rather than regression model coefficients, to measure the penalty and reward coefficients.

$$\begin{aligned} \beta_{\text{low}}^i &= \text{mean}(\sum_{k \in \text{low}} SHAP_{\text{low}}^{ik}) - \text{mean}(\sum_{k \in \text{medium}} SHAP_{\text{medium}}^{ik}) \\ \beta_{\text{high}}^i &= \text{mean}(\sum_{k \in \text{high}} SHAP_{\text{high}}^{ik}) - \text{mean}(\sum_{k \in \text{medium}} SHAP_{\text{medium}}^{ik}) \end{aligned} \quad (2)$$

In our model, an attribute is considered “low” when it scores one, “medium” when it scores three, and “high” when it scores five. Using classification techniques (e.g., index values), qualities can be classified into three groups: basic, performance, and excitement. The most common index value is “impact asymmetry” (IA) (Mikulić and Prebežac, 2008). On the one hand, the index value of IA spans from -1 to +1, making it easier to compare calculated indices. Furthermore, AIPA which applied to select hotel qualities including the IA index. Consequently, we used the IA index to categorize the attributes into separate groups. Based on the obtained  $\beta_{\text{low}}^i$  and  $\beta_{\text{high}}^i$ , the IA index can be calculated by **Equation (3)**.

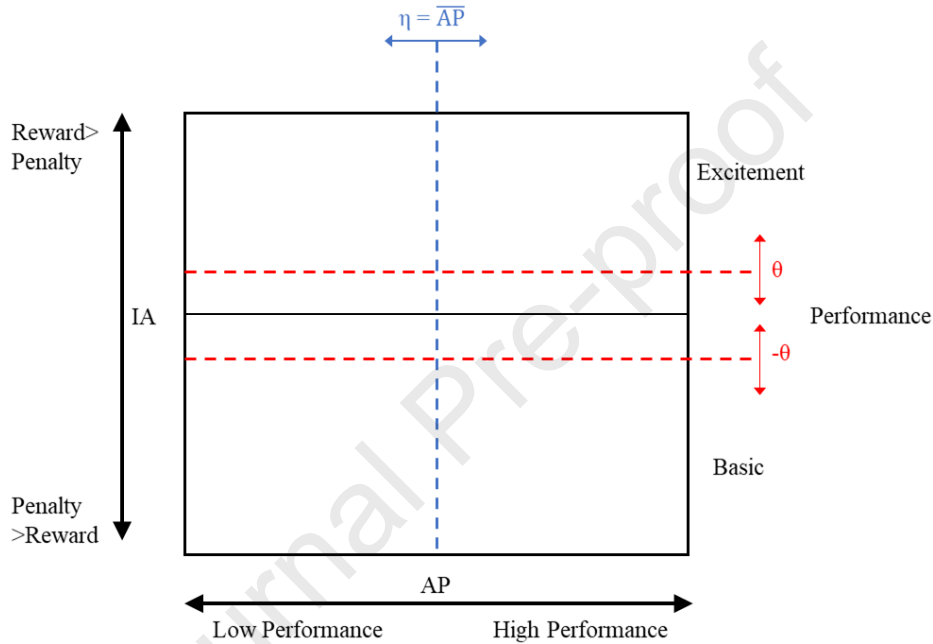
$$IA_i = \frac{|\beta_{\text{high}}^i| - |\beta_{\text{low}}^i|}{|\beta_{\text{low}}^i| + |\beta_{\text{high}}^i|} \quad (3)$$

The smaller the IA, the more likely it is that the attribute will lead to customer dissatisfaction, and the larger the IA, the more likely it is that the attribute will lead to CS. According to Albayrak and Caber (2015) and Mikuli and Prebeac (2008), a cut-off point should be set subjectively to divide attributes into distinct categories. After the specific analysis of the data in this work, we define  $\theta = 0.2$ . In this paper, this value is selected primarily for two reasons. First, previous studies often employed  $\theta = 0.1$ , but this approach overlooks the distribution characteristics of the IA value, leading to the majority of attributes being recognized as basic or excitement attributes and poor classification. Second, to address this issue, we must adjust the value to reflect the distribution characteristics. In this study, the majority of the IA values obtained through the improved PRCA method, which is based on the IML, were positioned close to 0.1 (see **Appendix 4**). Nonetheless, such attributes cannot be correctly classified if 0.1 is taken as the cut-off point. By selecting 0.2 as the cut-off point, observations of the dynamic changes in attribute categories over time are more precise.

### 3.2.3 Dynamic asymmetric analysis

We used the average user ratings to evaluate each attribute’s performance and the method

described in Section 3.2.2 to calculate IA. Based on the obtained  $AP_i$  and  $IA_i$ , we categorized the attributes into different types. In this study, two types of DAA were synthesized: AIPA and DAIPA. DAA aims to explain and investigate the dynamic asymmetry of relationships. AIPA is a method for understanding the asymmetry of the AP-CS relationship and for developing product/service improvement strategies (Albayrak and Caber, 2015; Caber et al., 2013). The horizontal axis of the AIPA graph represents AP, and the vertical axis represents the IA index. **Fig. 3** depicts the service quality classification. The six classes of attributes are high-performance excitement (HE), high-performance performance (HP), high-performance basic (HB), low-performance basic (LB), low-performance performance (LP), and low-performance excitement (LE).



**Fig. 3.** Categorization of attributes.

Note: The black vertical dotted line is the dividing line that classifies the attributes into two categories (i.e., high and low performance), and the position is determined by calculating the average performance of all attributes. The two red horizontal dashed lines are the demarcation lines ( $\theta$  and  $-\theta$ ), which are used to categorize the attributes into three categories (i.e., basic, performance, and excitement attributes). If  $-\theta \leq IA_i \leq \theta$ , it is considered a performance attribute. If  $\theta < IA_i \leq 1$ , it is regarded an excitement attribute. If  $-1 \leq IA_i < -\theta$ , the attribute is considered a basic attribute.

According to Kano's three-factor theory, basic attributes have considerable potential to cause dissatisfaction, whereas performance attributes have equal potential for both satisfaction and dissatisfaction. Finally, the excitement attribute has a high potential to generate satisfaction. To optimize CS with minimal expenditure, managers should enhance the quality of attributes with low performance and maintain the quality of attributes with high performance. Additionally, for attributes with equivalent performance levels, both low and high, resources should be allocated in the following order: basic, performance, and excitement. Combining these two aspects, the priority

order for allocating resources in the AIPA is  $LB > LP > LE > HB > HP > HE$  (Bi et al., 2020).

According to Albayrak (2016) and Bi (2019b), the attributes in the AIPA diagram belong to the following quadrants:

- (1) Q1 is referred to as the “major advantage.” The properties positioned in Q1 have a high AP and a positive IA, indicating that the property has a high-performance score and a positive impact on CS. Therefore, the attributes positioned in Q1 can be considered major strengths, and hotels should aim to maintain the performance level of these attributes.
- (2) Q2 is referred to as “optimization potential.” Attributes positioned in Q2 have a positive IA and a poor AP, indicating that the performance of the attribute is poor, despite the high performance of the attribute that elicits CS. For attributes positioned in Q2, the hotel can allocate more resources and transform them into superior attributes.
- (3) Q3 is called “urgent action.” The attributes positioned in Q3 have a negative IA and a poor AP, indicating that the attribute performs poorly and is prone to CS. Therefore, the attributes positioned in Q3 can be considered major weaknesses, and the hotel should take urgent action to improve them.
- (4) Q4 is labeled “no concern.” Properties positioned in Q4 have a negative IA and a high AP, indicating that the property is more prone to customer dissatisfaction but performs reasonably well. These attributes are not positioned as strengths nor weaknesses of the hotel in Q4 and thus do not require much attention.

It is crucial to consider the dynamics of the asymmetric relationship between AP and CS to examine the trends of significant attributes of a product or service. Therefore, this study suggests using DAIPA, which enables managers to track AP and IA of a product/service over time. To perform DAIPA, the time period  $T$  must first be established. Period  $T$  can be a year, season, or month. For the purposes of this article, we used one year for  $T$ , and we divided the online ratings into subsets according to the specified timeframe. Using the process outlined in Sections 3.1 and 3.2, the performance of attribute  $A_i$  at the  $t$ th time period and IA can be estimated, denoted as  $AP_i^t$  and  $IA_i^t$  ( $i = 1, 2, 3, \dots, I, t = 1, 2, 3, \dots, T$ ), respectively.

Thus, based on the obtained  $AP_i^t$  and  $IA_i^t$ , an AIPA plot concerning the  $t$ th period can be drawn, and a three-dimensional (3D) plot of DAIPA can be obtained by combining AIPA plots. An example DAIPA plot is shown in **Fig. 4**. Based on the DAIPA plot, we can comprehend alterations in AP and the uneven influence of attributes on CS over time, as well as changes in the category to which the attribute belongs.

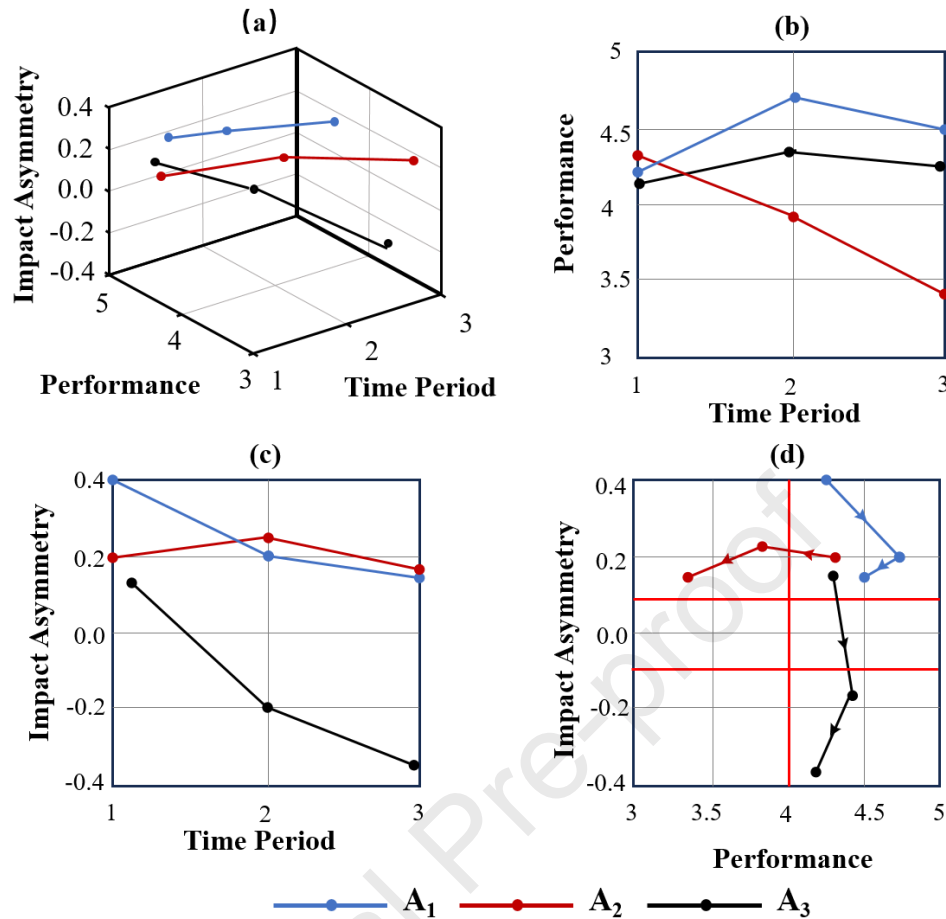


Fig. 4. An example of the DAIPA plot.

Note: (a) The first figure shows a three-dimensional plot with AP on the horizontal axis, IA on the vertical axis, and the year on the z-axis. (b) The second graph is an AP-year change graph. (c) The third graph depicts the IA-year change. (d) The fourth graph is an IA-AP graph. Each point represents a combination of years and categories. The arrow points in the direction of the year increase.

Based on these graphs, we selected three potential patterns or trends over time to demonstrate the use of the DAIPA graph.

- (1) If an attribute is consistently located in a quadrant (e.g., attribute  $A_1$  in **Fig. 4(d)**), this indicates that the attribute's performance and asymmetric impacts do not change significantly over time. The importance and management strategy of the attribute are the same as those of attributes located in the same quadrant of the AIPA diagram.
- (2) If the location of an attribute changes over time from Q1 (high performance) to Q2 (high performance) (e.g., attribute  $A_2$  in **Fig. 4(d)**), this indicates that the performance of the attribute declines over time. Therefore, additional funds should be allocated to boost this attribute's performance. Over time, base attributes should be modified and improved.
- (3) If the position of the attribute changes from Q1 (positive IA) to Q4 (negative IA) (e.g., attribute  $A_3$  in **Fig. 4(d)**), it shows that although people are becoming more aware of this attribute and demand for it is rising, the quality performance is now subpar or has worsened. Therefore,

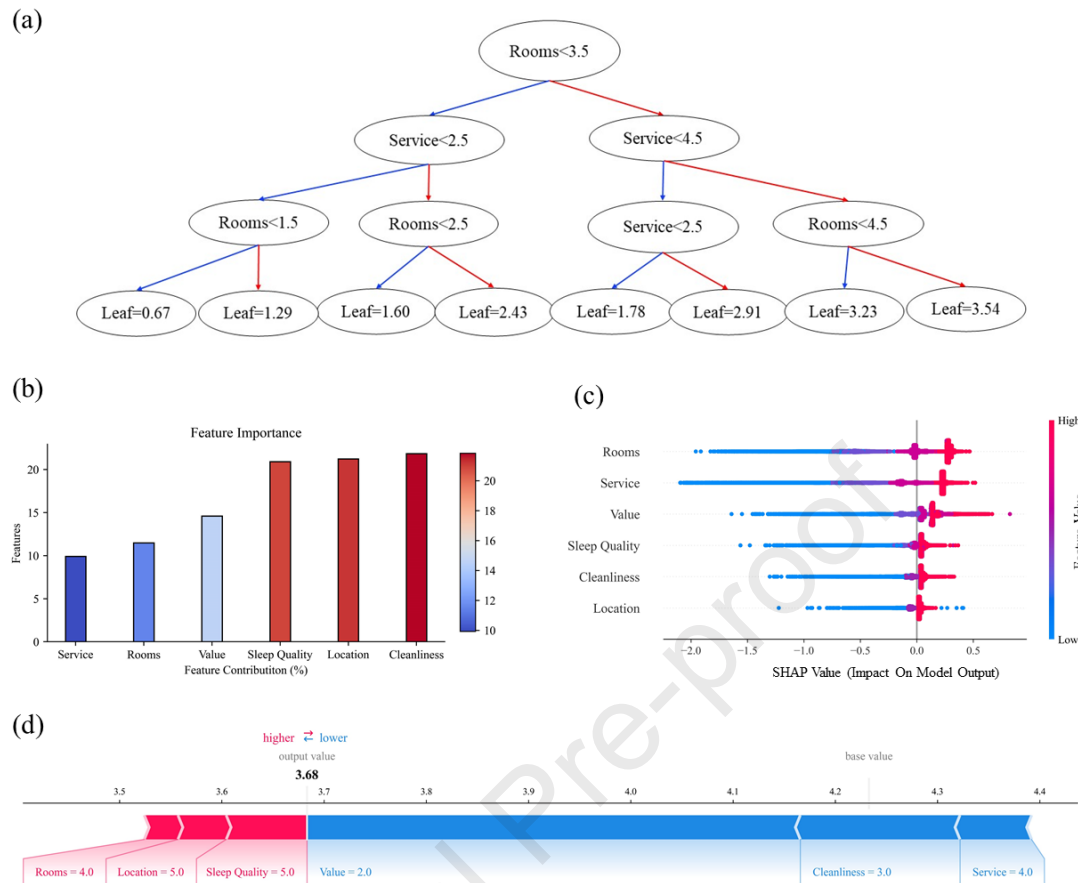
more attention should be paid to this attribute.

## 4. Results and discussion

### 4.1 Results of model interpretation

As described earlier in the model selection section, five machine learning algorithms were implemented in this study, and six evaluation metrics were used to assess the predictive ability of each model to determine the best-fitting regression prediction model. The results showed that the XGBoost model had the best fit, explaining more than 80% of the variability in hotel CS. In this model, the importance of each feature is measured based on the number of splits and split gain of each feature in the tree. **Fig. 5** shows the constructed trees and the analysis of feature importance.

Considering the frequency of the feature and split gain, XGBoost calculates the importance score for each feature. As shown in **Fig. 5** (b), the importance of each attribute feature to the model was analyzed for all hotels in New York City, with values ranging from 0% (least significant) to 100% (most significant) and the sum equal to 100%. Nearly half of the features have significance values greater than 15%, and most of the characteristics significantly contributed to the reliability of the model. Among them, “cleanliness” is the most important characteristic for the overall model prediction.



**Fig. 5.** Tree model and feature importance analysis.

Note: (a) Tree model constructed by XGBoost (b) Importance of each attribute in model. (c) Summary plot of the impacts of features across all samples in the model. The features are positioned on the y-axis, and each Shapley value determines the position on the x-axis. The x-axis shows the influence on the model output, where positive (negative) values increase (decrease) OCS. Colors represent feature values (red = high, blue = low), whereby changes in feature values can be matched with their impact on OCS. The features were sorted by importance based on the average absolute value of the SHAP values for each feature. (d) For this graph, the model output value is 3.68. Base value: The average of the model output and training data was 4.23. The numbers below the arrows represent the feature values in this instance; for example, Rooms = 4.0, Location = 5.0. Features that push the prediction higher (lower) are shown in red (blue). The longer the arrow, the greater the effect of the feature on the output. The magnitude of the decrease or increase in the impact can be seen from the value of the scale on the x-axis.

Based on the trained model, model interpretation was performed using the Python SHAP package. The SHAP force plot provides the interpretability of single-model predictions to find interpretations of instance-specific predictions. Furthermore, plotting the Shapley value of each feature for each sample provides an overall indication of the most important features and the extent

of their impact on the dataset, as shown in **Fig. 5 (c)**. According to this figure, it is evident that “service” is the most crucial feature of the model, with the largest impact on CS. The poor performance of this feature results in a small SHAP value, and consequently, significant customer dissatisfaction. On the other hand, the influence of “location” on satisfaction in New York City hotels is relatively low. This is consistent with the results of previous statistical analyses and provides further evidence of the effectiveness of using SHAP for model interpretation.

## 4.2 Classification of different attributes

The SHAP value distribution graph feature allows us to identify which attribute has the highest impact on CS; that is, which attribute is most significant to customers. However, it does not reveal whether the sentiment toward an attribute is positive or negative. In other words, an attribute may be significant to customers because it meets their expectations and thus has a positive impact, or because it fails to meet their expectations, resulting in negative customer feedback. To determine whether the impact of an attribute is positive or negative, we must compare its positive performance with its negative performance and assess the extent of its respective impact on OCS.

To achieve this, we calculated two coefficients (penalty and reward) for all attributes using all of the obtained SHAP values. These coefficients are based on the improved PRCA methodology described in Section 3.2 of this study. We then computed the IA values of the six attributes using **Equation (3)**. Finally, we categorized each attribute into Kano categories.

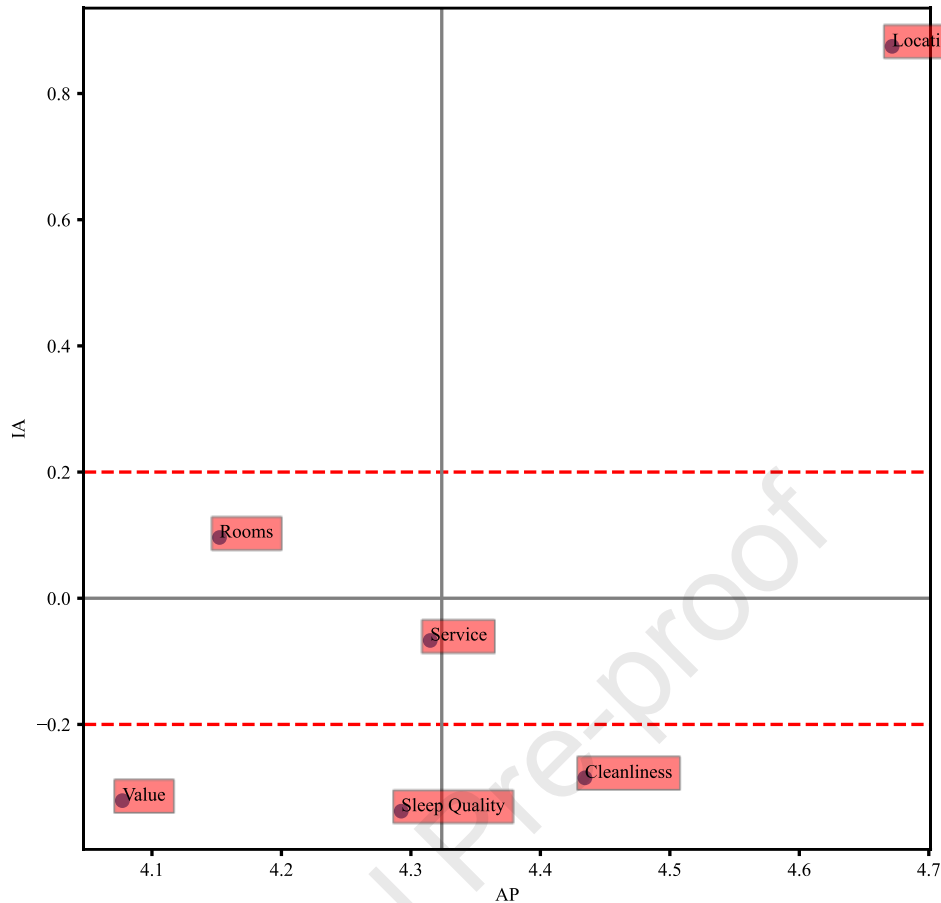
**Table 4** shows that among all the New York City hotels, “location” and “rooms” are excitement attributes, “service” is a performance attribute, and “value,” “sleep quality,” and “cleanliness” are the base attributes.

Further, based on the obtained AIPA graph, “location” is the HE attribute, “rooms” and “service” are the LP attributes, “cleanliness” is the HB attribute, and “value” and “sleep quality” are the LB attributes. Therefore, the order of prioritization of attributes for the future allocation of resources at New York City hotels can be determined as follows:

Value>Sleep quality>Service>Rooms>Cleanliness>Location

**Table 4.** Results from the asymmetric analysis of attribute performance and customer satisfaction in New York City hotels.

	AP	$\beta_{low}^i$	$\beta_{high}^i$	IA	Type
Value	4.08	-0.71	0.22	-0.53	B
Rooms	4.15	-0.46	0.84	0.30	E
Location	4.67	-0.02	0.18	0.81	E
Cleanliness	4.43	-0.36	0.18	-0.32	B
Service	4.31	-0.75	0.73	-0.02	P
Sleep Quality	4.29	-0.47	0.15	-0.53	B



**Fig. 6.** The AIPA plot for all New York City hotels.

Note: In the AIPA plot, the two red horizontal dashed lines are the demarcation lines (0.2 and -0.2), which are used to categorize the attributes into three categories: basic, performance, and excitement.

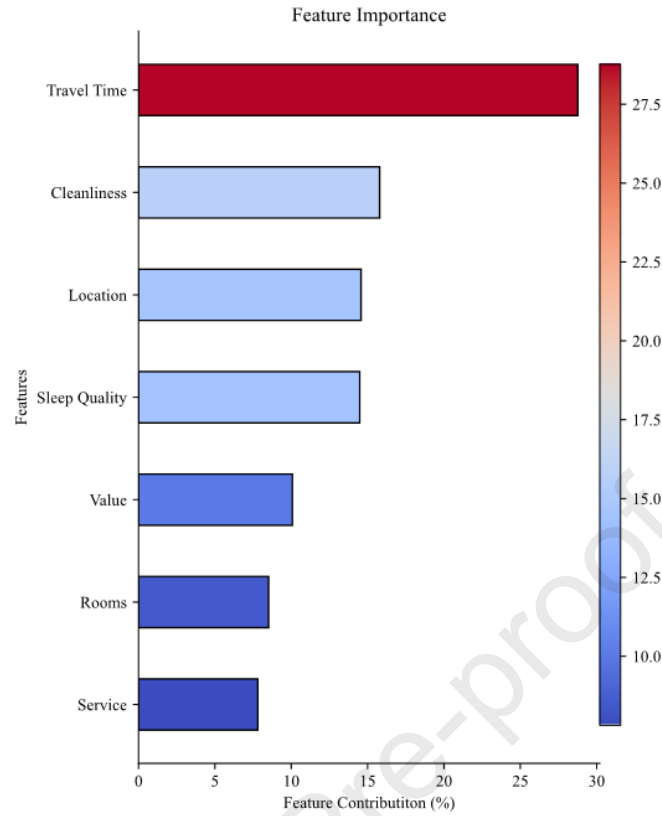
### 4.3 The dynamic asymmetric effect between AP and CS

To verify that the relationship between AP and CS changes over time, we first added the travel year as a feature to the dataset fed into the XGBoost model for training and observed whether the hotel star rating impacted CS based on feature importance. The accuracy of the model was obtained, as shown in **Table 5**. The results revealed that travel time does impact CS, and this feature contributes significantly to the model prediction (26%), which means that CS or customer expectation changes depending on travel time.

**Table 5.** Model result and feature importance after adding time variables.

	Train	Test
Evar	0.84	0.82
MAE	0.28	0.29
MAPE	0.08	0.09
MSE	0.17	0.19
RMSE	0.10	0.10
$R^2$	0.84	0.82

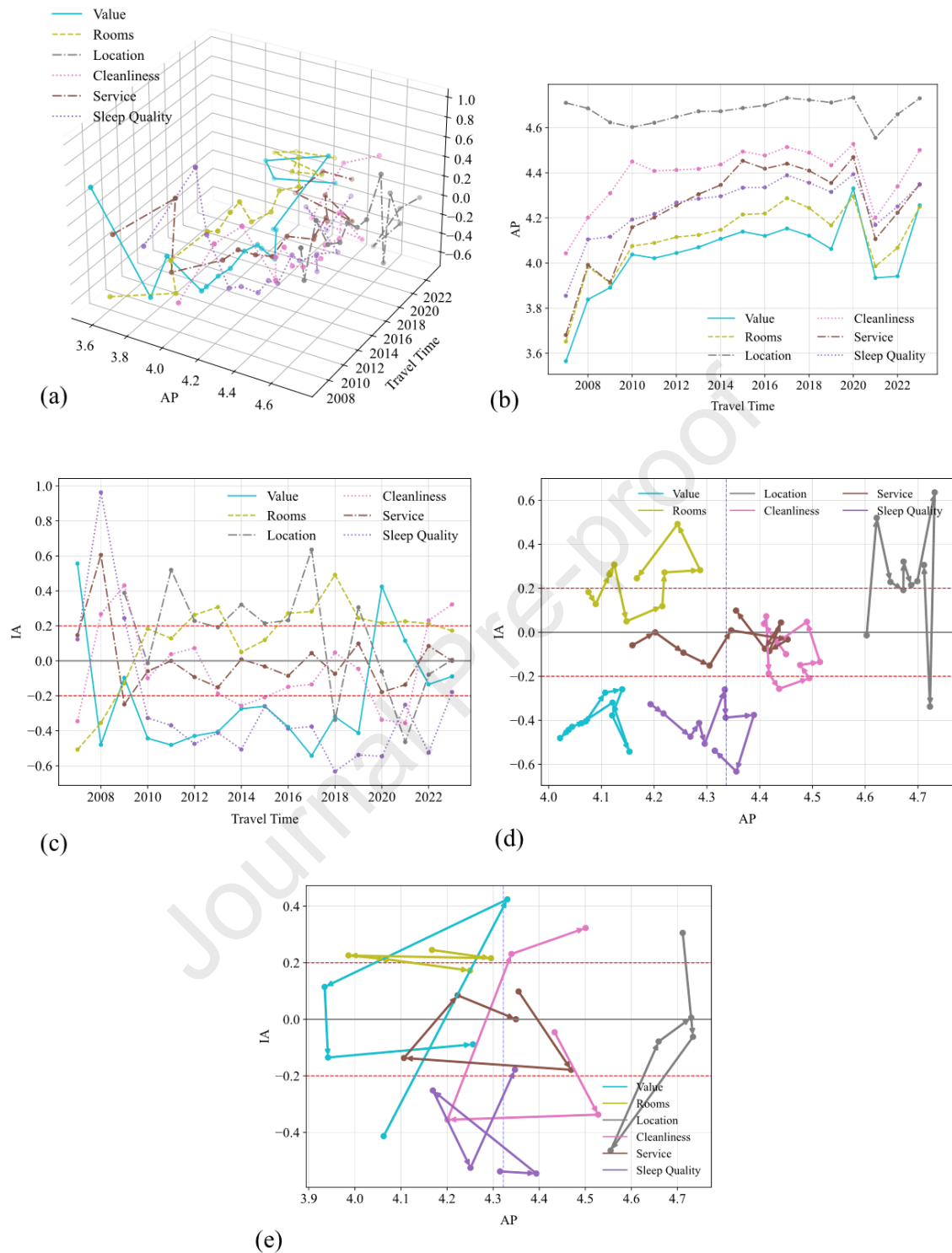




**Fig. 7.** Feature importance for travel time.

Based on the year of publication of online reviews, we divided the data into a total of 17 subsets from 2007–2023. Similarly, the training, validation, and test sets were divided according to 7:1:2. The accuracy of the XGBoost model on different datasets is presented in **Appendix 3**. The explained variance and resolvability coefficient ( $R^2$ ) of the model for both the training and test sets were greater than 80%.

Using the procedures in Sections 3.2.2 and 3.2.3, it is possible to determine the  $AP_i^t$  and  $IA_i^t$  of attribute  $A_i$  at the  $t$ th time period, where  $i = 1, 2, \dots, 6, t = 1, 2, \dots, 7$ . The DAIPA plot depicted in **Fig. 8** was created based on the obtained  $AP_i^t$  and  $IA_i^t$ .



**Fig. 8.** DAA of the AP-CS relationship for New York City hotels over the period 2007–2023. Notes: (a) 3D image of the DAIPA plot; (b), (c), and (d) are travel time-AP, travel time-IA, and AP-IA images of the 3D image, respectively. (b) and (c) show the changes in the performance and IA of each attribute over time, respectively. (d) This plot reflects the quadrant change in each attribute over time in the AIPA graph for the period 2010–2019, with arrows indicating the time series. (e)

Quadrant changes in attributes on the AIPA chart between 2019 and 2023.

This data relates to two quantities—the average score of hotels (AP) and the average impact of hotels (IA)—as well as six categories (value, rooms, location, cleanliness, service, and sleep quality) and years (2007–2010). The following are interpretations of this model.

From **Fig. 8 (b)**, it can be seen that AP shows large fluctuations in 2007–2010 and 2019–2023, which correspond to the financial crisis and the novel coronavirus epidemic, respectively; 2020–2022 is the period of the global novel coronavirus epidemic outbreak, the outbreak and spread of which greatly impacted the global hospitality industry. The figure shows that during this period, hotel ratings for each property dropped substantially. The following factors related to the epidemic may have contributed to the decrease in hotel attribute scores.

- (1) Effects on hotel business: The outbreak and postponement of the global pandemic led to restrictions on tourism and business travel, resulting in a sharp decrease in hotel businesses and an impact on hotel operations.
- (2) Effects on hotel service quality: During the epidemic, to prevent the spread of the epidemic, hotels took a variety of preventive and control measures, such as strict control over and restriction of services. These factors may have led to a decline in service quality and customer ratings may have fallen accordingly.
- (3) Effects on travelers' expectations: During the pandemic, travelers' expectations and psychological states changed; they paid more attention to hygiene and safety and were more aware and careful, altering their rating criteria for hotels.

In summary, the global epidemic of the novel coronavirus significantly impacted the hotel industry, resulting in a sharp drop in business, declines in service quality, and changes in customer expectations, which ultimately resulted in a substantial drop in hotel attribute ratings in all areas.

Based on the above analysis, we divided the time into three periods: 2007–2010, 2010–2019, and 2019–2023, in which AP showed rapid improvement, stable maintenance, and a sharp decline, respectively. After the epidemic was resolved in 2023, the hospitality industry gradually began to rebound, and AP displayed gradual improvement.

According to **Fig. 8 (b)**, location was the best-performing attribute in each period. The performance of attributes value and rooms was worse than that of cleanliness, service, and sleep quality. In addition, the performance of location was generally consistent over time.

As shown in **Fig. 8 (c)** the IA of each attribute varied depending on the time period. Overall, as with the travel time-AP graph, changes in this graph over the three time periods were evident, with dramatic changes in the IA values of the attributes in the 2007–2010 and 2019–2023 time periods.

The shift in the quadrants in which an attribute fell between 2010 and 2019 was depicted in **Fig. 8 (d)**. However, the regions to which each attribute belonged did not change significantly. Managers should be interested in the value attribute because it is in the LB region. The service attribute gradually shifted from LP to HP, indicating a gradual improvement in performance. The

cleanliness attribute mostly remained in the HP region, whereas the rooms attribute gradually moved from LP to LE, indicating that users were paying more attention to it. Its good performance was likely increasing user satisfaction, but there was room for improvement. Hotel managers could utilize this new attribute to enhance their performance. Although the location attribute changed frequently, they mostly fell within the HE zone. Given the rooms attribute's poor performance, hotel management should improve this attribute by using new technical means to increase user satisfaction.

**Fig. 8 (e)** illustrated a significant change in the quadrant to which the attribute belonged from 2019 to 2023. Cleanliness changed from HP to HB in 2020 and from HB to LB in 2021, demonstrating that users were becoming increasingly concerned about hygiene attributes during the novel coronavirus epidemic and their requirements were gradually increasing. Therefore, additional resources must be allocated to addressing these attributes. In 2020, both service and location ratings decreased from HE to HP, likely due to people needing more hotel services to minimize their interactions with others and needing the hotel's location to be suitable to avoid or minimize public transportation during the epidemic. Accordingly, the hotel location should be as suitable as possible to avoid public transportation. Additionally, value moved from the LB zone to the LE zone in 2020. This is most likely because if the price is appropriate while meeting customers' requirements for epidemic prevention, users will be satisfied. Finally, room and service ratings did not change significantly over this period.

Using the obtained SHAP values, we calculated the two coefficients (penalty and reward) for all attributes based on the improved PRCA methodology in Section 3.2. The IA values were calculated for the six attributes for all star-rated hotels according to Equation 2. Finally, we categorized each attribute into a Kano category. The PRCA results for all times are presented in **Appendix 4**, and verify the dynamic asymmetric effect of the AP-CS relationship from the perspective of time. Prior to the COVID-19 epidemic, "value" and "sleep quality" were gradually changing from excitement and performance attributes to basic attributes. As technology evolves and word of mouth spreads, certain attributes become the market standard (File et al., 1994), no longer have unlimited benefits, and negatively impact CS when the performance level of the attribute falls below customer expectations. As customers experience hotels, they become aware of their basic functions and find that they should be cost-effective and able to provide a good night's sleep. Based on the results presented in the figures and tables, we can see that the relationship between the AP and CS changes over time, particularly during major events, which can lead to more significant changes. This dynamic is reflected in various performance levels and categories of hotel attributes at different times. Based on this observation, hotel managers can predict changes in hotel attributes using historical data to better meet consumer needs.

## 5. Conclusions

This study proposes a method for constructing a hotel CS function based on IML that analyzes

the relationship between AP and CS from three perspectives— feature, AIPA and DAIPA—and fully verifies the dynamic asymmetric effect of the relationship between AP and CS.

## 5.1 Theoretical implications

This study analyzed 297,244 UGD points from 423 hotels in New York City using an improved PRCA method based on SHAP values and a constructed user satisfaction function. The aim was to validate the dynamic asymmetric effect of the AP-CS relationship, manifesting in two aspects. First, different APs perceive CS differently. Specifically, the positive and negative performances of an attribute cause different changes in CS. This difference was identified by categorizing attributes based on the three-factor theory and the PRCA method. Second, the same attribute can belong to different types at different stages, meaning that customer expectations and perceptions vary and are affected by time and events. DAA was used for analysis and validation of this finding.

Based on the analyses described above, the primary contributions of this work are as follows. First, this study proposes an IML-based method for assessing CS, using data from user ratings on TripAdvisor's website. Although previous studies have investigated the asymmetric effect between AP and CS, but in terms of data volume, they have primarily relied on "small data" from particular market groups, which restricts the generalizability of their results (Albayrak and Caber, 2015; Xu and Li, 2016; Zhou et al., 2014). Our findings are based on user data from hotels in New York City, consisting of 940,283 reviews from 928 hotels. The large sample size (encompassing different hotels, users, and time periods) confirms the dynamic asymmetrical relationship between AP and CS. New York City, being a highly internationalized city that attracts numerous domestic and international tourists, professionals, and immigrant communities, offers a diverse hotel market with various accommodation types including luxury, business, and resort hotels, making it suitable for a multifaceted study. We chose to utilize both the XGBoost model and SHAP explanatory mechanism due to their efficacy. Specifically, the XGBoost model yielded the best results among the evaluated models, and SHAP is a novel interpretation mechanism that can explain the prediction results of various machine learning models. Together, they outlined the contribution of each feature to the model output, enabling a better understanding of the model's prediction process. Based on the above models and mechanisms, we constructed a complete framework to analyze the relationship between AP and CS.

Second, based on the interpretation results of the satisfaction prediction model, we propose an improved PRCA method to estimate the importance of each attribute by combining the AIPA and DAIPA. The PRCA suggested an adaptive model that relied on SHAP values without considering any prior assumptions about the online rating distribution. Each attribute's IA value can be acquired precisely and consistently, enabling attribute categorization and dynamic analysis based on the IA value. The DAIPA evaluated each attribute's performance and the IA offered by a hotel over different periods, representing a useful effort to advance and enhance the IPA's theory and methodology. This study confirms that the asymmetric connection between the AP and CS varies over diverse travel

periods. Primarily, the analysis of the six attributes revealed interesting findings on the most critical aspects of guests. Consistent with previous research (Bi et al., 2020), for all hotel types, “value,” “cleanliness,” and “sleep quality” were identified as basic attributes, with the hotel's primary function being to provide a place to sleep. Out of all hotel factors, “location” had the least influence on OCS and ranked last, which is a unique feature of New York City hotels.

Third, based on variations in AP over different periods, we divided the timeframe into three segments: a period of swift growth from 2007–2010, followed by a stable phase of maintenance from 2010–2019, and ultimately a time of rapid decline from 2019–2023. During 2007–2010 and 2019–2023, the IA values of the attributes underwent significant changes. According to DAIPA, the quadrant to which the attribute belongs changed significantly from 2019 to 2023. One notable change includes “cleanliness” moving from HP to HB in 2020, and then to LB in 2021. This shift indicates that users were becoming increasingly concerned about hygiene attributes during the novel coronavirus epidemic, and their requirements were gradually increasing. In 2020, both “service” and “location” moved from HE to HP. One explanation for this is that during an epidemic, people desired more hotel services to minimize contact with others and sought hotel locations convenient enough to avoid public transportation. Most likely due to users being satisfied if the price is right while meeting the customer's epidemic prevention requirements, “value” moved from LB to LE. In contrast, “rooms” and “sleep quality” changed relatively little. The findings of this study are significant for measuring the impact of the COVID-19 pandemic on CS. It shows that several attributes changed, providing valuable insights for managers facing future significant public events and contributing to the limited prior research examining the pandemic's impact on CS through the AP-CS relationship.

In summary, this study comprehensively analyzes the dynamic and asymmetric effects of the AP-CS relationship, proposes a complete set of IML frameworks to study it, and explains it from different theoretical perspectives. The insights gained can help hotel managers make informed assessments and decisions.

## 5.2 Practical implications

Apart from the mentioned theoretical contributions, this study offers managerial insights for hotel practitioners, which are described below.

First, for New York City hotels, “location” is the HE attribute “rooms” and “service” are the LP attributes, “cleanliness” is the HB attribute, and both “value” and “sleep quality” are the LB attributes. Therefore, the order of prioritization of the attributes for the future allocation of resources at New York City hotels can be determined as:

Value>Sleep quality>Service>Rooms>Cleanliness>Location.

“Value” is a base attribute; thus low performance in this attribute is likely to lead to customer dissatisfaction. New York City hotels require efficient operations to minimize operating costs and satisfy customers. Therefore, they should focus on achieving efficient operations to provide the best

service at the best price. “Sleep quality” is a fundamental attribute that customers care about; for all hotels, ensuring customers’ sleep quality should be the first priority. To this end, hotels can take measures to improve the travelers’ quality of sleep, such as enhancing room soundproofing and providing different bedding (e.g., pillows) for travelers. In addition, “service” and “rooms” are performance attributes. Reasonable arrangements in these dimensions can effectively improve CS, such as a standardized room layout, ensuring the normal use of various facilities, and offering full amenities. Previous research has primarily analyzed CS from the perspective of a single hotel or specific customer segment (Zhang et al., 2021), with limited attention given to identifying the key factors impacting CS across different locations. Our study addresses this gap in the literature and contributes to understanding how business management practices of hotels in a certain region should be improved and optimized.

Second, hoteliers should distinguish between periods affected by major public health events, such as COVID-19, in which hotel customers' expectations of various service characteristics' change. DAIPA demonstrates that the corresponding quadrants of service attributes were modified dramatically between 2019 and 2023. The rating for “cleanliness” changed from HP to HB in 2020 and shifted again from HB to LB in 2021. These changes demonstrate that, during the COVID-19 pandemic, users became increasingly concerned about hygiene attributes, and their demands gradually increased to the point that hygiene has become a basic requirement. Considering the link between current events and changes in users' expectations, it is crucial to take measures to improve hygiene levels in hotels (e.g., disinfecting regularly and providing hand sanitizers) to effectively enhance traveler satisfaction. It is also essential to assign more resources to “cleanliness” in hotels. The elevated “value” shift from LB to LE implies that CS can be effectively improved by serving customers to the best of the hotel's ability while ensuring they feel they are getting value for their money during a major event. Thus, hotels must efficiently make reasonable arrangements in the case of significant events, such as proposing suitable preferential policies to enhance their reputation. As demonstrated above, studying the impact of AP on CS during the COVID-19 pandemic is crucial. These findings further demonstrate the dynamic asymmetry of the AP-CS relationship, emphasizing the need for managers to adapt hotel business strategies to accommodate dynamic changes in customers, the market, and the environment over time.

### **5.3 Limitations and avenues for future research**

Some limitations of this study may present opportunities for future research. First, only the asymmetric relationship between AP and CS was specifically examined in this study via six variables (location, cleanliness, room, service, sleep quality, and value). In the future, researchers may look into other characteristics, such as “food,” “check-in/check-out,” etc. Second, this study used only user-generated ratings. A significant number of user-generated reviews containing vast amounts of unstructured and valuable data, including images and text, are available to the public on the Internet and can provide a great deal of information about CS (Nie et al., 2020; Oender, 2017;



Xiao et al., 2020). However, hotel characteristics mentioned in various reviews may differ. Structured data obtained from user-generated reviews through feature extraction and sentiment analysis methods may have certain limitations in that it is difficult to combine text and photos in a structured manner. (Bi et al., 2019a). This limitation could have affected the outcomes of CS's asymmetric effects of CS on AP. Identifying the asymmetric effects of additional attributes based on user-generated reviews is a topic for future research. Third, the data used in this study were sourced solely from TripAdvisor, New York City. Therefore, it is important to re-evaluate the relationship between AP and CS across diverse cities for further verification. Finally, the distribution of attributes depends on the location of the crosshairs, and AIPA's categorization of attributes into one of three types depends on the definition of the cut-off point. Changing the cut-off point changed the distribution. Thus, although the subjective cut-off of  $\theta = 0.2$  seemed to have worked well for New York City hotels because the classification results validate the dynamic asymmetry of the AP-CS relationship well, a different cut-off point may need to be chosen for other cities.

## Acknowledgment

This research was partially supported by the National Key R&D Program of China (Grant No.: 2022YFF0903000) and the National Natural Science Foundation of China (Grant Nos. 72101197 and 71988101).

## Reference

- Albayrak, T., 2019. The inclusion of competitor information in the three-factor theory of customer satisfaction. *Int. J. Contemp. Hosp. Manag.* 31(4), 1924–1936.
- Albayrak, T., Caber, M., 2013a. Penalty-Reward-Contrast Analysis: a review of its application in customer satisfaction research. *Total Qual. Manag. Bus. Excell.* 24(11–12), 1288–1300.
- Albayrak, T., Caber, M., 2013b. The symmetric and asymmetric influences of destination attributes on overall visitor satisfaction. *Curr. Issues Tour.* 16(2), 149–166.
- Albayrak, T., Caber, M., 2015. Prioritisation of the hotel attributes according to their influence on satisfaction: A comparison of two techniques. *Tour. Manag.* 46, 43–50.
- Albayrak, T., Caber, M., 2016. Destination attribute effects on rock climbing tourist satisfaction: an Asymmetric Impact-Performance Analysis. *Tour. Geogr.* 18(3), 280–296.
- Albayrak, T., Caber, M., Bideci, M., 2016. Identification of hotel attributes for senior tourists by using Vavra's importance grid. *J. Hosp. Tour. Manag.* 29, 17–23.
- Back, K. J., 2012. Impact-range performance analysis and asymmetry analysis for improving quality of Korean food attributes. *Int. J. Hosp. Manag.* 31 (2), 535-543.
- Berezina, K., Bilgihan, A., Cobanoglu, C., et al., 2016. Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. *J. Hosp. Mark. Manag.* 25 (1), 1-24.
- Bi, J. W., Liu, Y., Fan, Z. P., et al., 2019a. Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *Int. J. Prod. Res.* 57 (22), 7068-7088.



- Bi, J. W., Liu, Y., Fan, Z. P., et al., 2019b. Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tour. Manag.* 70, 460-478.
- Bi, J. W., Liu, Y., Fan, Z. P., et al., 2020. Exploring asymmetric effects of attribute performance on customer satisfaction in the hotel industry. *Tour. Manag.* 77, 104006.
- Bigorra, A. M., Isaksson, O., Karlberg, M., 2019. Aspect-based Kano categorization. *Int. J. Inf. Manag.* 46, 163-172.
- Brandt, D. R., 1988. How service marketers can identify value-enhancing service elements. *J. Serv. Mark.* 2(3), 35-41.
- Caber, M., Albayrak, T., Loiacono, E. T., 2013. The classification of extranet attributes in terms of their asymmetric influences on overall user satisfaction: an introduction to asymmetric impact-performance analysis. *J. Travel Res.* 52 (1), 106-116.
- Chang, Y. C., Ku, C. H., Le Nguyen, D. D., 2022. Predicting aspect-based sentiment using deep learning and information visualization: The impact of COVID-19 on the airline industry. *Inf. Manag.* 59 (2), 103587.
- Chatterjee, D. S., 2019. Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. *Decis. Support Syst.* 119, 14-22.
- Chen, K., Kou, G., Shang, J., et al., 2015. Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches. *Electron. Commer. Res. Appl.* 14 (1), 58-74.
- Chen, L. F., 2014. A novel framework for customer-driven service strategies: A case study of a restaurant chain. *Tour. Manag.* 41, 119-128.
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, pp. 785-794.
- Conklin, M., Powaga, K., Lipovetsky, S., 2004. Customer satisfaction analysis: Identification of key drivers. *Eur. J. Oper. Res.* 154 (3), 819-827.
- Covert, I., Lundberg, S. M., Lee, S. I., 2020. Understanding global feature contributions with additive importance measures. In: *Advances in Neural Information Processing Systems*, 33, pp. 17212-17223.
- Cunningham, P., Delany, S. J., 2021. k-Nearest neighbour classifiers-A Tutorial. *ACM Comput. Surv. (CSUR)* 54(6), 1-25.
- Davras, O., Caber, M., 2019. Analysis of hotel services by their symmetric and asymmetric effects on overall customer satisfaction: A comparison of market segments. *Int. J. Hosp. Manag.* 81, 83-93.
- Deng, W. J., Chen, W. C., Pei, W., 2008. Back-propagation neural network based importance-performance analysis for determining critical service attributes. *Expert Syst. Appl.* 34 (2), 1115-1125.
- Dong, N., Huang, H., Zheng, L., 2015. Support vector machine in crash prediction at the level of traffic analysis zones: assessing the spatial proximity effects. *Accid. Anal. Prev.* 82, 192-198.
- File, K. M., Cermak, D. S., Prince, R. A., 1994. Word-of-mouth effects in professional services buyer behavior. *Serv. Ind. J.* 14 (3), 301-314.
- Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 29(5),

- 1189–1232.
- Gou, H., Su, L., Zhang, G., et al., 2022. A XGBoost method based on telecom customer satisfaction enhancement strategy. In: 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), IEEE, pp. 209-213.
- Hu, F., Li, H., Liu, Y., et al., 2020. Optimizing service offerings using asymmetric impact-sentiment-performance analysis. *Int. J. Hosp. Manag.* 89, 102557.
- Jain, A. K., Mao, J., Mohiuddin, K. M., 1996. Artificial neural networks: A tutorial. *Comput.* 29 (3), 31-44.
- Jang, S. S., Ha, A., Silkes, C. A., 2009. Perceived attributes of Asian foods: From the perspective of the American customers. *Int. J. Hosp. Manag.* 28 (1), 63-70.
- Ji, F., Cao, Q., Li, H., et al., 2023. An online reviews-driven large-scale group decision making approach for evaluating user satisfaction of sharing accommodation. *Expert Syst. Appl.* 213, 118875.
- Joung, J., Kim, H. M., 2022. Explainable neural network-based approach to Kano categorisation of product features from online reviews. *Int. J. Prod. Res.* 60(23), 7053–7073.
- Ju, Y., Back, K.J, Choi, Y., et al., 2019. Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction. *Int. J. Hosp. Manag.* 77, 342–352.
- Kano, N., Seraku, N., Takahashi, F., et al., 1984. Attractive quality and must-be quality. *J. Jpn. Soc. Qual. Control.* 14(2), 39–48.
- Lee, J.-S., Choi, Y., Chiang, C.-H., 2017. Exploring the dynamic effect of multi-quality attributes on overall satisfaction: The case of incentive events. *Int. J. Hosp. Manag.* 64, 51–61.
- Lee, J.-S., Min, C., 2013. Prioritizing convention quality attributes from the perspective of three-factor theory: The case of academic association convention. *Int. J. Hosp. Manag.* 35, 282–293.
- Li, H., Bruce, X. B., Li, G., et al., 2023. Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tour. Manag.* 96, 104707.
- Li, H., Liu, Y., Tan, C. W., et al., 2020. Comprehending customer satisfaction with hotels: Data analysis of consumer-generated reviews. *Int. J. Contemp. Hosp. Manag.* 32(5), 1713–1735.
- Liu, C. J., Huang, T. S., Ho, P. T., et al., 2020. Machine learning-based e-commerce platform repurchase customer prediction model. *Plos one* 15(12), e0243105.
- Liu, X., Chen, Y., Qiu, Z., et al., 2019. Forecast of the Tourist Volume of Sanya City by XGBoost Model and GM Model. In: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, pp. 166-173.
- Liu, Y., Bi, J. W., Fan, Z. P., et al., 2017. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Inf. Fusion.* 36, 149–161.
- Liu, Z., Jiang, P., Wang, J., et al., 2023. Hospitality order cancellation prediction from a profit-driven perspective. *Int. J. Contemp. Hosp. Manag.* 35(6), 2084-2112.
- Lundberg, S. M., Lee, S. I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30.
- Martilla, J. A., James, J. C., 1977. Importance-Performance Analysis. *J. Mark.* 41(1), 77–79.
- Mathe-Soulek, K., Slevitch, L., Dallinger, I., 2015. Applying mixed methods to identify what drives quick service restaurant's customer satisfaction at the unit-level. *Int. J. Hosp. Manag.* 50, 46–54.

- Matzler, K., Bailom, F., Sauerwein, E., et al., 1996. How to delight your customers. *J. Prod. Brand Manag.* 5(2), 6–18.
- Matzler, K., Sauerwein, E., 2002. The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis. *Int. J. Serv. Ind. Manag.* 13(4), 314–332.
- McKinney, V., Yoon, K., Zahedi, F., 2002. The measurement of web-customer satisfaction: An expectation and disconfirmation approach. *Inf. Syst. Res.* 13(3), 296–315.
- Mikulić, J., Krešić, D., Miličević, K., et al., 2016. Destination Attractiveness Drivers among Urban Hostel Tourists: An Analysis of Frustrators and Delighters. *Int. J. Tour. Res.* 18(1), 74–81.
- Mikulić, J., Prebežac, D., 2012. Using dummy regression to explore asymmetric effects in tourist satisfaction: A cautionary note. *Tour. Manag.* 33(3), 713–716.
- Mikulić, J., Prebežac, D., 2008. Prioritizing improvement of service attributes using impact range-performance analysis and impact-asymmetry analysis. *Manag. Serv. Qual.* 18(6), 559–576.
- Murdoch, W. J., Singh, C., Kumbier, K., et al., 2019. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U.S.A.* 116(44), 22071–22080.
- Nie, R. X., Chin, K. S., Tian, Z. P., et al., 2023. Exploring dynamic effects on classifying service quality attributes under the impacts of COVID-19 with evidence from online reviews. *Int. J. Contemp. Hosp. Manag.* 35(1), 159–185.
- Nie, R. X., Tian, Z. P., Wang, J. Q., et al., 2020. Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory. *Int. J. Hosp. Manag.* 88, 102495.
- Oender, I., 2017. Classifying multi-destination trips in Austria with big data. *Tour. Manag. Perspect.* 21, 54–58.
- Oliver, R. L., Rust, R. T., Varki, S., 1997. Customer delight: Foundations, findings, and managerial insight. *J. Retailing* 73(3), 311–336.
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Quinlan, J. R., 1986. Induction of decision trees. *Mach. Learn.* 1, 81–106.
- Radojevic, T., Stanic, N., Stanic, N., et al., 2018. The effects of traveling for business on customer satisfaction with hotel services. *Tour. Manag.* 67, 326–341.
- Ribeiro, M. T., Singh, S., Guestrin, C., 2016. "Why should I trust you?" Explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, pp. 1135–1144.
- Rita, P., Ramos, R., Borges-Tiago, M. T., et al., 2022. Impact of the rating system on sentiment and tone of voice: A Booking.com and TripAdvisor comparison study. *Int. J. Hosp. Manag.* 104, 103245.
- Schuckert, M., Liu, X., Law, R., 2015. A segmentation of online reviews by language groups: How English and non-English speakers rate hotels differently. *Int. J. Hosp. Manag.* 48, 143–149.
- Shapley, L. S., 1953. Stochastic Games\*. *Proc. Natl. Acad. Sci.* 39(10), 1095–1100.
- Strumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* 41(3), 647–665.
- Vavra, T.G., 1997. *Improving Your Measurement of Customer Satisfaction: A Guide to Creating,*

- Conducting, Analyzing, and Reporting Customer Satisfaction Measurement Program. ASQ Quality Press, Milwaukee, Wis.
- Voss, G. B., Parasuraman, A., Grewal, D., 1998. The roles of price, performance, and expectations in determining satisfaction in service exchanges. *J. Mark.* 62(4), 46–61.
- Wong, J. W. C., Lai, I. K. W., 2018. Evaluating value co-creation activities in exhibitions: An impact-asymmetry analysis. *Int. J. Hosp. Manag.* 72, 118–131.
- Xiao, X., Fang, C., Lin, H., 2020. Characterizing Tourism Destination Image Using Photos' Visual Content. *ISPRS Int. J. Geo-Inf.* 9(12), 730.
- Xu, X., Li, Y., 2016. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *Int. J. Hosp. Manag.* 55, 57–69.
- Yi, J., Oh, Y. K., 2022. The informational value of multi-attribute online consumer reviews: A text mining approach. *J. Retail. Consum. Serv.* 65, 102519.
- Zhang, C., Xu, Z., Gou, X., et al., 2021. An online reviews-driven method for the prioritization of improvements in hotel services. *Tour. Manag.* 87, 104382.
- Zhang, D., Shen, Z., Li, Y., 2023. Requirement analysis and service optimization of multiple category fresh products in online retailing using importance-Kano analysis. *J. Retail. Consum. Serv.* 72, 103253.
- Zhang, M., Zhang, Y., Zhao, L., 2020. What drives online course sales? Signaling effects of user-generated information in the paid knowledge market. *J. Bus. Res.* 118, 389–397.
- Zhou, L., Ye, S., Pearce, P. L., et al., 2014. Refreshing hotel satisfaction studies by reconfiguring customer review data. *Int. J. Hosp. Manag.* 38, 1–10.

## Appendix

### 1. Specific definitions of measurement indicators

1. Explained Variance (Evar):

$$Evar(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

where  $y$  represents the true values of the target variable and  $\hat{y}$  represents the model's predicted values. Evar ranges from 0 to 1, and a value closer to 1 indicates a better-performing model in explaining the target variable's variance.

2. Mean Absolute Error (MAE):

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

where  $n$  is the number of samples,  $y_i$  is the true value of the target variable for the  $i$ th sample, and  $\hat{y}_i$  is the model's predicted value. A smaller MAE value indicates better predictive performance of the model.

3. Mean Absolute Percentage Error (MAPE):

$$MAPE(y, \hat{y}) = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

A smaller MAPE value indicates better predictive performance of the model, but it is sensitive to samples with target values close to zero or small values.

4. Mean Squared Error (MSE):

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

A smaller MSE value indicates better predictive performance of the model.

5. Root Mean Squared Error (RMSE):

Root Mean Squared Error is the square root of the Mean Squared Error. It is similar to MSE.

$$RMSE(y, \hat{y}) = \sqrt{MSE(y, \hat{y})}$$

A smaller RMSE value indicates better predictive performance of the model.

6. R-squared ( $R^2$ ):

R-squared is a metric used to evaluate the goodness of fit of a regression model. It represents the proportion of the variance in the target variable that is explained by the model's predictions. The formula for  $R^2$  is:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$\bar{y}$  is the mean of the target variable.  $R^2$  ranges from 0 to 1, and a value closer to 1 indicates a better-fitting model, while a value closer to 0 suggests that the model does not explain much of the target variable's variance.

## 2. Parameter settings for grid search

grid\_search\_param.xlsx

## 3. The accuracy of the XGBoost model on different datasets

Data	Time	Evar	MAE	MAPE	MSE	RMSE	$R^2$
Test	2007	0.72	0.45	0.15	0.39	0.14	0.72
Test	2008	0.58	0.30	0.07	0.20	0.09	0.58
Test	2009	0.77	0.36	0.11	0.24	0.11	0.77
Test	2010	0.80	0.31	0.10	0.20	0.11	0.80
Test	2011	0.79	0.31	0.09	0.19	0.10	0.79
Test	2012	0.80	0.31	0.09	0.18	0.10	0.80
Test	2013	0.78	0.31	0.09	0.19	0.10	0.78
Test	2014	0.80	0.30	0.09	0.19	0.10	0.80
Test	2015	0.80	0.28	0.09	0.19	0.11	0.80
Test	2016	0.83	0.26	0.08	0.17	0.10	0.83
Test	2017	0.82	0.26	0.08	0.18	0.11	0.82
Test	2018	0.85	0.25	0.08	0.17	0.10	0.85
Test	2019	0.86	0.27	0.09	0.18	0.11	0.86
Test	2020	0.86	0.24	0.10	0.19	0.12	0.86
Test	2021	0.90	0.26	0.11	0.20	0.13	0.90
Test	2022	0.88	0.27	0.11	0.22	0.13	0.88
Test	2023	0.86	0.27	0.10	0.21	0.12	0.86

## 4. The category results of different service attributes at different times

Time		Value	Rooms	Location	Cleanliness	Service	Sleep Quality
2007	AP	3.57	3.65	4.71	4.04	3.68	3.86
	IA	0.56	-0.51	—	-0.35	0.15	0.12
	Type	E	B	—	B	P	P
2008	AP	3.84	3.98	4.69	4.20	3.99	4.10
	IA	-0.48	-0.35	—	0.27	0.61	0.96
	Type	B	B	—	E	E	E
2009	AP	3.89	3.91	4.62	4.31	3.91	4.12
	IA	-0.10	-0.13	0.39	0.43	-0.25	0.24
	Type	P	P	E	E	B	E
2010	AP	4.04	4.08	4.60	4.45	4.16	4.19
	IA	-0.44	0.18	-0.01	-0.10	-0.06	-0.33
	Type	B	P	P	P	P	B
2011	AP	4.02	4.09	4.62	4.41	4.20	4.22
	IA	-0.48	0.13	0.52	0.04	0.00	-0.37
	Type	B	P	E	P	P	B
2012	AP	4.04	4.12	4.65	4.41	4.26	4.27
	IA	-0.43	0.26	0.23	0.07	-0.09	-0.47
	Type	B	E	E	P	P	B
2013	AP	4.07	4.12	4.67	4.42	4.30	4.29
	IA	-0.41	0.31	0.19	-0.19	-0.15	-0.41
	Type	B	E	P	P	P	B
2014	AP	4.11	4.15	4.67	4.44	4.35	4.30
	IA	-0.27	0.05	0.32	-0.26	0.01	-0.51
	Type	B	P	E	B	P	B
2015	AP	4.14	4.22	4.69	4.49	4.45	4.33
	IA	-0.26	0.12	0.21	-0.21	-0.03	-0.26
	Type	B	P	E	B	P	B
2016	AP	4.12	4.22	4.70	4.48	4.42	4.34
	IA	-0.38	0.27	0.23	-0.15	-0.09	-0.39
	Type	B	E	E	P	P	B
2017	AP	4.15	4.29	4.73	4.51	4.44	4.39
	IA	-0.54	0.28	0.64	-0.13	0.04	-0.38
	Type	B	E	E	P	P	B
2018	AP	4.12	4.24	4.72	4.49	4.41	4.36
	IA	-0.32	0.49	-0.34	0.05	-0.07	-0.63
	Type	B	E	B	P	P	B
2019	AP	4.06	4.17	4.71	4.43	4.36	4.31
	IA	-0.41	0.25	0.31	-0.05	0.10	-0.54
	Type	B	E	E	P	P	B
2020	AP	4.33	4.30	4.73	4.53	4.47	4.39
	IA	0.42	0.22	-0.06	-0.34	-0.18	-0.55
	Type	E	E	P	B	P	B
2021	AP	3.93	3.99	4.55	4.20	4.11	4.17
	IA	0.11	0.23	-0.46	-0.36	-0.14	-0.25
	Type	P	E	B	B	P	B
2022	AP	3.94	4.07	4.66	4.34	4.22	4.25
	IA	-0.14	0.21	-0.08	0.23	0.08	-0.53
	Type	P	E	P	E	P	B
2023	AP	4.26	4.25	4.73	4.50	4.35	4.35
	IA	-0.09	0.17	0.01	0.32	0.00	-0.18
	Type	P	P	P	E	P	P

- An Interpretable Machine Learning-based Dynamic Asymmetric Analysis (IML-DAA) approach
- Explore the dynamic asymmetric effects between attribute performance (AP) and customer satisfaction (CS)
- Effect of COVID-19 and time on the relationship between the AP and CS

Journal Pre-proof

**Conflict of Interests**

The authors declare that there is no conflict of interest regarding the publication of this paper.

Journal Pre-proof