Equivalent: Jurnal Ilmiah Sosial Teknik
Volume 8, Issue 3, 787-799
e_ISSN: 2775-0833
https://jurnalequivalent.id/index.php/jequi/index
DOI: doi.org/10.59261/jequi.v8i3.357

AHP-SERVPERF Assessment of Five-Star Hotel Service Quality Using TripAdvisor Reviews in Surabaya

Keisya Tifany*
Universitas Ciputra, Indonesia

Agustinus Nugroho
Universitas Ciputra, Indonesia

Adrie Oktavio
Universitas Ciputra, Indonesia

Fabiola Leoparjo
Universitas Ciputra, Indonesia

***Corresponding author:**
Keisya Tifany, Universitas Ciputra, Indonesia.
✉keisyatifany08@gmail.com

Article Info:
Article history:
Received: May 19, 2026
Revised: June 21, 2026
Accepted: July 02, 2026

Keywords:
SERVPERF; AHP; 2-Tuple Linguistic Model.

Abstract

Background: Conventional SERVPERF-based hotel service quality assessments often rely on questionnaires, which may lead to midpoint bias, limited engagement, and subjective interpretation. Therefore, online customer reviews offer an alternative data source for capturing authentic service experiences.
Objective: This study aims to evaluate service quality in five-star hotels in Surabaya by integrating SERVPERF, Analytical Hierarchy Process (AHP), text mining, sentiment analysis, and the 2-tuple linguistic model.
Methods: A total of 1,200 TripAdvisor reviews from five five-star hotels in Surabaya were collected and analyzed. Expert judgments were obtained through AHP to determine the weight of each SERVPERF dimension. Review texts were segmented, classified into SERVPERF dimensions using a Large Language Model, and analyzed using TextBlob sentiment polarity. The results were then integrated into the 2-tuple linguistic model.
Results: Tangibles was the most frequently discussed and most positively perceived dimension in customer reviews. However, Reliability received the highest expert-based AHP weight and became the most influential dimension after weighted SERVPERF scoring, followed by Responsiveness, Assurance, Empathy, and Tangibles.
Conclusion: The proposed hybrid model provides a more comprehensive and data-driven approach to hotel service quality assessment by combining authentic customer reviews with expert-based weighting. This method helps hotel managers identify priority areas for service improvement more objectively.

To cite this article: Tifany, K., Nugroho, A., Oktavio, A., & Leoparjo, F. (2026). AHP-SERVPERF Assessment of Five-Star Hotel Service Quality Using TripAdvisor Reviews in Surabaya. *Equivalent: Jurnal Ilmiah Sosial Teknik*, 4(3), 787-799. https://doi.org/10.59261/jequi.v8i3.357

INTRODUCTION

The contribution of the tourism sector to Indonesia's national Gross Domestic Product (GDP) in 2024 is estimated to reach approximately 4.01-4.5% (Wisnubroto, 2025). Recognizing this, the Indonesian government has planned to further optimize this sector. As the second- largest city, Surabaya plays a significant role in the national economy, particularly within the tourism and hospitality sectors (Imaduddin, 2025). Surabaya supports tourism development through substantial infrastructure investments and collaborations with hotels, restaurants, and travel agencies (Pemkot Surabaya, 2025).

As part of the tourism industry, the success of the hospitality sector also heavily relies on customer satisfaction. According to Expectation-Confirmation Theory, customer satisfaction is a subjective evaluation resulting from the comparison between customer expectations and actual experiences, reflecting the extent to which products or services consistently meet expectations

(Islam et al., 2026; Oliver, 1980). When experiences align with or surpass expectations, customers feel satisfied. Furthermore, customer satisfaction represents a business approach that focuses on delivering value to customers, proactively understanding and managing customer expectations, and demonstrating the capability and responsibility to meet customer needs (Abate, 2026; Dominici & Guzzo, 2010).

One of the biggest challenges for management in the service sector is attaining and sustaining customer satisfaction (Baharieman et al., 2025). As the hospitality industry is also part of the service sector, customer satisfaction is one of the main indicators used to assess business performance and market positioning (Stephan et al., 2024), since hotels cannot effectively compete without meeting customers' expectations (Ndaguba & Van Zyl, 2026; Radojevic et al., 2015). Therefore, to expand and improve their businesses, hotels must have a clear understanding of the factors that provide greater value to customers. It is thus crucial to pay attention to the quality of service provided by hotels, as this is also a strong predictor of customer satisfaction.

Service quality refers to how well customers' needs are fulfilled and how effectively the delivered services meet customer expectations (Abu-Hasheesh et al., 2026; Lewis & Booms, 1983). Perceived service quality is determined by comparing the service that customers expect with the service they actually experience (Grönroos, 1984; Panesar et al., 2026). With the increasing popularity of online platforms and review sites, service quality is frequently measured through online reviews. SERVPERF is an instrument widely used to measure service quality. SERVPERF is a service quality measurement scale developed by Cronin and Taylor in 1992 (Cronin & Taylor, 1992; Shahi, 2025). This scale is a variant of SERVQUAL, utilizing only the perception component without comparing it with customer expectations (Gupta, 2022; Jain & Gupta, 2004). Several studies have demonstrated that SERVPERF outperforms SERVQUAL in terms of effectiveness and efficiency and has been applied across various industries.

However, nearly all research involving SERVPERF relies on perception-based questionnaires using Likert scales, instead of online customer reviews. This method has several drawbacks. First, questionnaires do not allow researchers to probe deeper into respondents' answers (Nayak & Narayan, 2019; Oladokun et al., 2025). If responses are unclear, too brief, or ambiguous, researchers cannot request further clarification. The same studies also indicate respondents often choose neutral responses (Midpoint Bias), possibly due to uncertainty or a lack of understanding of the questions. Furthermore, there is also a risk of duplicated responses or spamming, where an individual might fill out multiple surveys, manipulating results and thereby reducing data validity. There is no guarantee respondents will complete questionnaires seriously or provide accurate answers (Mishra et al., 2025), and there is significant subjectivity in interpreting questionnaire items. Likert-scale questions may be interpreted differently by respondents, causing inconsistent responses and challenges in ensuring reliable data analysis. Despite the growing body of research on hotel service quality, a significant methodological gap persists. Previous studies have either applied SERVPERF through traditional questionnaire surveys without leveraging the richness of user-generated online content (Alp et al., 2022; Mohebbi et al., 2024), or analyzed online reviews using sentiment analysis alone without integrating expert-based dimensional weighting (Bi et al., 2019; Wu & Yang, 2024). No existing study has simultaneously combined SERVPERF dimensional framework, AHP-based expert weighting, LLM-powered text classification, sentiment polarity analysis, and 2-tuple linguistic aggregation in a single integrated model, particularly in the context of five-star hotels in Surabaya. This gap is critical because it means that the relative importance of service dimensions, as perceived by industry experts, has not been systematically compared against customer-generated evaluations derived from large-scale online review data, potentially leading to misaligned service improvement priorities.

This research addresses these weaknesses by introducing a novel method to measure customer perceptions of hotel service quality by implementing the Analytical Hierarchy Process (AHP) to determine the importance weights of each SERVPERF dimension in hotel service quality. The Analytic Hierarchy Process (AHP) is a widely recognized method for handling complex information in Multi-Criteria Decision Making (MCDM) by structuring problems into a hierarchical framework (Saaty, 1987; Tavana et al., 2023). This research adopts an integrated approach, starting with text mining of customer reviews from TripAdvisor to obtain commentbased data. Subsequently, weights for each SERVPERF dimension are determined through pairwise comparison using questionnaires distributed to hospitality industry experts, including academics, hotel management, and frequent guests of five-star hotels. Afterward, text parsing is conducted to separate individual sentences, followed by text classification to identify whether these sentences contain any of the five SERVPERF dimensions using Large Language Models (LLM) (Kostina et al., 2025); sentences not matching these dimensions are ignored. This process utilizes Python scripts with the assistance of LLM, as such models effectively extract structured data from unstructured text (Schilling-Wilhelmi et al., 2024). Qualified sentences are then analyzed further to determine sentiment polarity scores (-1 to +1) using the TextBlob- based algorithm implemented through Python scripts, chosen for its ease of integration and capability in emotion identification, classifying texts as positive, negative, or neutral based on polarity (Kaur et al., 2024). This approach adopts the 2-tuple linguistic model to prevent information loss and enhance the interpretability of results (Herrera & Martínez, 2000; Shu et al., 2026). Each SERVPERF dimension's average sentiment score is then calculated and multiplied by the corresponding AHP weight (Alp et al., 2022), producing a standardized overall satisfaction score. The final outcome demonstrates that the proposed model effectively converts customer reviews into quantitative representations of the five SERVPERF dimensions, enabling more objective and systematic identification of key factors causing customer dissatisfaction.

TripAdvisor was chosen as the primary data source because of its extensive use and credibility among tourists worldwide, its globally increasing influence in recent years (Filieri et al., 2021), and its recognition as a reliable data source within the hospitality and tourism industries (Qin et al., 2025). This study selected five-star hotels due to their relevance within the hospitality industry and the high research interest associated with them. The dataset collected for this research comprises 1200 reviews (user IDs, months of user comments, and comments) from five different five-star hotels in Surabaya.

Online reviews have become valuable data sources for understanding and capturing customer needs (Bi et al., 2019). Compared to questionnaires, online reviews offer advantages such as larger scale, real-time updates, and richer information content. This research aims to overcome the limitations of conventional customer satisfaction measurement methods, such as questionnaire-based SERVQUAL and SERVPERF, by adopting a novel approach integrating the AHP method and the 2-tuple linguistic model. Utilizing online customer reviews from TripAdvisor, this study seeks to identify and evaluate key factors affecting customer satisfaction in five-star hotels in Surabaya. This approach is expected to provide deeper insights into the factors influencing customer satisfaction and significantly contribute to the development of strategies for service quality improvement in the hospitality industry. The research findings may also serve as a reference for hotel management to formulate policies that are more effective and responsive to customer needs and expectations.

METHOD

This section outlines the fundamental foundations underpinning the proposed model: SERVPERF, the Analytical Hierarchy Process method, and the 2-tuple linguistic model.

SERVPERF

SERVPERF (Service Performance) is the customer's perceived level of service performance, considered the most reflective indicator of service quality in a business (Mohebbi et al., 2024). In this research, the SERVPERF scale was used to measure service perceptions across five main dimensions: Tangibles, referring to physical facilities such as linens, furniture, and hotel amenities; Reliability, defined as the hotel's capability to deliver dependable and timely services, such as room readiness at check-in and the functionality of facilities; Responsiveness, which assessed the promptness and attentiveness of staff in responding to guest requests; Assurance, representing staff competence in building trust, for example by providing recommendations for activities or restaurants; and Empathy, reflecting the attention given to guests' needs and feelings, including the handling of special requests or complaints.

Analytical Hierarchy Process (AHP) Method

In this research, the AHP method was employed to determine the importance weights of each SERVPERF dimension in evaluating hotel service quality. The AHP technique has been widely applied across diverse fields, including human resource management, business management (Claus & Aldianto, 2024) and finance (Nobanee & Al-Suwaidi, 2021; Tavana et al., 2023), healthcare, transportation (Kriswardhana et al., 2025; Park & Lee, 2020), and many more.

AHP begins with defining the primary goal: determining the relative importance of the five SERVPERF dimensions: Tangibles, Reliability, Responsiveness, Assurance, and Empathy. Next, these five dimensions are established as criteria. The weight of each dimension is calculated using pairwise comparisons based on Saaty's 9-point scale (Dash et al., 2023), allowing decision-makers to systematically evaluate different dimensions (Khan et al., 2022). Table 1 illustrates the 9-point Saaty scale used in pairwise comparisons.

Table 1. Saaty's 9-Point Scale

Intensity of importance on an absolute scale	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance of one over another	Experience and judgment strongly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Very strong importance	An activity is strongly favored and its dominance demonstrated in practice
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2,4,6,8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals	If activity *i* has one of the above numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	
Rationals	Ratios arising from the scale	If consistency were to be forced by obtaining *n* numerical values to span the matrix

Linguistic Model 2-Tuple

The 2-tuple linguistic model is based on fuzzy logic and linguistic variables (Zadeh, 1975). This model allows aggregation of linguistic scales, enhancing interpretability (Shu et al., 2023). Introduced to overcome the problem of information loss during the fusion of linguistic data (Herrera & Martínez, 2000; Shu et al., 2026). Numerous studies have adopted this model because it offers results that are more intuitive (Giráldez-Cru et al., 2023) and easier to interpret (Chen & Wang, 2026; Labella et al., 2021; Zhang et al., 2012) compared to results obtained solely through numerical scales.

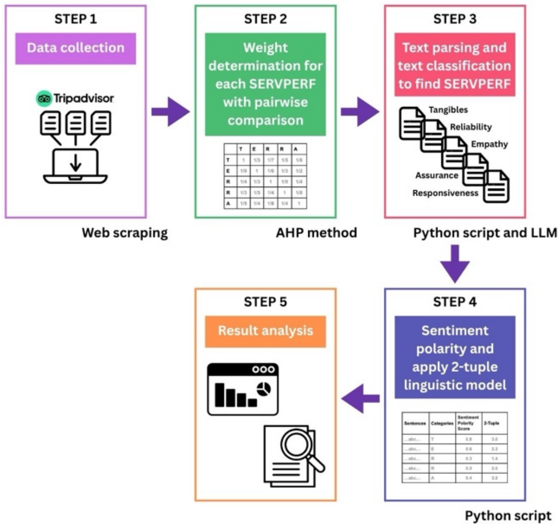

Figure 1. Flow diagram of review extraction to data analysis

Figure 1 illustrates the research method flow. The stages consist of: (1) Data Collection, (2) Weight Determination for each SERVPERF dimension, (3) Text Parsing and Classification, (4) Sentiment Polarity Analysis and Application of the 2-tuple Linguistic Model, and (5) Result Analysis. The details of each stage are explained in the subsequent subsections.

Data Collection via Web Scraping

The first step involved web scraping from the online review site TripAdvisor for 5-star hotels in Surabaya. Web scraping was performed to collect all available information, including images, user identities, review dates, and text comments. For this study, 1,200 online reviews were extracted from five selected 5-star hotels in Surabaya (The Westin Surabaya, Sheraton Surabaya Hotel, Shangri-La Surabaya, JW Marriott Hotel Surabaya, and DoubleTree by Hilton Surabaya), representing approximately 10% of each hotel's total online reviews. However, for the analysis of customer perceptions, text mining was conducted specifically on customer comments. These comments formed the basis for analyzing service quality as perceived naturally by customers. Web scraping was performed using Python's Selenium WebDriver library with Chrome browser automation, targeting the English-language review pages of each hotel. Review selection criteria included: (1) reviews with stay dates between January and December 2024; (2) reviews written in English to ensure consistency in sentiment analysis; and (3) reviews containing substantive textual content (excluding reviews with only star ratings and no accompanying text). Data cleaning involved removing duplicate reviews, eliminating reviews with fewer than 10 words, and standardizing text encoding. The LLM-based classification used GPT-4o with a structured prompt instructing the model to classify each sentence-level unit into exactly one SERVPERF dimension based on predefined keyword anchors and contextual meaning. Classification validation was performed through manual verification of a random 10% sample (approximately 120 review segments), achieving an inter-rater agreement of 89.2% between the LLM classification and human expert coding.

Weight Determination using AHP (Pairwise Comparison)

This step was performed concurrently with Step 1, employing the Analytical Hierarchy Process (AHP) to determine the importance weights for each SERVPERF dimension. This was conducted through a questionnaire distributed to hospitality industry experts (hospitality academics, hotel management professionals, and frequent guests of five-star hotels), who were asked to carry out pairwise comparisons among dimensions, resulting in a comparison matrix. This matrix was then processed to generate the weights for each dimension.

Text Parsing and Classification using LLM

After collecting the comments, the next step was text parsing to segment comments into individual sentences. Then, text classification was performed to determine whether each sentence belonged to one of the SERVPERF dimensions (Tangibles, Reliability, Responsiveness, Assurance, Empathy). This step employed a Python script assisted by a Large Language Model (LLM), specifically GPT, capable of accurately recognizing contextual and semantic classifications.

Sentiment Polarity and 2-Tuple Linguistic Model

The classified sentences were then analyzed for sentiment polarity using TextBlob (Python-based), where sentiments were categorized as positive (+1), neutral (0), and negative (−1). At this stage, the 2-tuple linguistic model was applied to prevent information loss during discretization and to improve interpretability and accuracy in linguistic representation. TextBlob was selected over alternative sentiment analysis tools such as VADER, BERT, or RoBERTa for three reasons: (1) TextBlob provides continuous polarity scores on a [−1, +1] scale, which is directly compatible with the 2-tuple linguistic model's requirement for numerical input; (2) TextBlob has demonstrated adequate performance for review-level sentiment classification in hospitality contexts (Kaur et al., 2024); and (3) its computational efficiency enables processing of large review datasets without requiring GPU infrastructure, making the methodology more reproducible for practitioners. However, transformer-based models such as BERT may offer higher accuracy for nuanced sentiment detection, representing a potential avenue for future methodological refinement.

Result Analysis and SERVPERF Scoring

Once each dimension received its average sentiment score in the 2-tuple format, these scores were multiplied by their respective AHP weights. The formula applied was: "Dimension Score = Average Sentiment × AHP Weight." The final result represented the overall quantitative hotel service quality score derived from real-world qualitative customer data. These results enabled the identification of service aspects that were strongest and weakest, supporting the formulation of service improvement strategies and informed decision-making based on actual customer feedback.

RESULTS AND DISCUSSION

Results

This study analyzed 1,200 TripAdvisor reviews of five five-star hotels in Surabaya with stay dates ranging from January 2024 to December 2024. Of the total reviews, 20.92% contained identifiable user location information.

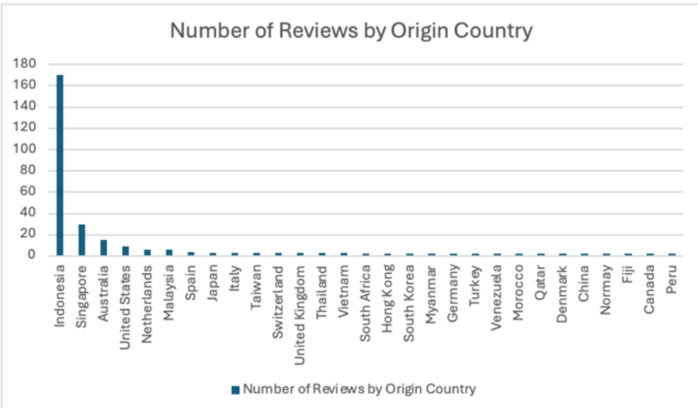

Figure 2. Number of Reviews by Origin Country
Source: Research Data

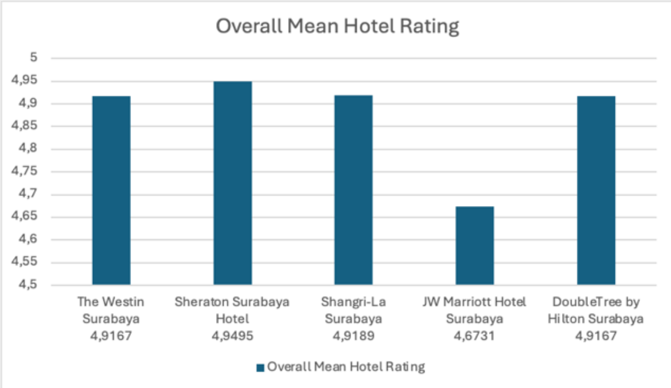

Figure 3. Overall Mean Hotel Rating
Source: Research Data

The descriptive statistics further show that the overall mean hotel rating was 4.9283, with a standard deviation of 0.4062. At the hotel level, DoubleTree by Hilton Surabaya had the highest mean rating (4.9547), followed by Sheraton Surabaya Hotel & Towers (4.9495), Shangri-La Surabaya (4.9189), The Westin Surabaya (4.9167), and JW Marriott Hotel Surabaya (4.6731). These results indicate that the reviewed hotels generally received very high ratings during the observation period.

Table 2. Pairwise Comparison Matrix.

	Tangibles	**Reliability**	**Responsiveness**	**Assurance**	**Empathy**
Tangibles	1	1/5	1/5	1	**1**
Reliability	5	1	5	4	**5**
Responsiveness	5	1/5	1	1	**1**
Assurance	1	1/4	1	1	**1**
Empathy	1	1/5	1	1	**1**

Source: Research Data

To determine the relative importance of SERVPERF dimensions, pairwise comparison questionnaires were completed by three experts representing different stakeholder perspectives. The experts consisted of Mr. LPB, a lecturer at Universitas Ciputra with multiple hospitality-related certifications; Mr. IP, a Front Office Director at a five-star hotel in Surabaya; and Mr. MFS, a university student who frequently stays at five-star hotels in Surabaya and surrounding areas. These experts were intentionally selected to represent academic, practitioner, and customer perspectives, respectively. The individual pairwise judgments were aggregated using the geometric mean before the AHP calculation was performed.

Table 3. Result and Consistency Check

SERVPERF Category	**Average Sentiment Score**	**Number of Sentences**	**AHP Score**	**Final Scores**	**Consistency Check**
Tangibles	3.73	1510	0.089	0.33197	Consistency OK 9%
Reliability	3.53	191	0.518	1.82854	
Responsiveness	3.44	702	0.170	0.5848	
Assurance	3.68	202	0.114	0.41952	
Empathy	3.69	1137	0.109	0.40221	

Source: Research Data

The AHP results showed that Reliability received the highest priority weight (0.518), followed by Responsiveness (0.170), Assurance (0.114), Empathy (0.109), and Tangibles (0.089). These AHP results were considered consistent, as the Consistency Ratio (CR) was ≤ 10% (Kheddache & Djadli, 2025), specifically 9%.

After the review dataset was collected, all review texts were segmented into sentence-level units using GPT-4o. The initial classification process assigned each sentence to one of the five SERVPERF dimensions or to a non-SERVPERF category. During this process, some outputs contained more than one SERVPERF category in a single sentence. These multi-aspect sentences were then cleaned by separating them into shorter phrases, so that each row in the final dataset represented only one phrase with one SERVPERF dimension. This cleaning step produced 3,742 valid categorized units for subsequent analysis.

The distribution of categorized phrases across SERVPERF dimensions. Tangibles had the largest number of categorized phrases (1,510), followed by Empathy (1,137), Responsiveness (702), Assurance (202), and Reliability (191). This shows that customer reviews most frequently referred to physical evidence and personalized attention.

Sentiment classification using GPT-4o and polarity computation using TextBlob were then applied to the cleaned phrase-level dataset. TextBlob polarity scores were rescaled from the original interval of -1 to 1 into a 1 to 5 scale. The resulting average sentiment scores for each SERVPERF dimension were 3.73 for Tangibles, 3.69 for Empathy, 3.68 for Assurance, 3.53 for Reliability, and 3.44 for Responsiveness. These results indicate that all dimensions were evaluated positively overall, although with varying strengths. The sentiment pipeline in the notebook also shows that the cleaned data were dominated by positive sentiment labels, with far fewer neutral and negative statements, confirming an overall favorable review pattern.

When the average sentiment score was multiplied by the AHP weight for each dimension, the weighted SERVPERF results changed substantially. Reliability obtained the highest weighted score (1.82854), followed by Responsiveness (0.58480), Assurance (0.41952), Empathy (0.40221), and Tangibles (0.33197). Therefore, although Tangibles had the highest unweighted average sentiment score and the largest number of categorized phrases, Reliability became the most influential dimension after expert-based importance weighting was applied.

Regarding data validity and reliability, several measures were implemented. First, the review data were filtered to include only reviews with stay dates in 2024, ensuring temporal consistency in the dataset. Second, unnecessary columns were removed during preprocessing to standardize the review data structure. Third, the sentence classification process was followed by a cleaning stage, in which responses containing multiple SERVPERF dimensions were decomposed into phrase-level units with a single category per row. Fourth, AHP consistency testing was used to verify that expert judgments were logically consistent. These steps were intended to reduce noise, improve classification clarity, and strengthen the reliability of the final SERVPERF measurement.

Discussion

The findings reveal an important contrast between customer-generated online reviews and expert judgment. From the perspective of raw review frequency and average sentiment, Tangibles emerged as the most visible and most positively evaluated dimension. It had the largest number of categorized phrases and the highest average sentiment score. This indicates that hotel guests most frequently discuss tangible elements such as room condition, cleanliness, ambiance, amenities, and other visible physical attributes. This pattern is understandable because physical evidence is the most immediate and observable part of hotel service encounters. Guests can easily notice and describe spacious rooms, comfortable beds, clean facilities, or attractive interiors in online reviews.

However, the AHP results show that experts considered Reliability to be the most important dimension in evaluating hotel service quality. After applying the AHP weights, Reliability produced the highest weighted SERVPERF score, even though its average sentiment score and number of categorized phrases were lower than Tangibles, Empathy, and Responsiveness. This result suggests that what customers mention most often is not necessarily what experts consider most critical for service quality performance. From the expert viewpoint, the ability of a hotel to provide services accurately, consistently, and as promised is more fundamental than visible physical aspects.

This difference between expert and review-based perspectives contributes to the gap identified in the Introduction regarding the limitations of questionnaire-only SERVPERF assessment. Conventional SERVPERF studies generally rely on direct perception scoring through

Likert-scale questionnaires. In contrast, this study shows that online reviews capture the spontaneous salience of service attributes from the customer side, while AHP captures normative importance from the expert side. The integration of both produces a richer assessment framework. In other words, online reviews highlight what customers naturally talk about, whereas AHP reveals which dimensions should be prioritized strategically.

The dominance of Tangibles in the review data also reflects the nature of online review behavior. Customers tend to write about aspects that are easiest to observe and easiest to articulate publicly, such as room cleanliness, facility comfort, building aesthetics, and atmosphere. By contrast, Reliability may be less frequently mentioned because it often becomes visible only when service delivery fails, such as when check-in is delayed, requests are not fulfilled, or promised services are inconsistent. This may explain why Reliability appeared less often in the text but still received the highest AHP weight from experts. These findings are consistent with Perdomo-Verdecia et al. (2024), who reported that physical aspects dominate customer satisfaction narratives in hotel reviews, and with Nunkoo et al. (2025), who found that the relative importance of service quality dimensions varies significantly across hotel star ratings. However, the present study extends these findings by demonstrating that expert-weighted priorities diverge substantially from review-frequency patterns a distinction not captured in previous questionnaire-based SERVPERF studies.

The relatively strong performance of Empathy and Assurance in the sentiment scores also indicates that guests positively perceived personal interactions with hotel staff. Friendly, attentive, and professional service remains a central source of positive customer experience in five-star hotels. Meanwhile, Responsiveness obtained the lowest average sentiment score among the five dimensions. Although still positive overall, this suggests that speed of response and prompt service may present more room for improvement compared to the other dimensions. Because Responsiveness also received the second-highest AHP weight, this dimension should be treated as a managerial priority area after Reliability.

These findings provide several managerial implications. First, hotel managers should not rely solely on high online praise for physical facilities as evidence of overall superior service quality. Strong Tangibles performance may enhance first impressions, but operational consistency remains more strategically important. Second, service improvement initiatives should prioritize systems and processes that strengthen Reliability, such as standardization of service delivery, more accurate fulfilment of guest requests, smoother check-in and check-out procedures, and better coordination across departments. Third, hotels should also improve Responsiveness by strengthening staffing readiness, response protocols, and frontline agility in handling guest requests and complaints. In short, the integrated model suggests that hotels should preserve strengths in Tangibles while prioritizing improvement in Reliability and Responsiveness. However, several data limitations should be acknowledged in interpreting these findings. First, TripAdvisor reviews exhibit a well-documented positivity bias, as satisfied guests are more likely to post reviews than dissatisfied ones (Ndaguba & Van Zyl, 2026; Radojevic et al., 2015), which may inflate average sentiment scores across all dimensions. Second, TripAdvisor users represent a specific demographic subset of hotel guests (predominantly international, technology-literate travellers) whose priorities may not fully represent the broader guest population, including domestic business travellers or conference attendees. Third, the overall mean hotel rating of 4.93 out of 5 suggests a ceiling effect that may compress the variance in sentiment scores, making it more difficult to detect meaningful differences between dimensions. Future studies should consider incorporating reviews from multiple platforms (e.g., Google Reviews, Booking.com, Traveloka) to mitigate platform-specific biases.

From a theoretical perspective, this study contributes to the hospitality service quality literature by demonstrating that SERVPERF dimensions can be operationalized using online review data rather than relying only on questionnaires. The use of GPT-assisted text parsing and classification enabled the transformation of unstructured customer narratives into structured SERVPERF categories. The addition of AHP weighting further addressed the issue that not all dimensions contribute equally to perceived service quality. Moreover, the rescaling of sentiment polarity and its integration with the 2-tuple linguistic framework improved interpretability while reducing information loss during linguistic-to-numeric conversion. Thus, this study extends prior

SERVPERF applications by combining text mining, multi-criteria weighting, and linguistic modeling in one analytical framework.

The study also has implications for future research. Since customer-generated reviews and expert-based weighting produced different emphases, future studies may compare results across hotel classes, cities, or platforms such as Google Reviews and Booking.com. Future research may also include larger expert panels, inter-rater validation, or comparisons with traditional SERVPERF questionnaire outcomes to further examine the robustness of this integrated approach. In addition, expanding the model with more advanced sentiment methods or multilingual review analysis may improve precision in contexts where customers write in mixed languages.

CONCLUSION

This study proposed an integrated approach for evaluating hotel service quality in five-star hotels in Surabaya by combining SERVPERF, AHP, text mining, sentiment analysis, and the 2-tuple linguistic model using TripAdvisor reviews. Based on 1,200 online reviews collected from five hotels, the findings show that Tangibles was the most frequently mentioned and the highest-rated dimension in customer reviews, while Reliability was identified as the most important dimension by experts through the AHP weighting process. After combining the average sentiment scores with the AHP weights, Reliability emerged as the most influential dimension, followed by Responsiveness, Assurance, Empathy, and Tangibles. These results indicate that customer-generated online reviews and expert judgments provide different but complementary perspectives on hotel service quality. On the one hand, customers tend to emphasize visible and directly experienced aspects, particularly physical facilities and hotel ambience. On the other hand, experts assign greater importance to the consistency and dependability of service delivery. Therefore, relying solely on either questionnaire-based expert judgment or online customer reviews may produce an incomplete assessment. The proposed hybrid model offers a more comprehensive framework by integrating authentic customer feedback with structured expert evaluation. From a practical perspective, the results suggest that hotel managers should not only maintain excellence in physical facilities, but also place stronger managerial focus on operational consistency and prompt service delivery. In this regard, Reliability and Responsiveness should become priority areas for service improvement, while Tangibles should continue to be preserved as a major source of positive customer perception. Academically, this study demonstrates that SERVPERF can be extended beyond conventional questionnaire-based measurement through the integration of online review analytics, AHP weighting, and linguistic-based sentiment modeling.

Nevertheless, this study has several limitations. First, the analysis was limited to TripAdvisor reviews of five-star hotels in Surabaya within one observation period. Second, the AHP weighting process relied on only three experts representing academic, practitioner, and customer perspectives. Although these experts were deliberately selected to provide diverse viewpoints, the number remains relatively small when compared to the scale of the online review dataset, which consists of 1,200 reviews and thousands of classified phrases from many different customers. This creates an imbalance between expert-based weighting and large-scale customer-generated data. Therefore, future studies are recommended to involve a larger and more diverse panel of experts in the AHP stage so that the resulting weights can better represent broader stakeholder perspectives and reduce the gap between expert judgment and large-scale online review evidence. Further research may also expand the analysis to other hotel categories, cities, and review platforms, as well as compare the results with conventional SERVPERF questionnaire findings to strengthen the generalizability of the proposed approach. The primary methodological contribution of this study lies in demonstrating that SERVPERF dimensions can be operationalized through automated analysis of unstructured online review text, bypassing the limitations of traditional survey-based measurement while preserving the theoretical rigor of the SERVPERF framework. The integration of AHP expert weighting with LLM-classified review data represents a novel hybrid approach that enables simultaneous comparison of customer perception patterns with expert-derived dimensional priorities.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to Universitas Ciputra for its academic support throughout this research. Appreciation is also extended to the hospitality experts, hotel practitioners, and frequent hotel guests who contributed valuable judgments during the AHP assessment process, as well as to the editorial team and reviewers for their constructive feedback.

AUTHOR CONTRIBUTION STATEMENT

Keisya Tifany conceptualized the study, collected and processed the TripAdvisor review data, conducted text mining and sentiment analysis, and drafted the manuscript. Agustinus Nugroho, Adrie Oktavio, and Fabiola Leoparjo contributed to the research design, methodological validation, data interpretation, manuscript revision, and final approval of the submitted version.

REFERENCES

Abate, A. (2026). The Effect of Tourism 4" P" S on Customer Satisfaction: The Case of Hotels in Dessie and Kombolcha Cities. *International Journal of Economics & Business Administration*, *14*(2), 3.

Abu-Hasheesh, S. J., Dandis, A., Eid, M. A. H., Alzyoud, S., & Al Khasawneh, M. (2026). The influence of service quality and marketing ethics on relationship marketing outcome in private hospitals. *Journal of Marketing Communications*, *32*(3), 254–281.

Alp, S., Yilmaz, F., & Geçici, E. (2022). Evaluation of the Quality of Health and Safety Services with SERVPERF and Multi-attribute Decision-Making Methods. *JOSE: International Journal of Occupational Safety and Ergonomics*, *28*(4), 2216–2226. https://doi.org/10.1080/10803548.2021.1984711

Baharieman, N. S. N. M., Azwan, L. H. S., Ngah, H. C., Nawawi, W. N. W., & Ishak, F. A. C. (2025). The Elements of Customer Relationship Management and Their Impact on Customer Satisfaction: A Study of Five-Star Hotels in Malaysia. *Information Management and Business Review*, *17*(1), 24–35. https://doi.org/10.22610/imbr.v17i1(I)S.4423

Bi, J.-W., Liu, Y., Fan, Z.-P., & Zhang, J. (2019). Wisdom of Crowds: Conducting Importance-Performance Analysis (IPA) through Online Reviews. *Tourism Management*, *70*(2), 460–487. https://doi.org/10.1016/j.tourman.2018.09.010

Chen, W., & Wang, L. (2026). A Behavioral and Relational Integrated Decision-Making Method Using 2-Tuple Linguistic Neutrosophic Sets: College Music Majors Teaching Quality Evaluation. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, *20*(1), 1–26.

Claus, G., & Aldianto, L. (2024). Application of the Analytical Hierarchy Process (AHP) for Innovative Technological Projects Evaluation and Prioritization in an FMCG Company. *International Journal of Current Science Research and Review*, *7*(6), 3612–3622. https://doi.org/10.47191/ijcsrr/V7-i6-10

Cronin, J. J., & Taylor, S. A. (1992). Measuring Service Quality: A Reexamination and Extension. *Journal of Marketing*, *56*(3), 55–68. https://doi.org/10.2307/1252296

Dash, N. S., Das, S., & Selvaraj, A. (2023). Revisiting the Concept of “Questionnaire” Used in Linguistic Field Surveys. *Indian Journal of Applied Linguistics*, *49*(1–2), 7–39.

Dominici, G., & Guzzo, R. (2010). Customer Satisfaction in the Hotel Industry: A Case Study from Sicily. *International Journal of Marketing Studies*, *2*(2), 3–12. https://doi.org/10.5539/ijms.v2n2p3

Filieri, R., Acikgoz, F., Ndou, V., & Dwivedi, Y. (2021). Is TripAdvisor Still Relevant? The Influence of Review Credibility, Review Usefulness, and Ease of Use on Consumers’ Continuance Intention. *International Journal of Contemporary Hospitality Management*, *33*(1), 199–223. https://doi.org/10.1108/IJCHM-05-2020-0402

Giráldez-Cru, J., Chica, M., & Cordón, O. (2023). An Integrative Decision-Making Mechanism for Consumers’ Brand Selection using 2-Tuple Fuzzy Linguistic Perceptions and Decision Heuristics. *International Journal of Fuzzy Systems*, *25*, 59–79. https://doi.org/10.1007/s40815-022-01385-x

Grönroos, C. (1984). A Service Quality Model and its Marketing Implications. *European Journal of Marketing*, *18*(4), 36–44. https://doi.org/10.1108/EUM0000000004784

Gupta, K. (2022). Re-Examining The Service Quality Measurement Tools–Servqual And Servperf. *Jamshedpur Research Review*, *1*(50), 1–184.

Herrera, F., & Martínez, L. (2000). A 2-Tuple Fuzzy Linguistic Representation Model for Computing with Words. *Brief Papers: IEEE Transactions on Fuzzy Systems*, *8*(6), 746–752.

Imaduddin, N. A. (2025). *Infrastruktur Surabaya Majukan Pariwisata dan Perhotelan*. ANTARA Jatim.

Islam, M. A., Somu, S., & Karatepe, O. M. (2026). Antecedents and consequences of quiet quitting: A systematic literature review. *The Service Industries Journal*, 1–36.

Jain, S. K., & Gupta, G. (2004). Measuring Service Quality: Servqual vs. Servperf Scales. *Vikalpa The Journal for Decision Makers*, *29*(2), 25–37. https://doi.org/10.1177/0256090920040203

Kaur, G., Kaur, A., Khurana, M., & Damasevicius, R. (2024). Sentiment Polarity Analysis of Love Letters: Evaluation of TextBlob, Vader, Flair, and Hugging Face Transformer. *Computer Science and Information Systems*, *21*(4), 1411–1433. https://doi.org/10.2298/CSIS240328040K

Khan, A. W., Khan, S. U., Alwageed, H. S., Khan, F., Khan, J., & Lee, Y. (2022). AHP-Based Systematic Approach to Analyzing and Evaluating Critical Success Factors and Practices for Component-Based Outsourcing Software Development. *Mathematics*, *10*(21), 3982. https://doi.org/10.3390/math10213982

Kheddache, F., & Djadli, S. (2025). A Proposed AHP-SERVQUAL Model for Hotel Service Quality Management: Evidence from Algeria. *JPM: Journal of Perspectives in Management*, *9*(e264674). https://doi.org/10.51359/2594-8040.2025.264674

Kostina, A., Dikaiakos, M. D., Stefanidis, D., & Pallis, G. (2025). Large Language Models for Text Classification: Case Study and Comprehensive Review. *Cornell University: Computer Science-Computation and Language*, 1–12. https://doi.org/10.48550/arXiv.2501.08457

Kriswardhana, W., Toaza, B., Esztergár-Kiss, D., & Duleba, S. (2025). Analytic Hierarchy Process in Transportation Decision-Making: A Two-Staged Review on The Themes And Trends of Two Decades. *Expert Systems with Applications*, *261*(125491). https://doi.org/10.1016/j.eswa.2024.125491

Labella, Á., Dutta, B., & Martinez, L. (2021). An Optimal Best-Worst Prioritization Method under a 2-tuple Linguistic Environment in Decision Making. *Computers & Industrial Engineering*, *155*(301–357), 107141. https://doi.org/10.1016/j.cie.2021.107141

Lewis, R., & Booms, B. (1983). *The Marketing Aspects of Service Quality: in Emerging Perspectives on Service Marketing*. American Marketing Association.

Mishra, L., Sharma, M., Singh, N. R., Dash, G., Misra, S. R., Sokolowski, K., Kumar, M., Das, R., Behera, S. K., & Lapinska, B. (2025). Oral Health Status, Behavior, and Knowledge of Patients with Cardiovascular Disease and Associated Risk Factors in Odisha: A Cross-Sectional Survey. *Dentistry Journal*, *13*(9), 401.

Mohebbi, H., Meybodi, A., & Eslami, S. (2024). Evaluation of Service Quality of Clinical Diagnostic Using the SERVQUAL and SERVPERF Models in Two Simple and Weighted Types. *Management Strategies in Health System*, *9*(1), 71–86. https://doi.org/10.18502/mshsj.v9i1.15785

Nayak, M., & Narayan, K. A. (2019). Strengths and weaknesses of online surveys. *Technology*, *6*(7), 0837–2405053138.

Ndaguba, E. A., & Van Zyl, C. (2026). The impact of sustainability practices on customer satisfaction and loyalty in hospitality: A systematic review. *International Journal for Tourism, Archeology & Hospitality (IJTAH)*, *6*(1).

Nobanee, H., & Al-Suwaidi, N. A. (2021). Performance Evaluation Using the Analytic Hierarchy Process. *Academy of Accounting and Financial Studies Journal*, *25*(2).

Nunkoo, R., Sharma, A., So, K. K. F., Hu, H., & Alrasheedi, A. F. (2025). Two decades of research on customer satisfaction: future research agenda and questions. *International Journal of Contemporary Hospitality Management*, *37*(5), 1465–1496.

Oladokun, B. D., Ajani, Y. A., & Tella, A. (2025). Traditional Surveys Versus Digital Surveys: Perspectives of Library and Information Science Researchers on Best Method for Data Collection in Research. *Folia Toruniensia*, *25*, 13–37.

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction

Decisions. *JSTOR: Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.2307/3150499

Panesar, A., Kakkar, A., Sood, R., & Santos, J. D. (2026). Understanding Q-commerce adoption: the role of technological, social, marketing stimuli, and service quality on usage intention. *European Journal of Management Studies*, 1–20.

Park, S. U., & Lee, J. D. (2020). Analyzing Factors that Influence the Efficiency of Airport Passenger Terminal by Using AHP (Analytic Hierarchy Process) Method (Case Study of Juanda Airport-Surabaya, Indonesia). *KSCE Journal of Civil Engineering*, *24*(1), 3856–386. https://doi.org/10.1007/s12205-020-0880-x

Pemkot Surabaya. (2025). *Pemkot Surabaya Beber Konsep Pengembangan Wisata dan Ekraf Tahun 2025*. Pemerintah Kota Surabaya.

Perdomo-Verdecia, V., Garrido-Vega, P., & Sacristán-Díaz, M. (2024). An fsQCA analysis of service quality for hotel customer satisfaction. *International Journal of Hospitality Management*, *122*, 103793.

Qin, Y., Luo, C., & Ngai, E. W. T. (2025). Deconstructing Customer Satisfaction Recipes: A Dynamic Configurational Framework Leveraging the Power of Online Reviews in Tourism Contexts. *Tourism Management*, *110*, 105181. https://doi.org/10.1016/j.tourman.2025.105181

Radojevic, T., Stanisic, N., & Stanic, N. (2015). Ensuring Positive Feedback: Factors That Influence Customer Satisfaction in the Contemporary Hospitality Industry. *Tourism Management*, *51*, 13–21. https://doi.org/10.1016/j.tourman.2015.04.002

Saaty, R. W. (1987). The Analytic Hierarchy Process — What it is and How it is Used. *Mathematical Modelling*, *9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Schilling-Wilhelmi, M., Ríos-García, M., Shabih, S., Gil, M. V., Miret, S., Koch, C. T., Márquez, J. A., & Jablonka, K. M. (2024). From Text to Insight: Large Language Models for Materials Science Data Extraction. *Cornell University: Condensed Matter-Materials Science*, 1–55. https://doi.org/10.48550/arXiv.2407.16867

Shahi, P. (2025). Measuring Service Quality in the Travel and Tourism Industry in Nepal. *Nepalese Journal of Economics*, *9*(2), 83–94.

Shu, Z., Galán Hernández, J. J., & Carrasco-Aguilar, Á. (2026). Profiling the application of the 2-tuple linguistic model in marketing: a comprehensive analysis and future directions. *Fuzzy Optimization and Decision Making*, *25*(2), 285–313.

Shu, Z., González, R. A. C., García-Miguel, J. P., & Sánchez-Montañés, M. (2023). Clustering using ordered weighted averaging operator and 2-tuple linguistic model for hotel segmentation: the case of TripAdvisor. *Expert Systems with Applications*, *213*(A), 118922. https://doi.org/10.1016/j.eswa.2022.118922

Stephan, U., Strauss, K., Gorgievski, M. J., & Wach, D. (2024). How Entrepreneurs Influence Their Employees' Job Satisfaction: The Double-Edged Sword of Proactive Personality. *Journal of Business Research*, *174*(114492), 1–21. https://doi.org/10.1016/j.jbusres.2023.114492

Tavana, M., Soltanifar, M., & Santos-Arteaga, F. J. (2023). Analytical hierarchy process: Revolution and evolution. *Annals of Operations Research*, *326*(2), 879–907.

Wisnubroto, K. (2025). *Geliat Sektor Pariwisata Pacu Pertumbuhan Ekonomi*. Indonesian GOID.

Wu, J., & Yang, T. (2024). Has COVID-19 Changed Consumers' Satisfiers and Dissatisfiers? Evidence from Online Reviews of 5-Star Hotels in Shanghai and Beijing. *Journal of China Tourism Research*, *20*(2), 430–452.

Zadeh, L. A. (1975). The Concept of a Linguistic Variable and its Application to Approximate R`easoning-I. *Information Sciences*, *8*(3), 199–249. https://doi.org/10.1016/0020-0255(75)90036-5

Zhang, Y., Ma, H., Liu, B., & Liu, J. (2012). Group Decision Making with 2-tuple Intuitionistic Fuzzy Linguistic Preference Relations. *Soft Computing*, *16*, 1439–1446. https://doi.org/10.1007/s00500-012-0847-z