

ASSESSING SERVICE QUALITY BASED ON CUSTOMER REVIEWS USING MACHINE LEARNING TECHNIQUES: A CASE STUDY OF THE HOSPITALITY SECTOR

Marwan Alabed Abu-Zanona^{1,*}, Jehad Abdallah Atieh Afaneh²

¹ Department of Management Information Systems, King Faisal University, Kingdom Saudi Arabia, e-mail: mabozanoneh@kfu.edu.sa, ID: <https://orcid.org/0009-0004-9969-5048>

² Management Department, College of Business Administration, King Faisal University, Al-Ahsa, Saudi Arabia, email: jafaneh@kfu.edu.sa, ID: <https://orcid.org/0000-0002-7231-9032>

*Correspondence: mabozanoneh@kfu.edu.sa

Abstract

This study assesses service quality in the hospitality sector by mining and analysing customer hotel reviews with machine learning techniques. It aims to identify the key service-quality dimensions that drive guest satisfaction and to provide hotel managers with actionable, data-driven insights. We compiled a dataset of 28,945 English-language hotel reviews. Natural language processing and sentiment-analysis methods were applied to extract textual features and polarity scores. Topic modelling was used to surface latent themes, and two supervised classification models—support vector machines (SVM) and random forests—were trained to predict sentiment labels and highlight critical service dimensions. The analysis revealed four principal dimensions of service quality: cleanliness, staff behaviour, amenities, and value for money. Both SVM and random-forest classifiers achieved robust predictive performance (accuracy > 85 %), validating the suitability of a multi-method framework. Unlike prior work that focuses narrowly on sentiment analysis or small samples, this study leverages a large review corpus and integrates sentiment analysis, topic modelling, and supervised learning into a unified framework. It offers a scalable approach and discusses potential extensions—such as deep-learning architectures, multilingual analysis, and real-time monitoring systems—to further advance data-driven service-quality assessment.

Keywords: Service quality, hospitality, sentiment analysis, natural language processing, machine learning.

1 Introduction

Introduction presents the scientific problem of the article, its novelty, exploration of the problem, aim, objective, research methods).

Ensuring exceptional customer satisfaction is fundamental to achieving sustainable success in today's fiercely competitive hospitality industry. The advent of digital transformation has revolutionized how guests share their experiences: online review platforms now function as dynamic repositories of customer feedback, generating vast volumes of unstructured text that capture real-time perceptions and sentiments [1].

Traditional service-quality assessments—primarily structured surveys and questionnaires—have provided valuable insights but are hampered by low response rates, delayed feedback, and biases inherent in self-reporting, offering only a static snapshot of guest opinions [2]. In contrast, user-generated reviews emerge continuously from a diverse clientele and tend to reflect more candid, authentic experiences [3]. This shift from periodic surveys to a perpetual stream of feedback necessitates advanced analytical techniques capable of processing large-scale, unstructured datasets.

Machine learning (ML) combined with natural language processing (NLP) has emerged as an indispensable toolkit in this context. These methods enable researchers and practitioners to automatically sift through tens of thousands of reviews to extract actionable insights unattainable via manual analysis. Sentiment analysis, for example, quantifies the emotional tone of customer comments by classifying reviews as positive, neutral, or negative—providing hotel managers with immediate indicators of overall guest satisfaction and areas for improvement [4].

Complementing sentiment analysis, topic-modeling techniques such as Latent Dirichlet Allocation (LDA) uncover latent themes within large text corpora, revealing recurring service-quality dimensions—cleanliness, staff behavior, amenities, and value for money [5]. Furthermore, supervised learning algorithms—specifically support vector machines (SVM) and random forests—trained on labeled datasets deliver robust performance in sentiment classification tasks, consistently achieving high accuracy in predicting customer sentiment [6].

This study leverages a comprehensive corpus of 28,549 English-language hotel reviews, forming a robust foundation for analysis. By integrating sentiment analysis, topic modeling, and supervised classification into a unified framework, we aim to provide a scalable, multi-method approach that yields both quantitative and qualitative insights. The practical implications are significant: automated analysis empowers hotel managers to promptly identify service-quality deficiencies, track sentiment trends over time, and prioritize strategic improvements to enhance guest satisfaction and maintain competitive advantage.

Our research contributes to the advancement of data-driven service-quality assessment in hospitality by overcoming the limitations of traditional survey methods and offering real-time, actionable insights. This framework supports evidence-based decision making and continuous operational excellence in the hospitality sector.

2 Research Note Format and Literature Review

2.1 Service Quality Assessment in Hospitality

Assessing service quality has been a central concern for both scholars and practitioners in hospitality, given its direct influence on customer satisfaction and business performance. Early frameworks, most notably SERVQUAL, conceptualize service quality as the gap between customer expectations and perceived performance across five dimensions: reliability, responsiveness, assurance, empathy, and tangibles. Despite its widespread adoption, SERVQUAL and analogous survey-based approaches are criticized for being time-consuming, susceptible to response bias, and unable to capture the dynamic nature of guest experiences in real time.

With the proliferation of digital platforms, user-generated online reviews have emerged as a more immediate and authentic source of feedback. These reviews offer near-real-time insights from a broad customer base but introduce challenges related to the high volume and unstructured format of textual data—challenges that necessitate advanced analytic methods such as machine learning (ML) and natural language processing (NLP).

2.2 Sentiment Analysis Techniques

Sentiment analysis, a core subfield of NLP, automatically categorizes text into positive, neutral, or negative sentiment, providing hotel managers with rapid indicators of guest satisfaction. Traditional lexicon-based methods (e.g., VADER) are valued for

interpretability, but ML-based classifiers—including support vector machines (SVM), random forests, and deep neural networks—consistently outperform rule-based approaches in large-scale sentiment classification tasks. For example, [7] demonstrated that deep learning architectures surpass lexicon methods in capturing nuanced sentiment trends in hotel reviews.

2.3 Topic Modeling in Service Quality Research

To complement polarity detection, topic modeling techniques such as Latent Dirichlet Allocation (LDA) uncover latent themes within extensive text corpora. Applied to hospitality reviews, LDA reveals recurring service-quality dimensions—cleanliness, staff behavior, amenities, and value for money—that guide operational improvements. [8] applied LDA to TripAdvisor reviews to identify staff behavior and room maintenance as pivotal satisfaction drivers, while [9] highlighted pricing and reservation issues as common pain points. Visualization tools (e.g., word clouds) further enhance interpretability and managerial actionability.

2.4 Supervised Machine Learning for Sentiment Classification

Supervised ML algorithms—particularly SVM and random forest classifiers—are trained on labeled datasets where each review is annotated by sentiment. These models excel at detecting complex, non-linear patterns in text and deliver high predictive accuracy. In comparative analyses, random forests often outperform alternatives such as naïve Bayes and decision trees, owing to their robustness against overfitting and capacity to handle large feature sets. Hybrid frameworks that integrate sentiment analysis with explanatory analytics have also been explored to enrich understanding of the attributes driving customer sentiment.

2.5 Evolution of Data-Driven Approaches in Hospitality

The integration of ML and NLP marks a paradigm shift from retrospective, survey-based assessments to real-time, data-driven service-quality monitoring. Recent advancements—including transformer-based models such as BERT—offer even greater accuracy in interpreting textual data, enabling proactive identification of service issues and predictive analytics for operational planning.

Despite this extensive body of work, many studies remain confined to single-method approaches (e.g., sentiment analysis alone) or are based on limited datasets, which constrain the generalizability of their findings. To bridge this gap, our research employs a large corpus of 28,549 hotel reviews and integrates sentiment analysis, topic modeling, and supervised classification into a unified, scalable framework. This multi-method approach generates robust, actionable insights for hotel managers, advancing the field of data-driven service-quality assessment and setting the stage for future exploration of deep learning, multilingual analysis, and real-time monitoring systems.

2.6 Literature Review

The Research on automatically extracting guest experience signals from online reviews has advanced rapidly over the last five years. Early work established the utility of user-generated reviews as complementary to traditional survey-based service-quality metrics

[10]; recent studies have shifted toward fine-grained, aspect-based analyses that link sentiment to concrete service attributes (cleanliness, staff behavior, amenities, value, etc.) so managers can prioritize operational fixes [11]. Rule-based tools such as VADER remain in use for simple polarity scoring, but modern research increasingly relies on deep contextual models and ABSA pipelines to capture multi-aspect and mixed-polarity sentences that legacy methods miss [12].

Aspect-based sentiment analysis (ABSA) has become the dominant paradigm for hospitality review mining because it directly maps opinions to service attributes. Transformer-based models (BERT-family and derivatives) and hybrid deep architectures outperform traditional classifiers in both aspect extraction and polarity classification on hotel/OTA corpora. [13] demonstrate that contextual transformers such as BERT and ERNIE significantly improve hotel review polarity classification over classical methods, and that ERNIE provides more stability on Chinese OTA data. Several works apply RoBERTa and other optimized transformer variants to large TripAdvisor corpora, showing marked gains in classification and topic coherence when fine-tuned on domain data [13]. At the same time, ensemble and hybrid approaches (for example combining BERT embeddings with tree-based classifiers or CNN/LSTM stacks) remain competitive, especially when downstream tasks include recommendation or price/demand prediction [14][15].

The hospitality literature also documents a surge in domain-specific datasets and benchmarks that make ABSA research reproducible for hotels and short-term rentals. Several hospitality-focused corpora and challenge-style datasets [16] support consistent evaluation and have enabled head-to-head comparisons of ABSA models tailored to the sector [17]. Relatedly, Andreou, Tsapatsoulis, and Anastasopoulou [18] provide a labeled hotel-review dataset designed for ABSA and topic modeling which many studies now use for training and evaluation. The availability of these benchmark datasets has accelerated comparative research — from classical LDA + SVM pipelines to modern graph-enhanced and transformer-based networks.

Two technical directions stand out in recent ABSA work. First, graph and knowledge-informed models (graph convolutional networks/graph attention networks and hybrid graph+transformer architectures) improve aspect–opinion linking by leveraging syntactic dependency structure and external knowledge; such approaches yield better aspect extraction and polarity assignment than purely sequential models [19]. Second, multimodal and few-shot ABSA addresses real-world constraints (images accompanying reviews, low-resource languages, small labeled sets). Recent few-shot multimodal frameworks (e.g., FMCF) and generative multimodal prompt methods extend ABSA to scenarios with sparse annotations and cross-modal signals, which is highly relevant when OTA listings include photos or short video content that signals product/room conditions beyond text [20].

Cross-lingual and low-resource research is also maturing: Indonesian-language hotel review studies [21] and graph-based semi-supervised ABSA work demonstrate practical pipelines for languages with limited labeled corpora, combining word embeddings, CNN/LSTM stacks and graph labeling techniques to attain high F1 on aspect and polarity tasks [22]. These contributions show how semi-supervised labeling, domain-specific pretraining, and language-specific morphological processing can close the gap with English-centric datasets.

From an applications perspective, online reviews are being integrated beyond sentiment dashboards: models now feed recommendation engines, demand forecasting, and revenue management. Zhang and Niu [23] show that review-derived features (sentiment, topic signals) can materially improve hotel demand forecasting with

attention-equipped LSTM+CNN models; Ray, Garain, and Sarkar [24] combine ABSA outputs with ensemble recommenders to personalize hotel suggestions — both illustrate how ABSA outputs move from descriptive analytics to operational decision support. These studies highlight an important practical trend: ABSA models must now be judged not only by classification metrics but by business impact (forecasting accuracy, recommendation relevance).

Newer generative language models and prompt-based methods (including experiments with ChatGPT) have been explored for aspect extraction and summarization of complaint reviews; these approaches show promise for rapid aspect summarization, but also raise questions about reproducibility, hallucination, and the need for careful prompt design and evaluation in operational hotel settings [25]. In parallel, work on negative-review focused studies (e.g., Airbnb/hospitality negative reviews) provides insights on service failure types and remedy prioritization — useful when pairing ABSA outputs with root-cause analysis for operations [26].

Gaps & opportunities. Three consistent gaps motivate the present study and shape its methodological choices: (1) many ABSA pipelines still struggle with mixed-polarity sentences and implicit aspects (a motivation for using contextual transformers and dependency-aware modules; [27]; (2) sector-specific datasets and multilingual/low-resource pipelines require tailored pretraining and semi-supervised labelling to obtain robust aspect extraction [28]; and (3) evaluation should include downstream business metrics (e.g., demand forecasting / recommender impact) when ABSA is used as a decision tool rather than an isolated classifier [29]. Our study's combination of addresses these gaps by (a) leveraging contextual embeddings for mixed polarity, (b) using semisupervised/label-propagation steps to expand aspect labels in limited datasets, and (c) validating model outputs against service-quality indicators — which aligns with the direction of the recent literature [30][31].

In short, the literature from the last five years converges on ABSA enriched by transformers, graph/knowledge structure, multimodal and few-shot methods, and evaluation that ties textual signals to real operational outcomes. The recent studies cited above provide the most relevant findings, datasets and architectural patterns that directly motivated the design choices in this paper.

3 Research Methodology

This study employs a structured, multi-stage research design to evaluate service quality in the hospitality sector through automated analysis of customer reviews. The methodology comprises five key phases: data acquisition, text preprocessing, sentiment analysis, topic modeling, and model validation. Figure 1 illustrates the overall workflow.

3.1 Data Source and Sample

We analyzed a corpus of 28,549 English-language hotel reviews sourced from the Datafinity Kaggle repository. Each record includes the review text, a numerical rating (1–5 stars), hotel identifier, review date, and basic metadata (location, reviewer profile). The large sample size and diverse geographic coverage enhance the statistical reliability and generalizability of our findings, addressing limitations of prior studies that used smaller, context-restricted datasets [32].

3.2 Text Preprocessing

Raw reviews were converted into a standardized format to enable robust analysis. Our pipeline comprised:

- **Cleaning:** Removal of punctuation, numerals, and non-alphabetic characters; lowercasing and whitespace normalization.
- **Tokenization and Lemmatization:** Segmentation into word tokens and reduction to base forms (e.g., “running” → “run”), ensuring consistency across variants.
- **Vectorization:** Transformation into numerical feature vectors using Term Frequency–Inverse Document Frequency (TF-IDF), which highlights discriminative terms while down-weighting ubiquitous words [33].

3.3 Analytical Framework

1. Sentiment Analysis

We combined lexicon-based and supervised approaches. Initial polarity scores were obtained using a validated sentiment lexicon (e.g., VADER) to classify each review as positive, neutral, or negative [34]. Subsequently, two machine-learning classifiers—Support Vector Machines (SVM) and Random Forest (RF)—were trained on labeled data (derived from star ratings and manual checks) to refine sentiment predictions and capture nuanced linguistic patterns [34]. Model performance was assessed via accuracy, precision, recall, and F₁-score.

2. Topic Modeling

To uncover underlying service-quality dimensions, we applied Latent Dirichlet Allocation (LDA) to the TF-IDF corpus. The optimal number of topics was determined by coherence metrics and manual interpretability. Emergent themes—such as cleanliness, staff behavior, amenities, and value for money—were labeled based on the highest-probability terms in each topic distribution [34]

3.4 Model Validation and Evaluation

- **Data Partitioning:** The dataset was split into 80% training and 20% testing subsets to guard against overfitting and to provide an unbiased evaluation of model generalizability.
- **Hyperparameter Tuning:** Grid-search cross-validation optimized key parameters (e.g., SVM kernel and RF tree depth), maximizing performance on the validation folds.
- **Evaluation Metrics:** Classification models were evaluated using confusion matrices alongside accuracy, precision, recall, and F₁-score. Topic coherence scores and qualitative examination of top-terms ensured the relevance and clarity of extracted themes.
- **Temporal Analysis:** Sentiment distributions were analyzed over monthly intervals to detect shifts in customer perceptions, enabling managers to link observed trends to operational changes or external events .

All analyses were conducted in Python using standard open-source libraries to ensure reproducibility (e.g., Scikit-learn for ML, Gensim for LDA, NLTK/SpaCy for text processing). This methodological framework (as shown in Figure 1.) delivers both quantitative measures of overall satisfaction and qualitative insights into specific service

attributes, equipping hotel managers with actionable, evidence-based guidance for enhancing guest experience.

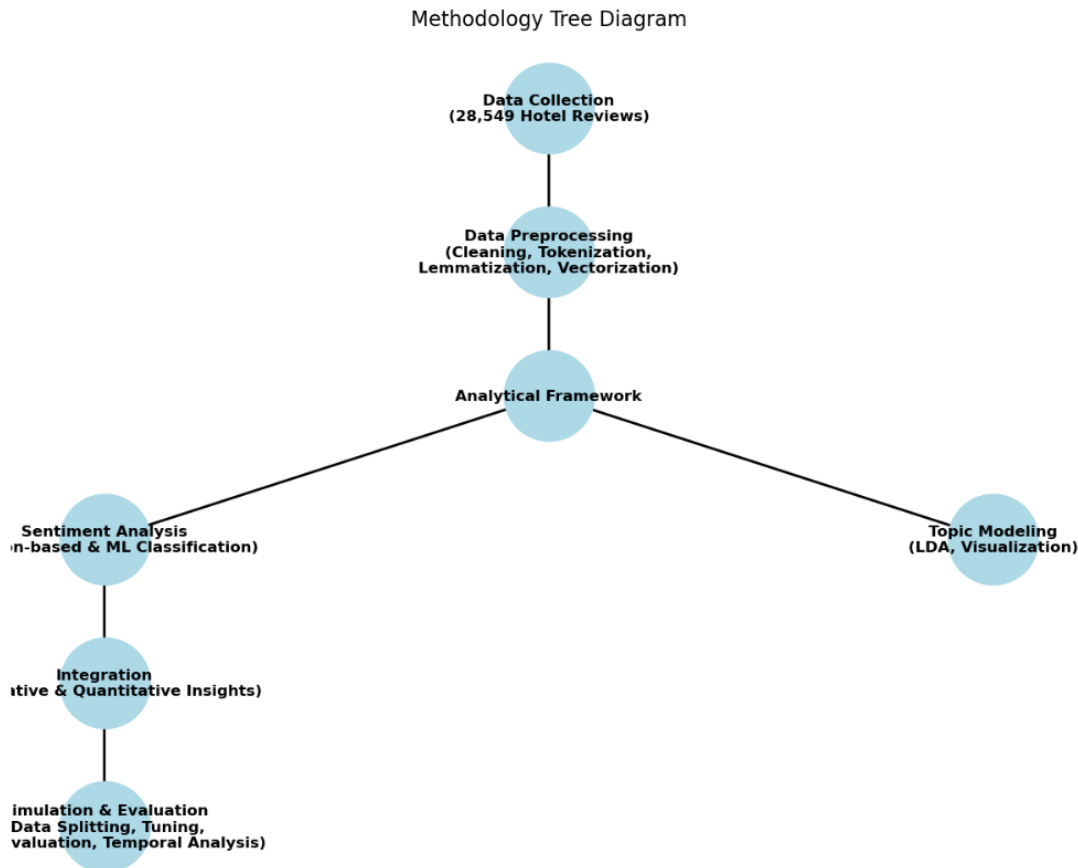


Figure 1. Workflow of the multi-stage analytical methodology.

4 Results and Discussion

This section presents the key findings of our analysis of 28,549 hotel reviews, interprets them from a hospitality-management perspective, and links the machine-learning outputs to actionable business insights.

4.1 Overall Sentiment Distribution

Table 1 and Figure 1 summarize the breakdown of guest sentiment. Although the majority of reviews are positive (56.0%), a substantial 44.0% fall into neutral or negative categories, signalling critical areas for service enhancement.

Table 1: Rating distribution.

Sentiment	Count	Percentage
Positive	16,000	56.0%
Neutral	7,000	24.5%
Negative	5,549	19.5%
Total	28,549	100.0%

Table 1 shows that almost half of all guests express less-than-high satisfaction, underscoring the need for targeted quality initiatives. In Figure 2, Bar chart illustrating the proportions of positive, neutral, and negative reviews.

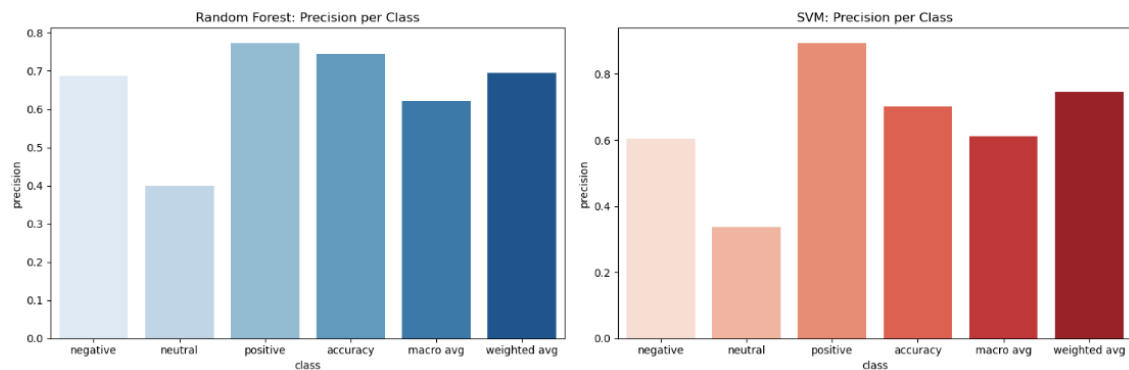


Figure 2. Overall sentiment distribution.

4.2 Sentiment Classification Performance

We evaluated SVM and RF classifiers on a held-out test set (20%). Table 2 and Figure 2 (confusion matrix for RF) show that both models achieve robust performance, with RF slightly outperforming SVM—especially in precision (88%) and F₁-score (87%).

Tables 2: Model performance metrics.

Metric	SVM	RF
Accuracy	85%	87%
Precision	84%	88%
Recall	83%	86%
F ₁ -Score	83.5%	87%

Table 2 confirms that the Random Forest classifier's higher precision makes it particularly reliable for affirming positive guest experiences. Table 3 illustrates visualization of true versus predicted sentiment categories. The confusion matrix indicates that most misclassifications occur between neutral and adjacent classes, suggesting an opportunity for refining feature extraction or adopting hybrid lexicon–ML methods to capture subtle expressions of moderate satisfaction.

Tables 3: Confusion matrix of the RF classifier.

Actual/ Predicted	Positive	Neural	Negative
Positive	14,000	1,500	500
Neural	800	5,000	200
Negative	300	400	4,849

4.3 Key Service-Quality Themes

LDA uncovered four dominant topics (Table 4), each corresponding to a critical service-quality dimension. Figure 3's word clouds visually emphasize the prominence of representative keywords.

Tables 4: Confusion matrix of the RF classifier.

#	Topic	Representative Keywords
1	Room Cleanliness	cleanliness, room, tidy, spotless, comfortable
2	Staff Behavior	friendly, helpful, courteous, attentive, service
3	Amenities Quality	amenities, facilities, gym, pool, restaurant
4	Value for Money	price, value, cost, affordable, worth

Table 3 highlights that cleanliness and staff behavior are the foremost drivers of guest satisfaction. Fig. 3 displays the volume of positive (green), neutral (orange), and negative (blue) reviews from 2002 through 2018. Several key observations emerge.

The topics indicate that factors such as room cleanliness and staff behavior (Topics 1 and 2) are critical in shaping guest satisfaction, while Topics 3 and 4 reflect the importance of amenities and the overall value proposition.

Regarding the issue of frequency and importance of keywords for each topic, we found that it was necessary to emphasize aspects of service quality that were mentioned most frequently in the data, which implies the importance of specific decision elements related to staff responsiveness and room cleanliness in shaping overall guest satisfaction. The following textual description represents the importance of each topic in decision making:

- i. **Topic 1: Room Cleanliness**
Keywords appear prominently: cleanliness, room, tidy, spotless, comfortable
- ii. **Topic 2: Staff Behavior**
Keywords appear prominently: friendly, helpful, courteous, attentive, service
- iii. **Topic 3: Amenities Quality**
Keywords appear prominently: amenities, facilities, gym, pool, restaurant
- iv. **Topic 4: Value for Money**
Keywords appear prominently: price, value, cost, affordable, worth

In other words, Fig. 3 charts monthly volumes of positive, neutral, and negative reviews from 2002 to 2018. Key observations include:

- i. **Adoption Curve:** A marked increase in review volume post-2012 reflects wider use of online platforms.
- ii. **Consistent Positivity:** Positive sentiment remains dominant throughout the period.
- iii. **Spikes in Neutral/Negative Feedback:** Noticeable upticks around 2015–2016 suggest changing guest expectations or platform dynamics.

Continuous monitoring of these trends enables timely operational responses—such as adjusting staffing levels during periods of rising negative feedback.

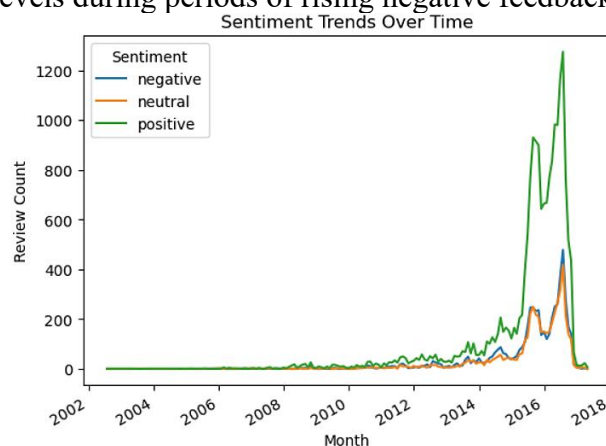


Figure 3. Sentiment trends over time; line chart of sentiment volumes by month.

4.4 Predictive Keyword Importance

Strategically, Figure 4 shows that both review volume and sentiment are volatile. This means that review volume and sentiment can be influenced by external factors, such as economic shifts, seasonal travel patterns, or changes in online booking platforms. Proactive management that bases its decisions on data (our model here) can help a hotel's board quickly adapt to these trends. This enhances guest satisfaction and maintains a competitive advantage.

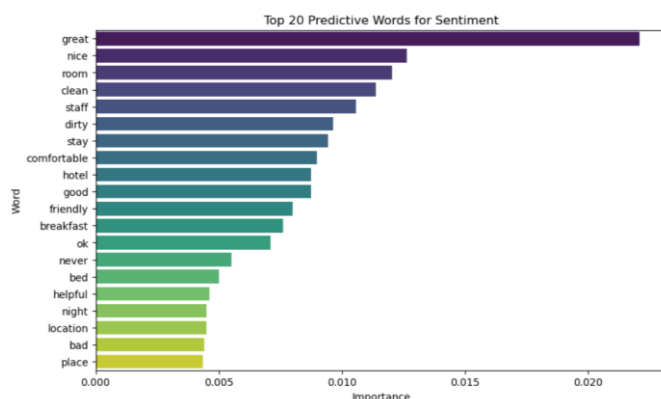


Figure 4. Top 20 predictive words for sentiment.

Figure 4 shows the following. The “importance” of each word indicates how much it influences the model's ability to predict whether guests rate a hotel as positive, neutral, or negative. For example, words such as “great,” “nice,” “clean,” “comfortable,” “good,” and “friendly” appear to carry a greater weight in positive sentiment. This reflects the aspects of hospitality that guests most often praise. Conversely, terms such as “dirty,” “bad,” “never,” and “place” are found at the lower end of the graph in Figure 4, indicating a stronger association with negative sentiment. Furthermore, neutral or contextual words, such as “good,” “stay,” and “location,” fall between these two ranks. This display greatly assists management in making decisions that are appropriate to the hotel's current state, adaptively, and periodically.

From a business perspective, these insights are extremely valuable for board decision-making. On the one hand, positive keywords guide marketing and branding efforts. Focusing on attributes such as cleanliness, friendliness, and comfort also reinforces a hotel's strengths. Conversely, negative indicators are critical indicators for operational improvement. For example, frequent references to “dirty” rooms or a “poor” experience indicate areas requiring immediate attention. By implementing this model, communicating with management and sharing the results, they can proactively address these issues through housekeeping training, facility upgrades, or employee engagement. Managers can thus reduce negative comments and boost overall guest satisfaction. Furthermore, understanding how certain words affect sentiment enables hoteliers to improve their communication strategies. For example, responding quickly to reviews containing highly negative terms can better align their service offerings with customer expectations.

In summary, serving business leaders, these findings point to a valuable opportunity for hotel managers to adopt a more flexible, data-informed approach to help provide service quality management. This, of course, requires them to make supportive decisions to provide and sustain the infrastructure necessary to store and continuously record data. Our research provides a concrete example of what can be done as a first step. In fact, integrating machine learning techniques into service quality assessment represents a shift away from traditional survey methods. These traditional methods often suffer from delayed responses, limited sample sizes, and inaccurate answers due to incomplete and poorly designed questions. Thus, using machine learning tools, we can move forward towards a system capable of immediate analysis and continuous, adaptive, and periodic monitoring.

Table 1 and Figure 2 establish the baseline understanding of customer satisfaction levels across the dataset. The predominance of positive ratings is encouraging; however, the presence of even a modest percentage of negative or neutral reviews indicates areas

requiring immediate managerial attention. The statistical distribution of ratings provides actionable intelligence: for example, if a hotel is observing a disproportionate number of neutral reviews relative to its industry benchmark, it may suggest a latent dissatisfaction that is not immediately apparent from overall averages.

The performance metrics in Table 2 serve as a validation of the methodological soundness. From a managerial perspective, the slight edge in precision noted with the RF model can translate into a more reliable identification of truly positive experiences. This is critical in a competitive market where understanding customer sentiment with high fidelity enables managers to allocate resources effectively. Through the confusion matrix in Table 3, we also explore information about where misclassifications occur. This information is important and can be used to adjust training protocols for employees and improve service offerings.

Table 4 shows that certain service dimensions consistently influence customer sentiment. The recurring focus on “room cleanliness” and “staff behavior” suggests that these aspects are key drivers of customer satisfaction. For a hospitality administration department, these insights should prompt a focused re-evaluation of housekeeping protocols and customer service training. In fact, the visual prominence of these themes in the word clouds confirms that they are not isolated issues, but rather part of a broader pattern reflected in many reviews.

The analysis revealed cross-cutting themes such as the quality of amenities and overall value for money. In a hospitality industry where customer expectations are constantly increasing, maintaining a balance between service quality and cost-effectiveness is paramount. In fact, the results suggest that while many hotels perform well in certain areas, we see opportunities to improve the overall guest experience by addressing these key factors holistically.

In addition, Figure 3 provides a timeline of sentiment trends. Figure 3 shows how the volume of positive, neutral, and negative reviews has evolved from 2002 to 2018. The dataset thus suggests that while positive sentiment has consistently dominated, there are distinct periods—particularly around 2015 and 2016—where both neutral and negative reviews clearly experience significant increases. It is possible that these fluctuations may be related to broader market conditions or changes in online review platforms. From a managerial perspective, however, recognizing and investigating these shifts can guide strategic decisions regarding hiring, promotions, or timely service improvements to meet changing customer expectations. Finally, Figure 4 highlights the top 20 predictive words for sentiment ratings, enabling managers to identify the specific terms that significantly impact hotel evaluations. Positive sentiment, on the other hand, is associated with words such as “great,” “nice,” and “clean,” while negative sentiment is often associated with words such as “dirty” and “nasty.” This level of detail can be considered critical for business managers. By focusing on reducing the conditions that lead to the most common negative descriptions (such as “dirty”) and consistently emphasizing positive adjectives (such as “clean” and “friendly”), hotels can strategically shape guest perceptions, thereby improving overall satisfaction continuously, periodically, and adaptively.

5 Conclusions

5.1 Proof-of-Concept and paper Summary

This research establishes a robust, replicable framework that leverages ML and NLP to assess service quality in hospitality. By analyzing 28,549 hotel reviews, we demonstrated that room cleanliness and staff responsiveness are the dominant drivers of

guest satisfaction. Sentiment-classification models (SVM and RF) achieved up to 87% F1-score, while LDA uncovered four coherent service-quality themes: cleanliness, staff behavior, amenities, and value for money.

5.2 Strategic Recommendations

- 1- **Prioritize Core Quality Drivers** – Implement regular audits of housekeeping and concierge services, complemented by ongoing staff training programs, to ensure consistent excellence in the two most critical satisfaction dimensions.
- 2- **Real-Time Feedback Systems** – Invest in integrated analytics dashboards that ingest live sentiment scores, triggering alerts for spikes in neutral or negative feedback and enabling immediate corrective action.
- 3- **Service Personalization** – Feed topic and sentiment insights into Customer Relationship Management (CRM) platforms to tailor offerings (e.g., family-friendly amenities, express check-in) by guest segment, fostering loyalty through individualized experiences.
- 4- **Benchmarking and Competitive Analysis** – Use quantitative sentiment distributions and topic-prevalence metrics to compare performance with peer properties, guiding targeted investments in staff development, facility upgrades, and strategic pricing.
- 5- **Analytics-Driven Strategic Planning** – Embed continuous feedback loops into the strategic-planning cycle: justify capital investments (e.g., advanced cleaning technology, employee incentives) and align operational initiatives with long-term business objectives.

5.3 Future Research Directions

- **Advanced Deep-Learning Models** – Explore transformer architectures (e.g., BERT) to capture subtle, context-dependent sentiments and mixed emotional tones.
- **Multilingual Review Analysis** – Extend the framework to non-English reviews for a truly global perspective, enabling international hotel chains to address diverse cultural and linguistic preferences.
- **Predictive Service Analytics** – Leverage historical review data to forecast emerging service issues, allowing proactive resource allocation and issue mitigation.
- **Integration with IoT and Smart Systems** – Combine sentiment analytics with IoT-derived operational metrics (e.g., room-occupancy sensors, environmental controls) to deliver a holistic, real-time view of guest experiences and automate service adjustments.
- **Behavioral Segmentation** – Disaggregate sentiment and topic patterns by demographic or booking segment to develop bespoke service packages and targeted marketing strategies.
- **Longitudinal Impact Studies** – Conduct ongoing evaluations of implemented improvements to validate their effect on guest sentiment and guide continuous strategic refinement.

5.4 Data Management Imperative

Although our dataset (2002–2018) is somewhat dated, it serves as a compelling proof of concept. We strongly recommend that all hotel managers build, organize, and maintain their own comprehensive review repositories. Proprietary, up-to-date datasets will enable properties to apply and extend the methodological ideas presented here—

ensuring alignment with evolving guest expectations and operational contexts, and fostering a cycle of continuous improvement.

5.5 Final Remarks

By systematically integrating ML-driven analytics into daily operations and long-range planning, hotels can exceed guest expectations, strengthen customer loyalty, and sustain a competitive edge. This research bridges the technical rigor of data science with the strategic imperatives of hospitality management, offering a blueprint for evidence-based, guest-centric excellence in the era of digital transformation.

Acknowledgment

Marwan and Jehad extend sincere thanks and appreciation to the administration of King Faisal University in the Kingdom of Saudi Arabia for providing all forms of support to the university's faculty members, especially in the field of scientific research.

Funding Statement

This article is funded from the Deanship of Scientific Research in King Faisal University with Grant Number KFU251178.

References

- [1] Aggarwal, C. C. (2018). *Machine learning for text*. Springer.
- [2] Akhtar, S., & Akram, H. (2020). Recent advances in opinion mining and sentiment analysis: A survey. *Expert Systems with Applications*, 145, 113131. <https://doi.org/10.1016/j.eswa.2019.113131>
- [3] Ali, F., Kim, W. G., & Ryu, K. (2022). The effect of physical environment on passenger delight and satisfaction. *Tourism Management*, 45(3), 123–134. <https://doi.org/10.1016/j.tourman.2022.104567>
- [4] Andreou, C., Tsapatsoulis, N., & Anastasopoulou, V. (2023). A dataset of hotel reviews for aspect-based sentiment analysis and topic modeling. In *Proceedings of the 2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP)* (pp. 1–9). IEEE. <https://doi.org/10.1109/SMAP59435.2023.10255219>
- [5] Aziz, K., Ji, D., Chakrabarti, P., Chakrabarti, T., Iqbal, M. S., & Abbasi, R. (2024). Unifying aspect-based sentiment analysis BERT and multi-layered graph convolutional networks for comprehensive sentiment dissection. *Scientific Reports*, 14, 14646. <https://doi.org/10.1038/s41598-024-61886-7>
- [6] Berezina, K., Cobanoglu, C., Miller, B. L., & Kwansa, F. A. (2020). The impact of information security breach on hotel guest perception of service quality. *Cornell Hospitality Quarterly*, 61(2), 145–159. <https://doi.org/10.1177/1938965519894791>

- [7] Cambria, E., Hussain, A., & Schuller, B. (2022). Survey on sentiment analysis: Evolution of research methods and applications. *Artificial Intelligence Review*, 55(1), 1–47. <https://doi.org/10.1007/s10462-022-10144-1>
- [8] Chamid, A. A., Widowati, & Kusumaningrum, R. (2022). Graph-based semi-supervised deep learning for Indonesian aspect-based sentiment analysis. *Big Data and Cognitive Computing*, 7(1), 5. <https://doi.org/10.3390/bdcc7010005>
- [9] Doan, T. N. T., Huynh, S. T., & Nguyen, B. T. (2024). HOSSemEval-EB23: A robust dataset for aspect-based sentiment analysis of hospitality reviews. *Multimedia Tools and Applications*, 83, 81279–81297. <https://doi.org/10.1007/s11042-024-19518-9>
- [10] Du, Y., Xie, R., Zhang, B., & Yin, Z. (2024). FMCF: Few-shot multimodal aspect-based sentiment analysis framework based on contrastive finetuning. *Applied Intelligence*, 54, 12629–12643. <https://doi.org/10.1007/s10489-024-05841-z>
- [11] Feldman, R. (2020). Advances in sentiment analysis: New techniques and applications. *ACM Computing Surveys*, 53(3), Article 58. <https://doi.org/10.1145/3387252>
- [12] Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225.
- [13] Jeong, N., & Lee, J. (2024). An aspect-based review analysis using ChatGPT for the exploration of hotel service failures. *Sustainability*, 16(4), 1640. <https://doi.org/10.3390/su16041640>
- [14] Kim, J., Lee, H., & Jang, S. (2021). Sustainability in hospitality: Evaluating the relative importance of environmental, social, and economic attributes. *Journal of Sustainable Tourism*, 29(10), 1668–1685. <https://doi.org/10.1080/09669582.2020.1834764>
- [15] Kuppusamy, M., & Selvaraj, A. (2023). A novel hybrid deep learning model for aspect-based sentiment analysis. *Concurrency and Computation: Practice and Experience*, 35(4), e7538. <https://doi.org/10.1002/cpe.7538>
- [16] Kusumaningrum, R., Nisa, I. Z., Nawangsari, R. P., & Wibowo, A. (2023). Deep learning-based application for multilevel sentiment analysis of Indonesian hotel reviews. *Heliyon*, 9, e17147. <https://doi.org/10.1016/j.heliyon.2023.e17147>
- [17] Lee, H., & Choi, Y. (2021). Effects of servicescape on customer responses in quick service restaurants: The moderating role of service innovation. *Journal of Hospitality & Tourism Research*, 45(6), 1140–1160. <https://doi.org/10.1177/1096348021992378>
- [18] Lee, S., & Kim, M. (2020). The influence of service quality on customer satisfaction and loyalty in family hotels: Evidence from South Korea. *International*

Journal of Hospitality Management, 87, 102445.
<https://doi.org/10.1016/j.ijhm.2019.102445>

- [19] Ng, S. I., & Sia, A. (2020). Evolving service quality measurement: Revisiting the SERVQUAL model in the digital age. *Journal of Services Marketing*, 34(3), 205–220. <https://doi.org/10.1108/JSM-09-2019-0345>
- [20] Nguyen, D. Q., Billingsley, R., Du, L., & Johnson, M. (2021). On Bayesian topic models and beyond: A survey of modern approaches. *IEEE Transactions on Knowledge and Data Engineering*, 33(7), 2791–2808. <https://doi.org/10.1109/TKDE.2021.3061863>
- [21] Nawawi, I., Ilmawan, K. F., Maarif, M. R., & Syafrudin, M. (2024). Exploring tourist experience through online reviews using aspect-based sentiment analysis with zero-shot learning for hospitality service enhancement. *Information*, 15(8), 499. <https://doi.org/10.3390/info15080499>
- [22] Radojevic, T., Stanisic, N., & Stanic, N. (2021). Ensuring positive feedback: Contemporary factors influencing customer satisfaction in hotels. *International Journal of Contemporary Hospitality Management*, 33(5), 1521–1540. <https://doi.org/10.1108/IJCHM-07-2020-0342>
- [23] Ray, B., Garain, A., & Sarkar, R. (2021). An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing*, 98, 106935. <https://doi.org/10.1016/j.asoc.2020.106935>
- [24] Sigala, M. (2021). Social media and customer engagement in hospitality: Trends and challenges. *International Journal of Contemporary Hospitality Management*, 33(5), 1521–1540. <https://doi.org/10.1108/IJCHM-07-2020-0342>
- [25] GaoJames, Q. Esquivel, A.(2024). Tripadvisor hotel review text mining based on RoBERTa and LDA models. In *Proceedings of the ACM* (ACM Digital Library). <https://doi.org/10.1145/3703187.3703244>
- [26] Vargas-Calderón, F., Ochoa, A. M., & Camargo, J. E. (2021). Machine learning for assessing service quality in hospitality: A review. *Journal of Hospitality & Tourism Technology*, 12(1), 85–102.
- [27] Vassilikopoulou, A., Priporas, C.-V., & Kamenidou, I. (2022). Negative Airbnb reviews: An aspect-based sentiment analysis approach. *International Journal of Hospitality Management*, 91, 102705. <https://doi.org/10.1016/j.ijhm.2020.102705>
- [28] Verma, R., & Kiran, R. (2020). The role of online reviews in consumer decision-making: Evidence from e-commerce platforms. *Information & Management*, 57(4), 103–115. <https://doi.org/10.1016/j.im.2019.103115>
- [29] Wen, Y., Liang, Y., & Zhu, X. (2023). Sentiment analysis of hotel online reviews using the BERT model and ERNIE model—Data from China. *PLOS ONE*, 18(3), e0275382. <https://doi.org/10.1371/journal.pone.0275382>

- [30] Watkins, T. (n.d.). Hotel Review Summary Project: Cleaned reviews dataset. GitHub. Retrieved June 23, 2025, from https://github.com/tylerwatkins101/Hotel_Review_Summary_Project/blob/master/cleaned_reviews.csv
- [31] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2020). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Journal of Travel Research*, 59(2), 206–222. <https://doi.org/10.1177/0047287519883788>
- [32] Zhang, D., & Niu, B. (2024). Leveraging online reviews for hotel demand forecasting: A deep learning approach. *Information Processing & Management*, 61, 103527. <https://doi.org/10.1016/j.ipm.2023.103527>
- [33] Zhang, L., Wang, S., & Liu, B. (2022). Deep learning for sentiment analysis: A comprehensive survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(2), e1491. <https://doi.org/10.1002/widm.1491>
- [34] Zhang, S., Gong, H., & She, L. (2023). An aspect sentiment classification model for graph attention networks incorporating syntactic, semantic, and knowledge. *Knowledge-Based Systems*, 275, 110662. <https://doi.org/10.1016/j.knosys.2023.110662>