International Journal of Computational Economics and Econometrics

ISSN online: 1757-1189 - ISSN print: 1757-1170
https://www.inderscience.com/ijcee

Unpacking customer feedback and brand equity dynamics in the hospitality industry through machine learning techniques

T.D. Dang, M.T. Nguyen

DOI: 10.1504/IJCEE.2025.10069873

Article History:

Received:	20 September 2023
Last revised:	23 September 2023
Accepted:	20 January 2024
Published online:	17 March 2025

Unpacking customer feedback and brand equity dynamics in the hospitality industry through machine learning techniques

T.D. Dang*

Department of MIS,
School of Industrial Management,
Ho Chi Minh City University of Technology (HCMUT),
VNU-HCM,
Ho Chi Minh City, Vietnam
Email: dtdoan.sdh221@hcmut.edu.vn
and
Eastern International University,
Thu Dau Mot, Binh Duong Province, Vietnam
Email: doan.dang@eiu.edu.vn
*Corresponding author

M.T. Nguyen

Department of MIS,
School of Industrial Management,
Ho Chi Minh City University of Technology (HCMUT),
VNU-HCM,
Ho Chi Minh City, Vietnam
Email: n.m.tuan@hcmut.edu.vn

Abstract: This study utilises Latent Dirichlet allocation (LDA) and latent semantic analysis (LSA) for advanced topic modelling in the hospitality sector, analysing customer feedback from Booking.com in Ho Chi Minh City, Vietnam. It highlights crucial aspects influencing brand equity: ambient noise levels, room standards, facility provisions, staff interactions, and strategic location advantages. Further, the research integrates an extensive suite of machine learning (ML) and deep learning (DL) techniques, including logistic regression (LR), random forest (RF), multinomial Naive Bayes (NB), long short-term memory (LSTM), convolutional neural network (CNN), and notably, the dense model. The dense model stands out, demonstrating remarkable performance with an accuracy rate of 0.95 and an F1-score of 0.97, validating the effectiveness of data-driven methodologies in extracting nuanced customer sentiments. These insights offer a multifaceted understanding, serving as a valuable resource for practitioners to refine service strategies, elevate customer satisfaction, and strengthen market presence.

Keywords: customer feedback; brand equity; sentiment analysis; topic modelling; hospitality industry; machine learning.

Reference to this paper should be made as follows: Dang, T.D. and Nguyen, M.T. (2025) 'Unpacking customer feedback and brand equity dynamics in the hospitality industry through machine learning techniques', *Int. J. Computational Economics and Econometrics*, Vol. 15, Nos. 1/2, pp.78–93.

Biographical notes: T.D. Dang is a PhD candidate in Business Administration at Ho Chi Minh City University of Technology (HCMUT), VNU-HCM, specialising in digital services and customer behaviour. He also serves as a Lecturer at Eastern International University, teaching business analytics-related courses.

M.T. Nguyen is a Senior Lecturer in Service Management, Knowledge Management, Data Science and Management Information Systems at the School of Industrial Management, HCMC University of Technology, Vietnam. He received his PhD in Business Administration from the HCMC University of Technology. He has published articles focusing on service management and technology, information/knowledge management and management information systems in journals like *The Service Industries Journal*, *International Journal of Information Technologies and Systems Approaches*, *International Journal of Knowledge and Systems Science*, *Management Research*, *International Journal of Services Technology and Management*, and *International Journal of Asian Business and Information Management*.

This paper is a revised and expanded version of a paper entitled 'Understanding customer perception and brand equity in the hospitality sector: integrating sentiment analysis and topic modeling' presented at International Conference on Applied Economics (ICOAE), Brno, Czech Republic, 29 June–1 July 2023.

1 Introduction

The hospitality industry plays a crucial role in worldwide economies, fostering job opportunities, economic expansion, and enriching tourist experiences (Raguseo et al., 2017; Sheehan et al., 2018). In today's interconnected global landscape, grasping customer perceptions and brand equity is crucial for the hospitality sector, influencing reputation, loyalty, and profitability in an ever-evolving market (Górska-Warsewicz and Kulykovets, 2020; Sürücü et al., 2019). Understanding customer sentiments and preferences has become paramount for organisations to effectively manage their brand reputation and enhance customer satisfaction (Serra-Cantallops et al., 2020). With the emergence of social media platforms and online review platforms like Booking.com, Agoda.com, and Airbnb.com, In the modern era, consumers have the capability to share their views, feedback, and experiences regarding products or services either by directly contacting the respective companies via email or by posting reviews on online platforms (Chaw and Tang, 2019; Raguseo et al., 2017). User-generated content (UGC) has become an abundant source of data that holds significant value for businesses in understanding customer perceptions. Sentiment analysis and topic modelling have emerged as powerful techniques to unlock the insights hidden within this vast textual data (Blasco-Arcas et al., 2022; Chaw and Tang, 2019).

The application of sentiment analysis through ML and DL models on UGC has emerged as a prominent trend in academic research. This approach offers a nuanced understanding of customer sentiment, drawing from the vast and rich reservoirs of organic feedback. For the hospitality industry, deciphering such sentiment holds paramount significance, as it provides actionable insights that can drive service improvement and foster enhanced guest experiences. Leveraging ML and DL for sentiment analysis on UGC, thus, promises to be a transformative avenue for research, offering substantial value to stakeholders in the hospitality sector (Liu and Park, 2015).

Topic models serve as instrumental tools in uncovering the intricate intricacies of social phenomena. They bridge the gap between social science and (un)structured analysis, integrating diverse reasoning methodologies and big data analytics (Hannigan et al., 2019). In the realm of text mining, myriad methods exist, each possessing unique techniques for discerning latent topics within textual data. In this study, we amalgamate insights from both Latent Dirichlet allocation (LDA) a probabilistic topic modelling approach and latent semantic analysis (LSA) an advanced natural language processing method. By applying topics modelling techniques to customer reviews, businesses can gain a qualitative understanding of customer interests, concerns, and preferences (Albalawi et al., 2020; Blasco-Arcas et al., 2022).

This study aims to contribute to the existing literature by integrating sentiment analysis and topic modelling techniques to comprehensively understand customer perception and brand equity in the hospitality sector. Leveraging a dataset of 31,989 customer reviews from Booking.com in Ho Chi Minh City, Vietnam, this research employs a range of ML and DL models, including RF, LR, NB, long short-term memory (LSTM), convolutional neural network (CNN), and dense (Blasco-Arcas et al., 2022; Essien and Chukwukelu, 2022). Through the combination of sentiment analysis and topic modelling using LDA (Blei et al., 2003), the study aims to identify underlying patterns, evaluate their impact on brand equity, and provide valuable insights for hospitality managers to enhance service quality, improve customer satisfaction, and strengthen their brand's competitive position.

This research contributes to academic theory and practical managerial decision-making in the hospitality sector by harnessing the power of sentiment analysis and topic modelling. The findings from this study will inform the development of effective marketing strategies and assist hospitality managers in making data-driven decisions to meet and exceed customer expectations. Furthermore, this research sets the stage for future investigations in the domain of customer perception and brand equity, emphasising the need to explore novel approaches to UGC analysis and address the challenges faced by the industry in an ever-evolving digital landscape.

2 Literature review

2.1 Customer perception and brand equity

Branding plays a vital role in services and customer perception plays a crucial role in shaping brand equity in the hospitality sector (Iglesias et al., 2019; Sürücü et al., 2019). Positive customer experiences and favourable perceptions contribute to developing a solid brand reputation, increased customer loyalty, and higher customer satisfaction (Liat et al., 2014). Understanding customer perceptions enables businesses to tailor their offerings, improve service quality, and create targeted marketing strategies (Blasco-Arcas et al., 2022; Sheehan et al., 2018).

On the other hand, brand equity encompasses the intangible assets associated with a brand, including customer perceptions, brand awareness, brand loyalty, and brand associations (Brahmbhatt and Shah, 2017). Customer perceptions directly impact brand equity, as positive perceptions enhance brand value and competitive advantage (Martínez and Nishiyama, 2019). Studies have shown that a positive online reputation from customer reviews and sentiments strongly correlates with higher brand equity (Brahmbhatt and Shah, 2017; Iglesias et al., 2019; Martínez and Nishiyama, 2019; Sürücü et al., 2019).

2.2 User-generated content data

The emergence of online platforms and social media has revolutionised how customers share their experiences and opinions (Lamberton and Stephen, 2016). UGC has become a valuable source of information for businesses in various industries, including the hospitality sector (Ye et al., 2011). UGC encompasses customer reviews, ratings, and feedback on TripAdvisor, Yelp, and Booking.com (Mariani and Borghi, 2020). These platforms provide a wealth of data that reflects customer experiences, preferences, and sentiments. Researchers have recognised the significance of UGC data in understanding customer perceptions, identifying trends, and informing decision-making processes (Chaw and Tang, 2019; Essien and Chukwukelu, 2022; Mariani and Borghi, 2020; Ye et al., 2011).

2.3 Sentiment analysis and topic modelling

Sentiment analysis, or opinion mining, utilises ML and DL algorithms to extract and quantify subjective information from text data. ML and DL models like LR, RF, NB, LSTM, Dense and CNN have been widely employed for sentiment analysis in the hospitality sector (Abkenar et al., 2021; Debortoli et al., 2016; Ye et al., 2011; Zarezadeh et al., 2022). These algorithms classify sentiments into positive, negative, or neutral categories, providing valuable insights into customer attitudes toward different attributes. Deep learning (DL) algorithms like LSTM and CNN capture sequential and local patterns in text, while LR, RF, and NB leverage statistical relationships to classify sentiments accurately (Ullah et al., 2020).

In the domain of text analytics, topic modelling algorithms play a pivotal role by statistically deciphering themes embedded in unstructured texts. Among such algorithms, LDA stands out as a robust method to automatically categorise textual collections into overarching themes 4. LDA's foundational hypothesis suggests that a text document encapsulates multiple themes. Designed as a three-tiered hierarchical Bayesian model, LDA posits each text item as a mixture of various topics, where each topic is conceived as a mixture of topic probabilities. Significantly, these topic probabilities offer an explicit document representation. LDA not only furnishes an automatic summary of topics via discrete word probability distributions for each topic but also projects per-document topic distributions. The utility of LDA, as evident in its application to accident reports, is its proficiency in automatically clustering words into salient topics, thereby potentially highlighting the most recurrent types of incidents documented.

The LSA has emerged over the past decades as a fundamental computational model for extracting semantic content from textual data. Central to LSA is the premise that words of similar semantic meaning frequently appear in analogous contexts within text. By analysing patterns of word co-occurrence across extensive textual samples, LSA produces semantic representations of words. The approach identifies which words are likely to appear in congruent documents, thereby mapping them closer in a semantic space 3. After training on a voluminous document collection, LSA yields a semantic space populated by vectors – termed semantic vectors – which encapsulate semantic features of each word. Intuitively, the richness of a training corpus improves the model's capability to discern or equate words semantically 3.

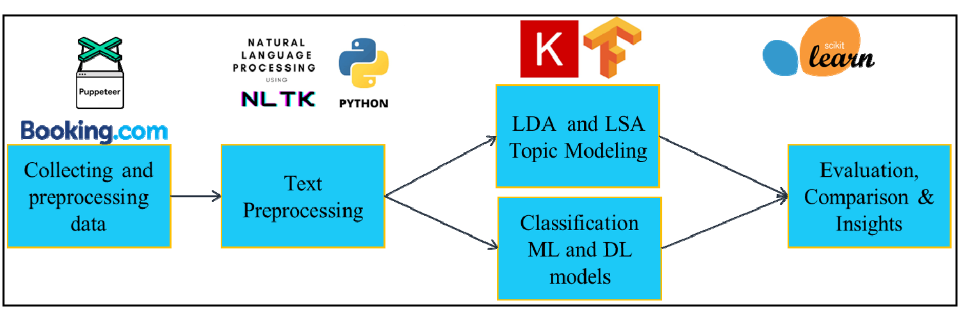
Integrating sentiment analysis and topic modelling enables researchers and businesses to gain comprehensive insights into customer perceptions and brand equity (Brahmbhatt and Shah, 2017; Martínez and Nishiyama, 2019; Sürücü et al., 2019). By applying sentiment analysis, businesses can understand sentiment distribution towards different attributes, while topic modelling allows for identifying key topics prevalent in customer reviews (Abkenar et al., 2021; Martínez and Nishiyama, 2019; Ullah et al., 2020; Zarezadeh et al., 2022). These insights help businesses identify areas for improvement, align their services with customer preferences, and make data-driven decisions to enhance customer satisfaction and strengthen their brand's competitive position in the hospitality sector. These models facilitate a deeper understanding of customer sentiments and preferences, aiding businesses in improving service quality, enhancing customer satisfaction, and effectively managing brand equity (Mustak et al., 2021).

3 Methodology

3.1 Framework for analysing UGC

This research uses a framework for analysing UGC from Booking.com to understand customer preferences and enhance experiences in the tourism sector (Abkenar et al., 2021; Zarezadeh et al., 2022). The process comprises three stages: data collection and organisation, data preparation, sentiment classification, and topic modelling (Essien and Chukwukelu, 2022). For details, see Figure 1.

Figure 1 Summary flow uses ML and DL for analytics (see online version for colours)



Data from Booking.com concerning hotels in Ho Chi Minh City, Vietnam, was collected in February 2023 using Node.js and Puppeteer (Chang, 2022). The dataset included over

156,256 records and was pre-processed using Python libraries, retaining only English comments for uniform analysis. The data was cleaned, encoded, and explored to identify patterns and insights, laying the foundation for further steps; final dataset comprised 31,989 records. Finally, ML and DL models were evaluated and tuned for sentiment classification and topic modelling. The model's performance was assessed by testing data with metrics like accuracy, precision, and recall. LDA was used for topic modelling.

3.2 ML and DL models for sentiment analysis

The study utilised six different algorithms to classify sentiments expressed in comments, comprising three traditional ML algorithms (RF, LR, NB) (Salmony and Faridi, 2021; Ullah et al., 2020) and three DL models (dense, LSTM, CNN) (Abid et al., 2020). The inclusion of traditional ML algorithms allowed leveraging established methods widely used in sentiment analysis; DL models, on the other hand, offer the ability to capture complex patterns and dependencies in data (Mehraliyev et al., 2022). By incorporating DL models alongside traditional ML algorithms, the study sought to evaluate their respective effectiveness in sentiment analysis. This comprehensive approach provides insights into the strengths and limitations of both ML and DL techniques for sentiment classification tasks (Albalawi et al., 2020; Mustak et al., 2021).

3.3 Topic modelling techniques

Furthermore, the research used LDA and LSA for topic modelling, helping identify underlying themes in hotel reviews by analysing word frequencies.

The LDA and LSA are advanced topic modelling techniques employed to unveil hidden thematic structures within extensive text datasets. LDA is grounded on the notion that documents are mixtures of topics, which in turn are mixtures of words. Through its probabilistic approach, LDA assigns words in a document to certain topics based on the probability and distribution of words among various topics. Conversely, LSA uses singular value decomposition (SVD) on the term-document matrix to discern topics by identifying patterns of word co-occurrence.

The study choice of LDA and LSA is underpinned by their proven efficacy in dissecting and interpreting large volumes of unstructured text data. LDA offers a robust framework for identifying the latent thematic structure within the customer reviews, while LSA contributes by uncovering the underlying semantic patterns, enabling a more nuanced understanding of customer sentiments. This combination is particularly potent in the context of the hospitality industry, where customer feedback is rich in detail and complexity. The comparative advantage of these techniques over other alternatives lies in their ability to capture the subtle nuances in customer feedback, which is paramount for deriving actionable business insights.

In a Python-based implementation, the initial step involves pre-processing the text data. Libraries such as sklearn provide tools for this. The CountVectorizer or TfidfVectorizer from sklearn feature_extraction.text can be used to tokenise the text and convert it into a document-term matrix. After this transformation, topic modelling can commence. For LDA, one can utilise LatentDirichletAllocation from sklearn.decomposition. For LSA, the appropriate method would be TruncatedSVD from the same module.

In essence, the following Python snippets demonstrate this:

```
# Python implementation
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.decomposition import LatentDirichletAllocation, TruncatedSVD
# Text preprocessing
vectorizer = TfidfVectorizer(stop_words = 'english')
data_vectorized = vectorizer.fit_transform(data)
# LDA
lda_model = LatentDirichletAllocation(n_components = num_topics)
lda_output = lda_model.fit_transform(data_vectorized)
# LSA
lsa_model = TruncatedSVD(n_components = num_topics)
lsa_output = lsa_model.fit_transform(data_vectorized)
```

This simplified workflow elucidates how Python, equipped with libraries like sklearn, can efficiently facilitate topic modelling through LDA and LSA.

3.4 Model evaluation

Model performance was assessed using accuracy, precision, recall, and F1-score metrics (see Table 1). The research used these metrics to evaluate six ML and DL algorithms on the hotel reviews dataset, providing insights into their predictive power and areas for improvement (Hossin and Sulaiman, 2015; Ullah et al., 2020).

Table 1 Classification performance metrics

Performance metric	*Description*	*Formula*
Confusion matrix	It is a 2 × 2 table that includes true positive (TP), false positive (FP), false negative (FN) and true negative (TN) metrics.	
Accuracy	It measures the proportion of correct predictions made by the model.	Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision	It measures the fraction of positive predictions.	Precision = TP / (TP + FP)
Recall	It measures the fraction of actual positive instances correctly predicted by the model.	Recall = TP / (TP + FN)
F1-score	The harmonic means of precision and recall.	F1-score = 2 * (precision * recall) / (precision + recall)

4 Experiment results

4.1 Data collection and pre-processing

The crawler gathered hotel names, star ratings, locations, addresses, and UGC-like reviews and ratings (see Figure 2). After collecting, only English reviews with complete attribute values were retained, then encoded for machine readability (Ullah et al., 2020).

Before analysing the dataset, crucial pre-processing techniques were implemented on the textual data to guarantee precise and significant outcomes (Kirilenko et al., 2018; Puh and Bagić Babac, 2022). These techniques encompassed converting all text to lowercase to ensure consistency and prevent duplication based on case sensitivity. Tokenisation was performed to split the text into individual words or tokens, while stopwords, commonly uninformative words, were removed. Additionally, lemmatisation was applied to reduce words to their base or dictionary form (Bird, 2006). These pre-processing steps were vital in preparing the text data for subsequent sentiment and topic modelling analysis by reducing noise, improving interpretability, and optimising algorithmic efficiency (Debortoli et al., 2016). The 'sentiment' attribute was created for sentiment analysis, classifying reviews into three categories representing positive, negative, and neutral customer experiences (Puh and Bagić Babac, 2022). The top popular words extracted from comments show in Figure 3.

Figure 2 Sample of review (see online version for colours)

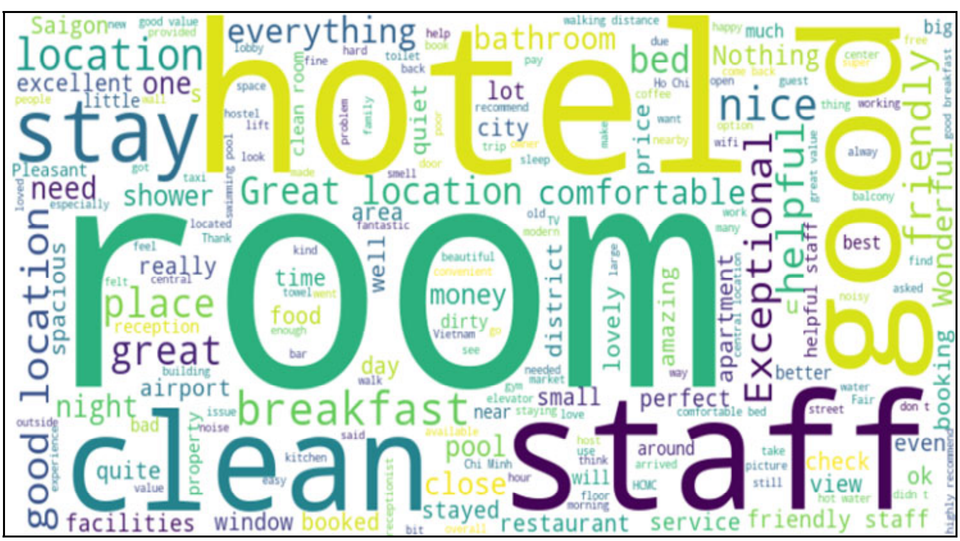

S Société
United States of America
Suite Double
5 nights · August 2022
Solo traveler

Reviewed: August 1, 2022
6.0
Decent but not a great value

· Good location and very friendly staff. Room was very spacious and reasonably clean.

· Upon checkin, my key card didn't work. Took three attempts (and going back and forth with reception) before they realized the battery in the lock was dead. A couple of days later, my key card stopped working again and they had to send an engineer to fix it. Guests should not have to waste time dealing with this kind of inconvenience.
Soundproofing of the room wasn't great.
Laundry service indicated that laundry would be returned by 7 pm same day if given to staff before 9 am, yet I had to call after 8 pm because I had still not received my clothes.
Overall, I don't think the guest experience delivered is consistent with the price I paid.

Figure 3 The word cloud (see online version for colours)



4.2 Sentiment analysis

The study aimed to classify sentiments expressed in comments using six different algorithms, three of which were traditional ML algorithms RF, LR, and NB (Salmony and Faridi, 2021; Ullah et al., 2020), and the other three were DL models: dense, LSTM, and CNN (Abid et al., 2020; Ullah et al., 2020). The objective was to develop models that could automatically analyse the textual content of comments and accurately predict the corresponding sentiment category. The performance metrics of the classification models are summarised in Table 2. The table presents accuracy, precision, recall, and F1-score for each sentiment category (negative, neutral, and positive) across the different models.

Multinomial Naive Bayes (NB) achieved an accuracy of 0.88 with high precision for the negative sentiment category while having low precision and recall for the neutral and positive categories. Random forest (RF) achieved an accuracy of 0.90 with balanced precision, recall, and F1-scores for all sentiment categories. Logistic regression (LR) achieved an accuracy of 0.94 with relatively high precision, recall, and F1-scores across all sentiment categories. The DL models also performed well. LSTM achieved an accuracy of 0.93, CNN achieved an accuracy of 0.94, and dense achieved the highest accuracy of 0.95. These models showed competitive precision, recall, and F1-scores for all sentiment categories, indicating their effectiveness in sentiment classification tasks.

Table 2 Summary classification performance metrics

Model	*Accuracy*	*Precision*			*Recall*			*F1-score*		
		Neg	*Neu*	*Pos*	*Neg*	*Neu*	*Pos*	*Neg*	*Neu*	*Pos*
NB	0.88	0.97	0	0.88	0.05	0.00	1	0.10	0.00	0.94
RF	0.91	0.96	0.94	0.90	0.24	0.20	1	0.38	0.32	0.95
LR	0.94	0.85	0.88	0.95	0.65	0.34	0.99	0.74	0.50	0.97
LSTM	0.93	0.65	0	0.96	0.77	0	0.98	0.70	0	0.97
CNN	0.94	0.77	0.83	0.96	0.76	0.32	0.99	0.77	0.46	0.97
Dense	0.95	0.80	0.97	0.96	0.74	0.39	0.99	0.77	0.55	0.97

Notes: Random forest (RF), logistic regression (LR), multinomial Naive Bayes (NB), long short-term memory (LSTM), and convolutional neural network (CNN).

4.3 Topic modelling

For a comprehensive understanding of customer perceptions derived from reviews, the LDA and LSA models were employed.

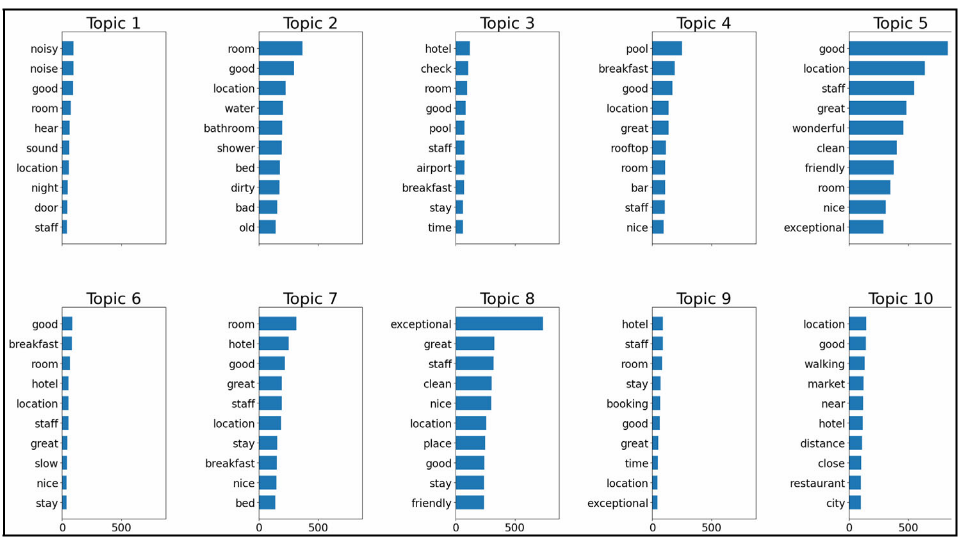
4.3.1 Thematic analysis of reviews using the LDA model

From the topics identified via the LDA model, several central themes emerge which provide valuable insights into the customer experience. Here's a consolidated thematic breakdown:

- *Theme 1 – room quality and comfort:* Topics like Topic #1 (concerning noise) and Topic #2 (highlighting cleanliness and amenities) provide insights into guests' perceptions of room quality. Factors like sound insulation, room cleanliness, and the state of bathroom amenities play crucial roles in guest comfort.

- *Theme 2 – hotel amenities and leisure facilities:* Topic #3 and Topic #4 underscore the significance of hotel facilities, ranging from general amenities such as pools and breakfast to leisure offerings like rooftop bars. The guest experience is often shaped by these added comforts and conveniences.
- *Theme 3 – overall positive reception and staff efficiency:* Topic #5, Topic #7, Topic #8, and Topic #9 are overwhelmingly positive, touching on the friendly and efficient staff, cleanliness, and overall exceptional hotel services. Staff interactions seem to significantly impact the overall guest experience.
- *Theme 4 – service speed and efficiency:* While Topic #6 overall presents a positive sentiment, the mention of 'slow' suggests that service speed or potential issues like slow Wi-Fi could be areas that need attention.
- *Theme 5 – location and proximity to attractions:* Topic #10 encapsulates the value of the hotel's location, emphasising its proximity to significant attractions, markets, and other city conveniences.

Figure 4 Topics extracted from LDA (see online version for colours)



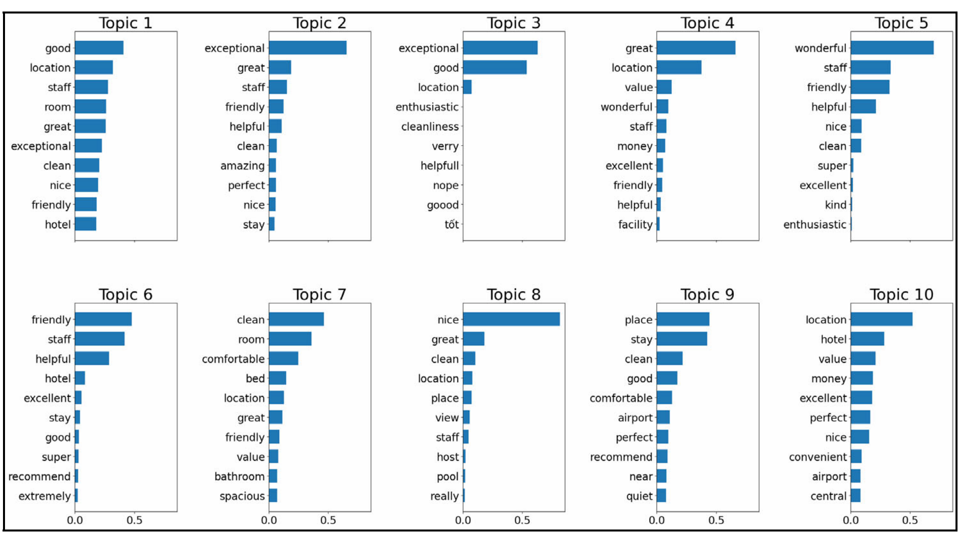
To summarise, the consolidated themes provide a snapshot of the primary drivers of customer satisfaction and areas that might require attention (see Figure 4 and 5 for details). The overall sentiment leans positive, especially concerning staff efficiency and hotel services. However, addressing specific concerns, particularly around room quality and service speed, could further elevate the guest experience.

4.3.2 Thematic analysis of reviews using the LSA model

Drawing from the topics discerned via the LSA model, we can group the insights into central themes that provide a deeper understanding of the customer experience. Here's a thematic breakdown based on the provided LSA topics:

- *Theme 1 – quality of guest experience:* Topics like Topic #1, Topic #5, and Topic #8 emphasise elements like ‘good’, ‘exceptional’, ‘wonderful’, and ‘great’, which are strong indicators of a generally positive guest experience. These topics collectively touch on the ambiance, service quality, and the overall impression guests have of their stay.
- *Theme 2 – room conditions and comfort:* Topic #2 and Topic #7 provide insights into the condition and comfort of rooms. With mentions of ‘bed’, ‘bathroom’, ‘dirty’, and ‘old’, it is evident that guests are keenly observant of room cleanliness, the state of amenities, and the overall comfort level.
- *Theme 3 – hotel facilities and amenities:* Topic #3, Topic #4, and Topic #6 collectively highlight the broader facilities the hotel offers. These range from general amenities like pools and breakfasts to other offerings such as rooftop bars. The recurrence of ‘breakfast’ and ‘pool’ across topics underlines their importance in the guest experience.
- *Theme 4 – staff efficiency and friendliness:* Topic #2, Topic #5, Topic #6, Topic #7, Topic #8, and Topic #9 all touch on the staff’s role in shaping the guest experience. Terms like ‘friendly’, ‘helpful’, ‘staff’, and ‘exceptional’ emphasise the significant impact that staff interactions and service quality have on guest satisfaction.
- *Theme 5 – location and accessibility:* Topic #10 captures the essence of the hotel’s location and its significance. Phrases like ‘walking distance’, ‘near’, ‘market’, and ‘close’ showcase the value guests place on the hotel’s proximity to city attractions, conveniences, and overall accessibility.

Figure 5 Topics extracted from LSA (see online version for colours)



In summation, the themes derived from the LSA model paint a comprehensive picture of the customer experience. While the hotel appears to be performing well in areas like staff interaction and general guest experience, there’s a consistent emphasis on room conditions and hotel facilities, signalling potential areas for enhancement.

5 Discussion and implication

5.1 Discussion

A multitude of studies have recognised the pivotal role of data analytics in enhancing business processes. This research distinguishes itself by demonstrating how sentiment analysis and topic modelling critically contribute to deciphering customer perceptions and bolstering brand equity within the hospitality domain. The blend of methodological innovation and empirical evidence in this study underscores the significant capabilities of machine learning (ML) and DL techniques in the analysis of customer feedback. The research not only highlights the inherent potential of these techniques but also exemplifies their practical application in guiding data-informed decision-making and strategic development in the hospitality industry. Furthermore, the quantitative analysis in this study provides a systematic approach for converting customer insights into substantial improvements in operational efficiency and financial outcomes, setting a new standard for data-driven research in this field.

Table 2 presents the performance metrics of various ML and DL models in sentiment analysis. Intriguingly, while traditional ML models like RF and LR showcased impressive results, the DL models, specifically dense, LSTM, and CNN, exhibit marginally superior performance. With accuracy rates approaching 0.95, these DL models emphasise their enhanced ability to decipher intricate patterns in UGC, often outperforming their traditional ML counterparts. Incorporating these advanced DL models into the hospitality sector could be a game-changer. By integrating such state-of-the-art models, an automated system can offer real-time insights into emerging customer sentiments, giving hospitality managers a competitive edge. By harnessing the predictive prowess of these models, managers can proactively identify potential issues, ensuring timely interventions and enhancing customer satisfaction – a testament to the increasing importance of DL in the domain of sentiment analysis.

In the journey to decipher the complex terrain of customer perception and brand equity in the hospitality industry, the collaborative insights from LDA and LSA topic modelling emerge as invaluable. Central themes like the quality of guest experiences (quality of guest experience) serve as robust pillars, with both models illuminating their weight in boosting brand equity. Delving into specifics, room conditions and cleanliness (room conditions and comfort) stand out prominently. These consistent findings underscore the urgency for establishments to be unwavering in their commitment to excellence in these domains. Any decline can have repercussions, creating a ripple effect in eroding brand perception.

Diving deeper, the spotlight on hotel facilities and the suite of amenities they offer (hotel facilities and amenities) is illuminating. In the fiercely competitive arena of hospitality, these amenities, when unique or superior, can be game changers, enhancing the allure of a hotel and solidifying its brand stature. However, one element resonates with unmatched intensity across both modelling techniques: the undeniable importance of staff interactions (staff efficiency and friendliness). The models resonate in harmony about the instrumental role of staff in shaping guest experiences. This insight is a clarion call for establishments to relentlessly focus on staff development, nurturing an ethos that is relentlessly guest oriented.

Lastly, a theme that unfailingly emerges is the strategic advantage of location (location and accessibility). Both LDA and LSA underline its pivotal role, indicating that the geographical positioning of a hotel, especially its closeness to attractions and essential conveniences, is a significant influencer of guest choice and satisfaction.

5.2 Theoretical implications

The theoretical insights of this study align with established frameworks such as the SERVQUAL model and the brand equity pyramid, underscoring the pivotal roles of service quality, customer experience, and brand perception in fostering customer satisfaction and brand equity (Górska-Warsewicz and Kulykovets, 2020; Macieira et al., 2020). These foundational models offer a systematic lens through which the nuanced contributions of the current study are magnified. While acknowledging the well-documented benefits of data analysis in business, this research extends beyond the theoretical norm by presenting a quantifiable connection between customer feedback analysis and tangible business outcomes (Bisoi et al., 2020). By employing robust data analysis techniques, the study demonstrates the direct impact of customer-centric strategies on financial performance indicators, offering a concrete assessment of return on investment in customer satisfaction initiatives (Kolomoyets and Dickinger, 2023). This approach not only reaffirms the significance of customer-focused approaches within the hospitality industry but also provides empirical evidence for the financial viability and strategic value of such initiatives. The integration of detailed customer feedback analysis and operational metrics sets this research apart, marking a novel contribution to the literature and offering a valuable, data-driven roadmap for enhancing business performance in the hospitality sector (Mehnaz et al., 2023).

5.3 Practical implications

Practically, these findings provide actionable insights for hotel managers aiming to enhance customer satisfaction and improve brand equity. Emphasising room quality, cleanliness, and overall experience is crucial for meeting customer expectations and fostering positive perceptions. By maintaining high standards in these areas, hotels can consistently deliver quality experiences, resulting in increased customer satisfaction and loyalty. Investing in hotel facilities, services, and staff interactions is crucial for shaping customer perceptions. Providing superior amenities, streamlining check-in/out processes, and cultivating friendly and professional staff interactions contribute to positive customer experiences and enhance brand equity (Parzych and Brkić-Vejmelka, 2020; Sürücü et al., 2019; Zarezadeh et al., 2022).

Moreover, the location and convenience of a hotel play a significant role in customer satisfaction. Hotels strategically located near attractions, markets, restaurants, and city centres provide added value and convenience, contributing to positive customer perceptions and enhancing brand equity (Topic #10) (Hanaysha, 2016).

The study transcends the theoretical realm, delivering actionable insights that entrepreneurs and hospitality managers can leverage to refine customer engagement strategies and operational frameworks. Insights gleaned from the analysis illuminate pathways for enhancing customer satisfaction, elevating service quality, and optimising the overall experience. For example, integrating these findings into strategic planning enables hotel managers to prioritise investments in areas highly valued by customers, such as room quality and cleanliness. This strategic alignment directly influences customer perceptions and fosters brand loyalty. Furthermore, the study serves as a valuable resource for policymakers, offering empirical evidence to inform policies that promote industry standards and bolster the competitive edge of the hospitality sector.

6 Limitations and future works

This research provides a detailed exploration of customer perceptions and brand equity in the hospitality sector, primarily drawing from Booking.com reviews centred on Ho Chi Minh City. However, the study's reliance on a singular platform and location may not capture the diverse customer viewpoints found on different platforms or regions. Additionally, the exclusive use of six ML and DL sentiment analysis models alongside LDA and LSA for topic modelling may miss certain intricate sentiments or thematic details in the reviews, and the primary focus on textual data overlooks other UGC attributes.

For subsequent investigations, expanding to multiple data sources and considering diverse geographical contexts can enrich our understanding. Delving into other advanced sentiment analysis and topic modelling techniques, incorporating non-textual UGC, and examining the role of temporal factors on sentiments can offer more comprehensive insights. Such refinements in future research can significantly bolster findings and inform best practices in hospitality management.

7 Conclusions

This study demonstrates the potential of sentiment analysis and topic modelling in understanding customer perception and brand equity in the hospitality sector. Using ML and DL techniques on Booking.com data, this study identified vital aspects impacting on customer satisfaction. While the study is limited to one city and a specific set of models, it underscores the power of data-driven approaches in enhancing customer experience and brand equity. Future research could expand on this foundation, encompassing a broader geographical scope and leveraging various analytical tools, ultimately striving to improve the quality of hospitality services.

Acknowledgements

We acknowledge Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

References

Abid, F., Li, C. and Alam, M. (2020) 'Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks', *Computer Communications*, Vol. 157, pp.102–115.

Abkenar, S.B., Kashani, M.H., Mahdipour, E. and Jameii, S.M. (2021) 'Big data analytics meets social media: a systematic review of techniques, open issues, and future directions', *Telematics and Informatics*, Vol. 57, p.101517.

Albalawi, R., Yeap, T.H. and Benyoucef, M. (2020) 'Using topic modeling methods for short-text data: a comparative analysis', *Frontiers in Artificial Intelligence*, Vol. 3, p.42.

Bird, S. (2006) 'NLTK: the natural language toolkit', Paper presented at the *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*.

Bisoi, S., Roy, M. and Samal, A. (2020) 'Impact of artificial intelligence in the hospitality industry', *International Journal of Advanced Science and Technology*, Vol. 29, No. 5, pp.4265–4276.

Blasco-Arcas, L., Lee, H-H.M., Kastanakis, M.N., Alcañiz, M. and Reyes-Menendez, A. (2022) 'The role of consumer data in marketing: a research agenda', *Journal of Business Research*, Vol. 146, pp.436–452.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) 'Latent Dirichlet allocation', *Journal of Machine Learning Research*, January, Vol. 3, pp.993–1022.

Brahmbhatt, D. and Shah, J. (2017) 'Determinants of brand equity from the consumer's perspective: a literature review', *IUP Journal of Brand Management*, Vol. 14, No. 4, pp.33–46.

Chang, Z. (2022) 'A survey of modern crawler methods', Paper presented at the *6th International Conference on Control Engineering and Artificial Intelligence*.

Chaw, L.Y. and Tang, C.M. (2019) 'Online accommodation booking: what information matters the most to users?', *Information Technology & Tourism*, Vol. 21, No. 3, pp.369–390.

Debortoli, S., Müller, O., Junglas, I. and Vom Brocke, J. (2016) 'Text mining for information systems researchers: an annotated topic modeling tutorial', *Communications of the Association for Information Systems (CAIS)*, Vol. 39, No. 1, p.7.

Essien, A. and Chukwukelu, G. (2022) 'Deep learning in hospitality and tourism: a research framework agenda for future research', *International Journal of Contemporary Hospitality Management*, Vol. 34, No. 12, pp.4480–4515.

Górska-Warsewicz, H. and Kulykovets, O. (2020) 'Hotel brand loyalty – a systematic literature review', *Sustainability*, Vol. 12, No. 12, p.4810.

Hanaysha, J. (2016) 'Restaurant location and price fairness as key determinants of brand equity: a study on fast food restaurant industry', *Business and Economic Research*, Vol. 6, No. 1, pp.310–323.

Hannigan, T.R., Haans, R.F., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S. and Jennings, P.D. (2019) 'Topic modeling in management research: rendering new theory from textual data', *Academy of Management Annals*, Vol. 13, No. 2, pp.586–632.

Hossin, M. and Sulaiman, M.N. (2015) 'A review on evaluation metrics for data classification evaluations', *International Journal of Data Mining & Knowledge Management Process*, Vol. 5, No. 2, p.1.

Iglesias, O., Markovic, S., Singh, J.J. and Sierra, V. (2019) 'Do customer perceptions of corporate services brand ethicality improve brand equity? Considering the roles of brand heritage, brand image, and recognition benefits', *Journal of Business Ethics*, Vol. 154, pp.441–459.

Kirilenko, A.P., Stepchenkova, S.O., Kim, H. and Li, X. (2018) 'Automated sentiment analysis in tourism: comparison of approaches', *Journal of Travel Research*, Vol. 57, No. 8, pp.1012–1025.

Kolomoyets, Y. and Dickinger, A. (2023) 'Understanding value perceptions and propositions: a machine learning approach', *Journal of Business Research*, Vol. 154, p.113355.

Lamberton, C. and Stephen, A.T. (2016) 'A thematic exploration of digital, social media, and mobile marketing: research evolution from 2000 to 2015 and an agenda for future inquiry', *Journal of Marketing*, Vol. 80, No. 6, pp.146–172.

Liat, C.B., Mansori, S. and Huei, C.T. (2014) 'The associations between service quality, corporate image, customer satisfaction, and loyalty: evidence from the Malaysian hotel industry', *Journal of Hospitality Marketing & Management*, Vol. 23, No. 3, pp.314–326.

Liu, Z. and Park, S. (2015) 'What makes a useful online review? Implication for travel product websites', *Tourism Management*, Vol. 47, pp.140–151.

Macieira, F., Oliveira, T. and Yanaze, M. (2020) 'Models of satisfaction antecedents: a brief review. An integrative literature review of the most discussed satisfaction models in marketing studies', *International Journal of Services and Operations Management*, Vol. 36, No. 3, pp.348–359.

Mariani, M. and Borghi, M. (2020) 'Environmental discourse in hotel online reviews: a big data analysis', *Journal of Sustainable Tourism*, Vol. 29, No. 5, pp.829–848.

Martínez, P. and Nishiyama, N. (2019) 'Enhancing customer-based brand equity through CSR in the hospitality sector', *International Journal of Hospitality & Tourism Administration*, Vol. 20, No. 3, pp.329–353.

Mehnaz, J.J., Ahmad, W. and Hussain, A. (2023) 'Factors predicting customer satisfaction in online hotel booking using machine learning technique: evidence from developing countries', *International Journal of Applied Decision Sciences*, Vol. 16, No. 6, pp.741–759.

Mehraliyev, F., Chan, I.C.C. and Kirilenko, A.P. (2022) 'Sentiment analysis in hospitality and tourism: a thematic and methodological review', *International Journal of Contemporary Hospitality Management*, Vol. 34, No. 1, pp.46–77.

Mustak, M., Salminen, J., Plé, L. and Wirtz, J. (2021) 'Artificial intelligence in marketing: topic modeling, scientometric analysis, and research agenda', *Journal of Business Research*, Vol. 124, pp.389–404.

Parzych, K. and Brkić-Vejmelka, J. (2020) 'Guests' assessment of hotel facilities and services: Zadar case study', *European Journal of Tourism, Hospitality and Recreation*, Vol. 10, No. 3, pp.241–250.

Puh, K. and Bagić Babac, M. (2022) 'Predicting sentiment and rating of tourist reviews using machine learning', *Journal of Hospitality and Tourism Insights*, Vol. 6, No. 3, pp.1188–1204.

Raguseo, E., Neirotti, P. and Paolucci, E. (2017) 'How small hotels can drive value their way in infomediation. The case of 'Italian hotels vs. OTAs and TripAdvisor'', *Information & Management*, Vol. 54, No. 6, pp.745–756.

Salmony, M.Y.A. and Faridi, A.R. (2021) 'Supervised sentiment analysis on Amazon product reviews: a survey', Paper presented at the *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*.

Serra-Cantallops, A., Ramon Cardona, J. and Salvi, F. (2020) 'Antecedents of positive eWOM in hotels. Exploring the relative role of satisfaction, quality and positive emotional experiences', *International Journal of Contemporary Hospitality Management*, Vol. 32, No. 11, pp.3457–3477.

Sheehan, M., Grant, K. and Garavan, T. (2018) 'Strategic talent management: a macro and micro analysis of current issues in hospitality and tourism', *Worldwide Hospitality and Tourism Themes*, Vol. 10, No. 1, pp.28–41.

Sürücü, Ö., Öztürk, Y., Okumus, F. and Bilgihan, A. (2019) 'Brand awareness, image, physical quality and employee behavior as building blocks of customer-based brand equity: consequences in the hotel context', *Journal of Hospitality and Tourism Management*, Vol. 40, pp.114–124.

Ullah, M.A., Marium, S.M., Begum, S.A. and Dipa, N.S. (2020) 'An algorithm and method for sentiment analysis using the text and emoticon', *ICT Express*, Vol. 6, No. 4, pp.357–360.

Ye, Q., Law, R., Gu, B. and Chen, W. (2011) 'The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings', *Computers in Human Behavior*, Vol. 27, No. 2, pp.634–639.

Zarezadeh, Z.Z., Rastegar, R. and Xiang, Z. (2022) 'Big data analytics and hotel guest experience: a critical analysis of the literature', *International Journal of Contemporary Hospitality Management*, Vol. 34, No. 6, pp.2320–2336.