

Article

A Deep Learning-Based Analysis of Customer Concerns and Satisfaction: Enhancing Sustainable Practices in Luxury Hotels

Tiantian Pang, Juan Liu *, Li Han, Haiyan Liu and Dan Yan

School of Management, Zhengzhou University, No. 100 Science Avenue, Gaoxin District, Zhengzhou 450001, China; pangtiantian@gs.zzu.edu.cn (T.P.); pyhanli@126.com (L.H.); lhyliuhaiyan09@163.com (H.L.); yandan@zzu.edu.cn (D.Y.)

* Correspondence: liujuanjune@zzu.edu.cn

Abstract: Hotels are one of the fastest-growing sectors in the tourism industry, and sentiment analysis plays a vital role in improving business performance and supporting sustainable practices. This paper proposes a novel framework combining topic mining and aspect-based sentiment analysis to examine 29,334 hotel reviews in Henan province in China, with the aim of informing strategies for sustainable hotel development. Our results reveal six key attributes of customer concern, particularly emphasizing family experiences, which reflect Henan's appeal as a family tourism destination. Additionally, we uncover sentiment quadruples, including categories, aspect terms, opinion terms, and polarities, thus enabling a dual-dimensional evaluation of factors influencing customer satisfaction. The results reveal that service mainly influences overall category-level satisfaction, while bed, front desk, and breakfast primarily drive aspect-level satisfaction. This study provides valuable insights into customer feedback, offering empirical support for optimizing services and guiding the sustainable strategic development of regional hotels.

Keywords: hotel online reviews; latent Dirichlet allocation; deep learning; topic mining; aspect-based sentiment analysis; hotel sustainability



Academic Editor: Robert E. Pitts

Received: 9 February 2025

Revised: 26 March 2025

Accepted: 3 April 2025

Published: 16 April 2025

Citation: Pang, T.; Liu, J.; Han, L.; Liu, H.; Yan, D. A Deep Learning-Based Analysis of Customer Concerns and Satisfaction: Enhancing Sustainable Practices in Luxury Hotels. *Sustainability* **2025**, *17*, 3603. <https://doi.org/10.3390/su17083603>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rapid development of the internet and social media has fundamentally reshaped consumer decision-making, particularly in the hospitality sector [1,2]. Online reviews, as a primary source of customer feedback, have become indispensable for both potential customers and hotel managers [3]. Consumers increasingly rely on these reviews to inform their choices, while hotel operators leverage them to gain valuable insights for service improvement [4]. The emotions conveyed in these reviews not only reflect customers' satisfaction with hotel services but also reveal their underlying needs and expectations [3]. In the post-COVID-19 era, shifting consumer behaviors and emotional expressions have led to a profound transformation in the drivers of customer satisfaction [5,6]. As a result, identifying the key determinants of customer satisfaction in this new context, and utilizing sentiment analysis to uncover the nuanced emotional needs of customers, is crucial for promoting the sustainable development of the hospitality sector.

Traditional sentiment analysis methods typically focus on document-level or sentence-level sentiment classification, categorizing online reviews as either positive or negative [7]. However, this approach oversimplifies the complexity of customer feedback. Even in predominantly negative reviews, customers often express positive sentiments about certain aspects of their experience [8]. For instance, a customer might write, "The room was clean, and the breakfast was delicious, but my biggest complaint is...". Although the

overall rating may be low, the comments about the room and breakfast are clearly positive. Traditional sentiment analysis techniques often fail to capture these subtle emotional variations [9]. In contrast, aspect-based sentiment analysis can identify multiple emotional dimensions within a single review, allowing for a more precise understanding of customer sentiment [9,10]. This approach provides deeper insights into customer emotions, enabling hotel managers to refine their service optimization strategies more effectively.

Aspect-based sentiment analysis, a sophisticated method of text analysis, enables a more thorough exploration of customer reviews through a structured analytical framework [7]. Initially, large-scale online review data are collected through web scraping to ensure that the sample is both comprehensive and representative [11]. Data preprocessing follows to clean and filter out irrelevant or noisy data, maintaining the integrity of the dataset for further analysis [11]. Topic mining techniques are then employed to identify key themes and focal points within the reviews, revealing customers' perceptions of various aspects of hotel services, amenities, and overall experience [12]. During the data labeling and model training phase, labeled data are used to train sentiment classification models, enhancing the accuracy and precision of sentiment analysis [13]. Additionally, advanced techniques such as entity recognition and relation extraction are used to uncover relationships between specific emotional expressions and hotel service attributes, offering a deeper understanding of customer sentiment [11]. This multi-step analytical process facilitates a more accurate and nuanced interpretation of customer emotions, providing valuable insights for service optimization and the development of sustainable strategies for the hotel industry.

Previous studies have explored the determinants of customer sentiment in hotel services through various methods of topic mining, including both manual content analysis and computer-based text analysis [14]. Manual content analysis, however, is often subjective, relying heavily on the researcher's interpretation and requiring substantial human efforts to categorize themes [8]. This results in limited replicability and small sample sizes, which may undermine the robustness of findings [15]. In contrast, computer-based text analysis techniques, such as word frequency analysis and traditional topic modeling, focus primarily on the frequency of terms without considering the contextual meaning of words [8]. This often leads to generalized conclusions that may overlook important emotional nuances. In response to these limitations, probabilistic topic models, such as Latent Dirichlet Allocation (LDA), have emerged as the dominant approach for analyzing hotel reviews [8]. LDA is capable of identifying latent topics within large-scale text data, uncovering deeper structures of customer sentiment, and providing more reliable and scalable results.

Furthermore, aspect-based sentiment analysis has traditionally relied on machine learning algorithms like support vector machines and random forests. These methods often require extensive manual feature engineering and struggle to capture the contextual relationships between words, which can lead to inaccurate sentiment classification, especially in complex reviews. Recent advancements in deep learning have addressed these issues with models like BERT-BiLSTM-CRF, which combine the strengths of Bidirectional Encoder Representations from Transformers (BERTs), Bidirectional Long Short-Term Memory (BiLSTM), and Conditional Random Fields (CRFs). BERT captures rich semantic features and contextual information, BiLSTM processes long-range dependencies, and CRF enhances entity recognition accuracy by constraining label predictions. This integrated approach has shown promise in sentiment analysis tasks, particularly for processing complex emotional content in customer reviews [16]. Furthermore, it enables the assessment of customer aspect-based sentiment with larger sample sizes, reduced costs, and greater data collection flexibility compared to traditional methods.

In the post-COVID-19 era, luxury spending has steadily increased as consumer demands evolve. The concept of luxury has garnered increasing attention, particularly within the tourism and hospitality industries. The Asia-Pacific region constitutes the largest market for luxury travel, accounting for approximately 33% of the market share [17]. Henan Province, located in central China, has experienced remarkable growth in the tourism industry in recent years [18]. Five-star hotels in the region have become increasingly pivotal in attracting both domestic and international visitors [19]. These hotels represent the upper echelon of the Chinese hospitality industry and serve as benchmarks for service quality, brand image, and market competitiveness. Therefore, analyzing luxury hotels in Henan can provide insights representative of the broader Chinese luxury hotel sector [19]. Despite the growing demand for luxury hotels in Henan, systematic research on customer sentiment and service optimization in this region remains limited. By focusing on five-star hotels in Henan, this study aims to fill this research gap and provide actionable insights to help hotel managers improve customer satisfaction, thereby contributing to the sustainable development of the hotel industry in Henan and beyond.

The objective of this manuscript is to bridge the existing gaps by analyzing 29,334 online reviews of five-star hotels in Henan Province, collected from the Ctrip platform between 2022 and 2024. To accomplish this objective, a comprehensive research framework was developed that integrates the LDA topic model with the deep learning-based BERT-BiLSTM-CRF model, combining both topic mining and aspect-based sentiment analysis. First, LDA and content analysis were employed to identify and summarize key themes and dimensions within the online reviews. Based on these findings, the data were annotated to align with the identified themes. Finally, a BERT-BiLSTM-CRF model was trained to perform aspect-based sentiment analysis, quantifying customer satisfaction and identifying actionable insights. This framework enables the automatic extraction of dimensions from customer reviews, quantifies satisfaction and importance, and pinpoints areas for improvement. Beyond the hotel industry, this method serves as a versatile, automated tool for analyzing consumer feedback across various economic domains, providing insights into sustainable development practices.

2. Literature Review

2.1. Satisfaction in Luxury Hotels

Numerous theories and frameworks have been developed to explain customer satisfaction. Oliver [20] validated the Expectation-Disconfirmation Model through experimental and survey-based research, proposing that consumers form expectations before purchasing a product and later compare their actual experience with those expectations. While this model provides a solid foundation for understanding cognitive discrepancies, it may not fully capture the emotional and subjective aspects of customers' experiences. Parasuraman, et al. [21] proposed the SERVQUAL model while identifying five key service quality dimensions, which may not fully capture the nuances of specialized sectors like luxury hospitality due to its focus on standardized services. Fornell, et al. [22] developed the Customer Satisfaction Index model, which integrates expectations, perceived quality, and perceived value to measure overall satisfaction; while comprehensive, the applicability of this model may vary across regions and industries. In contrast, this study leverages sentiment analysis of online reviews, a model that extracts customer sentiment from unstructured textual data [23]. This approach addresses the limitations of traditional survey-based methods by capturing the nuanced and often unsolicited feedback expressed in large volumes of online reviews, providing a more comprehensive and contextually relevant understanding of customer satisfaction.

The sustained growth of the tourism accommodation sector, coupled with the increasing demand for luxury experiences, has fueled scholarly interest in luxury hotel brands [24,25]. Within the luxury hotel context, satisfaction can be conceptualized as an overall evaluation of brand performance [26]. Recent studies have increasingly focused on luxury hotel satisfaction within developing economies. For example, Shin and Jeong [27] explored the impact of technology on luxury hotel guest satisfaction and loyalty. Khoi and Le [17], Padma and Ahn [25] investigated the influence of perceived “coolness” on satisfaction within the luxury hotel context. Padma and Ahn [25] analyzed big data in the form of online reviews, using content analysis to identify key themes related to luxury hotel service quality and applying critical incident techniques to examine the antecedents and consequences of guest satisfaction and dissatisfaction. Ismail, Zahari, Hanafiah and Balasubramanian [26] collected experiential data from 482 customers of luxury hotel restaurants through structured surveys and used structural equation modeling to validate that customer brand personality had a positive and significant impact on their dining experience, thereby influencing their satisfaction with luxury hotels.

While previous studies have often relied on questionnaires to understand the antecedents of consumer satisfaction with hotel brands [28], the digital age has witnessed the rise in online reviews and sentiment analysis as critical sources for studying customer satisfaction [29]. In this context, big data analytics and deep learning techniques offer promising opportunities for gathering online consumer insights [29]. However, the application of big data analytics tools remains relatively underexplored [30]. Therefore, developing innovative approaches to monitor online reviews is essential for understanding the unique characteristics of luxury hotels. Deep learning techniques hold substantial potential in facilitating this monitoring process [29].

2.2. Sentiment Analysis of Online Reviews

Online reviews, published by consumers on e-commerce platforms or social networking applications, serve as third-party evaluations that facilitate user interaction and assist potential buyers in making informed decisions [31,32]. Sentiment analysis is widely applied in online review research, spanning various domains such as recommendation systems, e-commerce, agriculture, tourism, crowdsourcing service provider selection, online course evaluations, and hospital reviews [9]. For instance, it examined Japanese restaurant reviews in English at Yelp.com and those in Japanese at Yelp.co.jp, demonstrating that Japanese customers have significantly different sentiment distribution patterns on four basic attributes of dining experience than Western customers. A sentiment analysis method was used to extract the sentiment index from the content of each online review, and historical sales data were developed for product sales forecasting. It proposed an improved word representation method that incorporates sentiment information into the traditional TF-IDF algorithm, generating weighted word vectors to more effectively capture contextual information and improve the representation of comment vectors. Chan and Chong [33] proposed a sentiment analysis engine, which takes advantage of linguistic analyses based on grammar and addresses key questions related to the explosion of interest in how to extract insight from unstructured data and how to determine if such insight provides any hints concerning the trends of financial markets. Additional questions were how to identify the structure of online restaurant reviews and examine the influence of review attributes and sentiments on restaurant star ratings.

2.3. Review of Aspect-Based Sentiment Analysis

Aspect-based sentiment analysis, an important branch of natural language processing, focuses on extracting specific sentiment information related to different aspects or entities

within a text [7]. Unlike traditional coarse-grained sentiment analysis, which primarily examines the overall sentiment polarity of a text, aspect-based sentiment analysis delves deeper into the emotional attitudes expressed towards specific aspects or attributes within the text. Aspect-based sentiment analysis has evolved significantly over the years, with three main phases of development: lexicon-based, traditional machine learning-based, and deep learning-based approaches [34].

Early aspect-based sentiment analysis methods relied on lexicon-based techniques, using predefined sentiment dictionaries to identify sentiment-related words associated with specific aspects [9]. These methods were straightforward but limited in their ability to handle emerging vocabulary, sarcasm, and contextual nuances. As a result, the field shifted towards machine learning-based approaches, which employed feature engineering and supervised learning algorithms (e.g., Naive Bayes and SVM) to classify sentiment related to specific aspects [7]. While these models improved scalability and flexibility, they still struggled with capturing complex and contextual sentiment expressions, prompting further advancements.

The most significant breakthrough in aspect-based sentiment analysis came with deep learning-based approaches. Deep learning-based approaches leverage contextual sequential information to extract textual features and semantic information for sentiment analysis. For instance, a stacked ensemble method was proposed for predicting the degree of intensity for emotion and sentiment by combining the outputs obtained from several deep learning and classical feature-based models using a multi-layer perceptron network. A deep learning-based sentiment analysis method was introduced that combines TF-IDF weighted Glove word embeddings with a CNN-LSTM architecture to analyze product reviews from Twitter. In a previous study, an efficient sentiment classification system was developed for instructor evaluation reviews by adopting a deep learning paradigm. A layered deep learning model was designed to capture sequential information within market snapshot series, constructed from technical indicators and news sentiment, for stock prediction purposes. Bello, et al. [35] extended this research by applying BERT for text classification in natural language processing tasks, demonstrating that BERT and its variants can effectively express sentiment based on user context.

3. Materials and Methods

This study employed a systematic five-step approach. Firstly, we collected and preprocessed a comprehensive dataset of online reviews from five-star hotels in Henan Province via the Ctrip platform. Secondly, we applied the LDA model to identify the optimal number of topics and extract the core dimensions of luxury hotels. Thirdly, we manually annotated 2000 reviews to train a BERT model, creating an aspect-based sentiment analysis framework specifically designed for Henan's hotel reviews. Fourthly, we deployed the BERT-BiLSTM-CRF model across the entire dataset to compute satisfaction and importance scores for each dimension. Finally, we analyzed the findings to identify customer priorities and propose targeted recommendations for enhancing hotel management and service quality. The overall framework is presented in Figure 1.

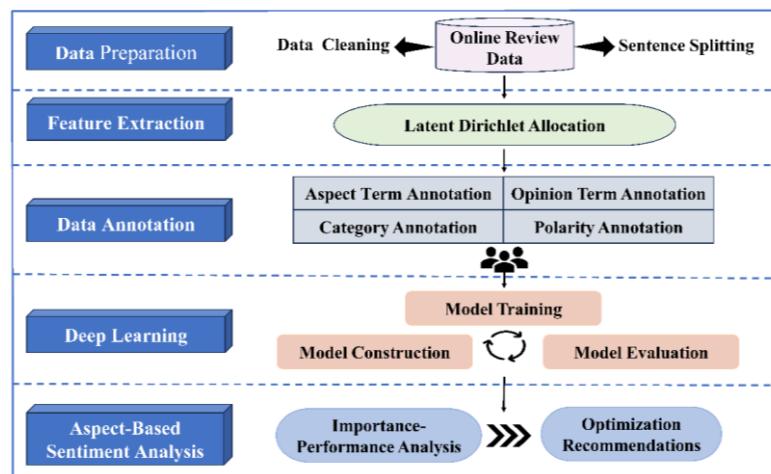


Figure 1. A conceptual framework for aspect-based sentiment analysis of hotels.

3.1. Latent Dirichlet Allocation (LDA)

3.1.1. LDA Model

LDA is a Bayesian generative probabilistic model widely used for topic mining in large-scale textual datasets [36]. The model operates on three hierarchical layers: documents, topics, and words. Each document is conceptualized as a mixture of latent topics, and each topic is represented by a probabilistic distribution of words. Both document-to-topic and topic-to-word relationships follow multinomial distributions, governed by parameters derived from Dirichlet distributions.

The generative process of LDA begins by sampling parameters for topic distributions using the Dirichlet distribution, followed by randomly generating a topic distribution for a document. For each word position within a document, a topic is selected based on this distribution, and a word is then generated using the topic's word distribution. This process is repeated for all words in the document and extended to the entire corpus. By efficiently capturing context-specific topics, LDA is particularly suited for analyzing unstructured online reviews, providing a robust foundation for extracting meaningful insights. Figure 2 outlines the steps of the LDA model.

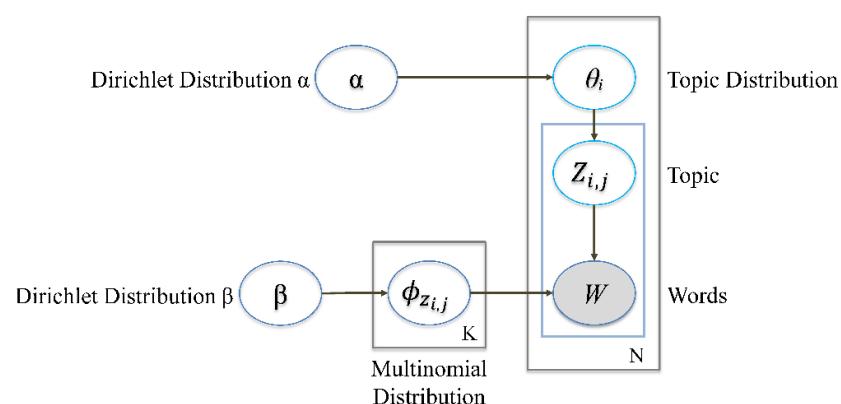


Figure 2. The basic working principle of LDA.

3.1.2. LDA Parameters

Topic perplexity measures the uncertainty of assigning a document to a given topic [37]. A lower perplexity indicates a better fit of the topic model to the data. However, as the number of topics increases, perplexity typically decreases before plateauing, suggesting

that further increases in topics no longer improve model performance. The perplexity calculation is defined as

$$\text{Perplexity} = \exp\left(-\frac{\sum_{d=1}^D \log P(W_d)}{\sum_{d=1}^D N_d}\right) \quad (1)$$

where D represents the total number of documents, $P(W_d)$ is the probability of generating document d , and N_d is the total number of words in document d .

Topic coherence evaluates the semantic consistency of high-probability words within a topic and the distinctiveness between different topics [37]. A higher coherence score reflects better differentiation and interpretability of the topics generated by the model. Coherence is calculated using the following formula:

$$C_v = \frac{1}{\text{pairs}} \sum_{w_i, w_j \in T} \text{PMI}(w_i, w_j) \quad (2)$$

$$\text{PMI}(w_i, w_j) = \log \frac{P(w_i, w_j)}{P(w_i)P(w_j)} \quad (3)$$

where T is the set of keywords for a given topic, w_i and w_j are two words within the same topic, and $\text{PMI}(w_i, w_j)$ (Pointwise Mutual Information) quantifies the likelihood of their co-occurrence. High PMI values indicate a stronger association between the words within the same topic.

3.2. Deep Learning

This study proposes a deep learning approach for aspect-based sentiment analysis by utilizing the BERT-BiLSTM-CRF model to automatically extract sentiment-related aspects and opinions from text. The model combines named entity recognition (NER) and relation extraction (RE) tasks, aiming to improve the extraction of relevant sentiment information at the aspect level. The model architecture is illustrated in Figure 3.

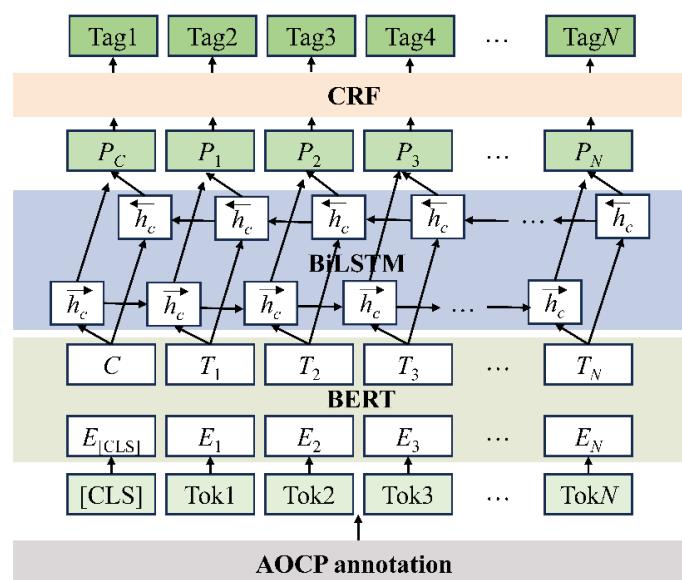


Figure 3. Structure diagram of deep learning for aspect-based sentiment.

3.2.1. AOCP Annotation

In aspect-based sentiment analysis, it is crucial not only to identify the categories and polarities within a sentence but also to accurately align aspect terms with opinion terms [38].

To address this need, we developed an AOCP annotation system, where “A” represents aspect terms, “O” stands for opinion terms, “C” refers to the category, and “P” indicates the polarity of the sentence. This system supports multi-task joint learning, allowing various sub-tasks of aspect-based sentiment analysis to be integrated into a unified model. This integration ensures comprehensive optimization rather than local optimization, thus enhancing the effectiveness of aspect-based sentiment analysis. The AOCP annotation process comprises four main steps:

What (aspect terms annotation): This step involved extracting specific attributes mentioned in the text to understand the focal points of user attention.

Why (opinion terms annotation): We extracted sentiment-related vocabulary associated with these attributes to gauge users' attitudes or feelings toward each aspect.

How (polarities annotation): The extracted sentiment terms were classified by polarity, determining whether the sentiment expressed was positive or negative, thereby allowing for an assessment of user satisfaction.

Where (categories annotation): This dimension involved identifying the specific categories to which the aspect and sentiment terms belong, enabling a more systematic analysis and organization of information to better understand user feedback characteristics across different categories.

3.2.2. BERT Model

BERT is a pre-trained language model based on the transformer architecture. Each transformer layer contains a multi-head self-attention mechanism and a feedforward neural network that work together to enhance the model's ability to handle complex semantics. This structure not only allows the model to better understand the nuances of language but also achieves higher accuracy in sentiment analysis tasks. Figure 4 illustrates the basic principles of BERT for sentiment analysis. In the figure, “Class Label” represents the classification label, which is the target category that the model needs to predict, while “Sentence 1” and “Sentence 2” represent different input sentences [39].

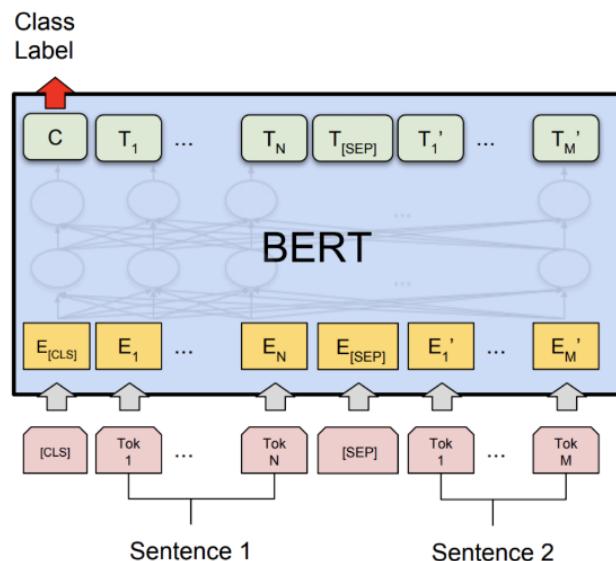


Figure 4. The basic working principle of BERT.

The self-attention mechanism in the Transformer layer can be mathematically expressed as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where Q, K, V are the Query, Key, and Value matrices, obtained through different linear transformations of the input.

3.2.3. BiLSTM Model

LSTM is a type of neural network model designed to process sequential data. It consists of an input gate, a forget gate, an output gate, and memory cells, as shown in Figure 5. Unlike traditional neural networks, LSTM is more suitable for handling long sequences, such as those found in online review data. However, a unidirectional LSTM model can only process information in one direction, which limits its ability to capture context from both the past and future. To overcome this limitation, this study employs a BiLSTM model. The BiLSTM model consists of two LSTM units, one processing information in the forward direction and the other in the backward direction, allowing the model to capture context from both directions and better understand long-range dependencies in the input sequence [40].

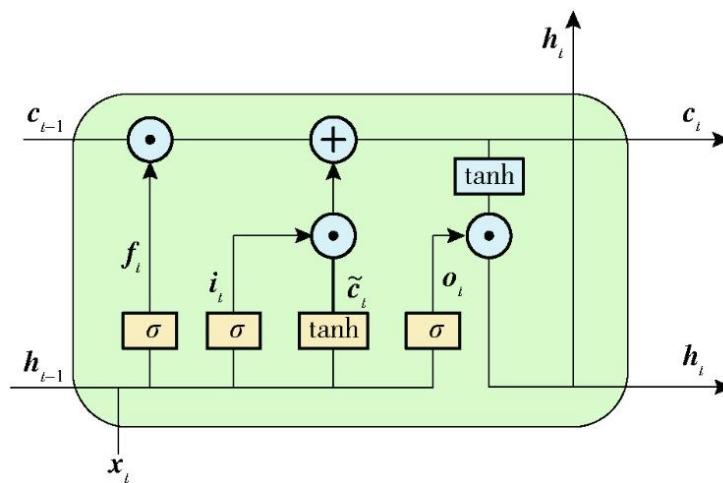


Figure 5. The basic working principle of LSTM.

The computation formulas for the gates in the BiLSTM layer are as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \\ z_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \\ c_t &= f_t c_{t-1} + i_t z_t \\ o_t &= \tanh(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \\ h_t &= o_t \tanh(c_t) \end{aligned} \quad (5)$$

where i_t , f_t and o_t control the flow of information into, out of, and within the memory cell c_t , which stores long-term information, and h_t represents the short-term output. z_t introduces updates at each time step. The weights W_{xi} , W_{xc} , W_{xf} , W_{xo} , etc., connect the input layer, hidden layer, and gates, while the bias terms b_i , b_c , b_f , b_o adjust computations. Activation functions sigmoid σ and tanh ensure non-linearity and stability.

3.2.4. CRF Model

The BiLSTM model effectively captures long-range dependencies in sequences but struggles to account for dependencies between adjacent labels. To address this, a CRF model is incorporated. The CRF models the dependencies between labels by defining transition scores and ensures that the predicted label sequences are globally optimal [41].

The probability of a label sequence $Y = \{y_1, y_2, \dots, y_n\}$ given an input sequence $X = \{x_1, x_2, \dots, x_n\}$ is defined as

$$P(Y|X) = \frac{\exp(\sum_{i=1}^n (A_{y_{i-1}, y_i} + P_{i, y_i}))}{\sum_{Y'} \exp(\sum_{i=1}^n (A_{y_{i-1}, y'_i} + P_{i, y'_i}))} \quad (6)$$

where P_{i, y_i} is the score from the BiLSTM layer for the label y_i at position i , and A_{y_{i-1}, y_i} represents the transition probability from label y_{i-1} to y_i . By leveraging dynamic programming, the CRF layer computes the globally optimal label sequence, capturing dependencies between labels and improving sequence labeling performance.

3.2.5. Evaluation Metrics

Aspect-based sentiment analysis can be considered a specialized form of text classification, so its evaluation metrics are similar to those used in text classification tasks. This study selected four key metrics to evaluate the performance of the models: accuracy, precision, recall, F1-score, and loss. Together, these metrics form a comprehensive evaluation framework for assessing the models.

Accuracy measures the overall correctness of predictions and is defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision indicates the proportion of correctly predicted positive samples among all samples predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

Recall represents the proportion of correctly predicted positive samples among all actual positive samples:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

F1-score is the weighted harmonic mean of precision and recall:

$$F1 - score = \frac{2PR}{P + R} \quad (10)$$

where TP represents true positives (correctly predicted positive cases), FP represents false positives (incorrectly predicted positive cases), FN represents false negatives (incorrectly predicted negative cases), TN represents true negatives (correctly predicted negative cases).

The loss function quantifies the discrepancy between the model's predictions and the actual labels, guiding the optimization process. In this study, the Cross-Entropy Loss is employed, defined as follows:

$$\mathcal{L}_{start} = -\frac{1}{N} \sum_{i=1}^N y_i^{start} \log(p_i^{start}) \quad (11)$$

$$\mathcal{L}_{end} = -\frac{1}{N} \sum_{i=1}^N y_i^{end} \log(p_i^{end}) \quad (12)$$

where N denotes the sequence length in the sample, y_i^{start} and y_i^{end} represent the ground truth labels for the start and end positions, respectively, while p_i^{start} and p_i^{end} are the

predicted probability distributions for the start and end positions. The total loss is computed as the sum of the start and end losses:

$$\mathcal{L} = \mathcal{L}_{start} + \mathcal{L}_{end} \quad (13)$$

3.3. Study Area and Data

Henan Province, one of China's most populous regions and a key tourist destination in China, has experienced remarkable growth in its tourism sector, driven by its rich historical and cultural heritage (see Figure 6). Strategically located in central China, Henan serves as a major transportation hub, making its hotel industry increasingly vital for meeting the demands of both domestic and international travelers. Existing literature on hotel customer satisfaction predominantly focused on China's developed coastal regions, leaving a research gap concerning inland provinces. As a representative inland region, Henan offers a valuable context for examining hotel industry dynamics. However, despite the rise in the number of five-star hotels in Henan, studies on customer satisfaction and sentiment analysis, particularly aspect-based sentiment analysis that captures detailed customer preferences, remain scarce. This study investigated customer satisfaction in five-star hotels within Henan Province, employing the LDA topic model and BERT-BiLSTM-CRF model to identify key themes and sentiment characteristics in customer reviews. The findings are likely to provide empirical insights that support service improvements in regional hotels and contribute to a deeper, more nuanced understanding of customer preferences in this underexplored context.

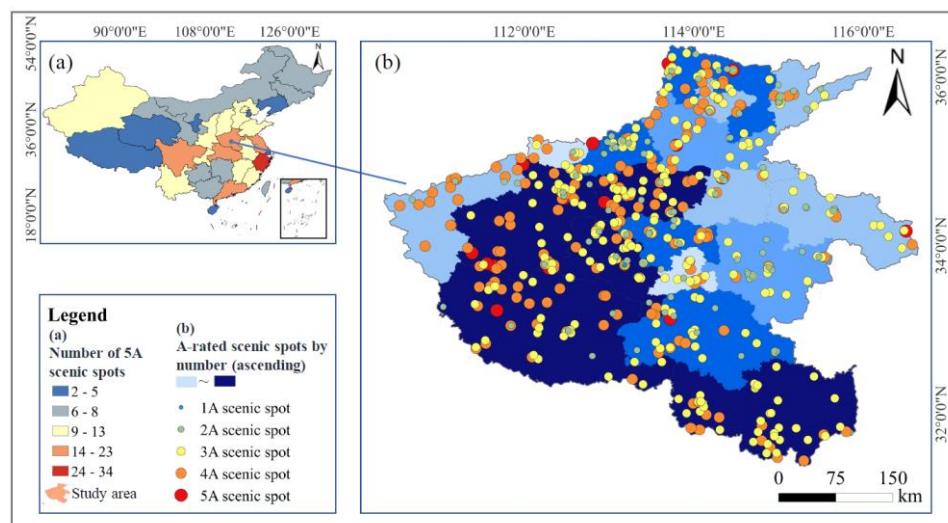


Figure 6. Spatial distribution of A-rated scenic spots in Henan Province, China.

The data used in this study were sourced from 29,334 reviews of five-star hotels in Henan Province, collected from the Ctrip website between January 2022 and May 2024. The dataset was first preprocessed, including data cleaning, tokenization, and stop-word removal. Data cleaning removed irrelevant and erroneous entries, while tokenization segmented the text into individual words or phrases. Stop-word removal filtered out common terms that might interfere with sentiment analysis. Feature extraction and manual annotation were then carried out, where key features were identified, categorized, and labeled for sentiment analysis.

4. Results

4.1. Topic Mining of Hotels Based on LDA Model

4.1.1. Determining the Number of Topics

The number of topics is a crucial parameter in the LDA model, directly influencing the effectiveness of topic mining and the interpretability of the results. In this study, the number of topics was determined using the Gensim natural language processing library [42]. We computed perplexity and coherence values for different topic numbers and plotted them to evaluate the optimal topic count. Generally, lower perplexity values indicate better model performance in predicting text sequences, while higher coherence values suggest better interpretability and topic differentiation.

Figure 7 clearly shows that the highest coherence value was achieved with 15 topics, indicating that the topics were highly interpretable. While the perplexity value was relatively high, coherence is prioritized in LDA topic modeling, particularly when the goal is to enhance topic interpretability. Therefore, we selected 15 topics as the optimal number for this study, based on the balance between coherence and perplexity, which provided the most meaningful topic structure for the reviews.

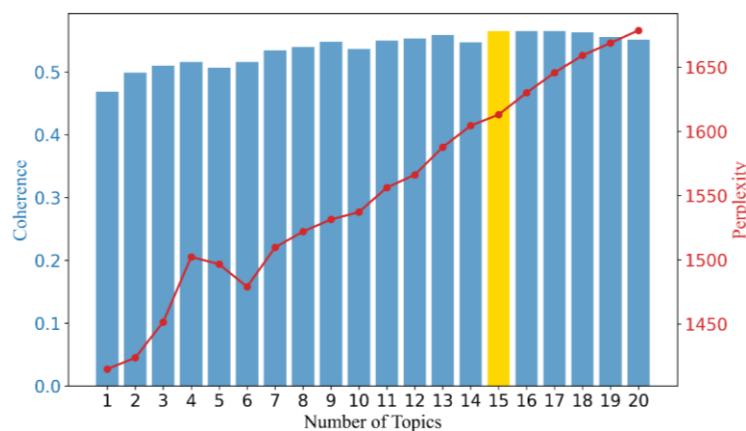


Figure 7. Perplexity–topic and coherence–topic line charts.

4.1.2. Topic Mining

Topic mining was performed using the LDA model, implemented through the Gensim library, to automatically extract latent semantic themes from the corpus of online reviews. Following Luo, et al. [43], the initial 15 topics identified by the LDA model were thematically aggregated into six overarching attributes characterizing the luxury hotel experience: service, facilities, environment, dining, convenience, and family experience. The correspondence between the initial topics and these aggregated attributes is presented in Table 1.

Service Dimension: This dimension includes aspects closely related to service quality, such as front desk reception, check-in and check-out efficiency, as well as staff attitude and problem-solving capabilities. Keywords such as “service”, “front desk”, “attitude”, and “arrangements” reflect customer concerns regarding service processes and staff interactions.

Facilities Dimension: This dimension focuses on both hardware and software aspects of the hotel, including beds, air conditioning, décor, and bathrooms, reflecting customer expectations for comfort and the availability of amenities.

Environment Dimension: This dimension emphasizes the overall ambiance and cleanliness of the hotel. Keywords such as “cleanliness”, “quiet”, “hygiene”, and “air” suggest that customers have high expectations for the comfort, tranquility, and tidiness of the environment.

Table 1. The six core dimensions and their corresponding topics and keywords.

No.	Dimension	Relevant Topic	Keywords Included	Explanation
1	Service	2, 3, 6, 14, 15	Service, front desk, check-in, handling, arrangement, pick-up, etc.	Covers keywords related to service quality, such as front desk and attitude, involving check-in experience, problem-solving, and more.
2	Facilities	8, 12	Facilities, air conditioning, equipment, completeness, decoration, bed, etc.	Primarily includes hardware and software facilities, as well as in-room accommodation amenities.
3	Environment	7, 9	Environment, cleanliness, hygiene, ambiance, air, cleanliness, etc.	Focuses on the overall environment and atmosphere of the hotel, including cleanliness, quietness, comfort, and more.
4	Dining	4, 11, 13	Breakfast, buffet, late-night snacks, taste, restaurant, flavor, etc.	Involves the quality and experience of breakfast and dining services, including the variety and taste of breakfast items.
5	Convenience	1, 5, 13	Location, transportation, high-speed rail, airport, tourist attractions, surroundings, etc.	Pertains to the convenience and accessibility of the hotel, including its location, transportation, proximity to tourist spots, and transport hubs.
6	Family Experience	10, 7	Kids, little ones, family, children, amusement park, baby, etc.	Relates to the experience of family customers, assessing whether the hotel is suitable for families and what child-related facilities are provided.

Dining Dimension: This dimension focuses on the quality and taste of meals, including breakfast, buffet, and restaurant services, reflecting customer feedback on the variety of dishes, freshness of ingredients, and dining services.

Convenience Dimension: This dimension reflects the hotel's location and transportation convenience, including proximity to transportation hubs, tourist attractions, and nearby amenities. Keywords such as "location", "transportation", and "parking" indicate customer emphasis on the ease of access to the hotel.

Family Experience Dimension: Primarily aimed at family tourists, this dimension evaluates whether the hotel is suitable for children and families, such as offering play facilities and children's meals. Keywords such as "children", "family", and "kids" highlight the demand for services tailored to family and child-friendly accommodations.

4.1.3. Data Annotation

Based on the six key dimensions identified earlier, we classified hotel reviews into six distinct categories: convenience, service, dining, environment, facilities, and family experience. A random sample of 2000 reviews was selected, which underwent categorization, cleaning, and sentence segmentation. The samples were then annotated using the AOCP annotation system, resulting in a total of 5981 annotated sentences. Table 2 presents a subset of the annotated data for reference.

To present the annotated data in a more structured format that facilitates subsequent input into models for training and prediction, a crucial step involved standardizing the AOCP annotation format into a structured annotated dataset. To achieve this, we utilized Python (PyCharm 2021.3) to transform the data to meet the requirements of the annotated data format. The sequence labeling format separates the text and labels, employing an array of labels to indicate the category of each word. Labels are represented in a tagging format (e.g., "O", "B-Service"), making it suitable for NER and sentiment analysis tasks.

In contrast, the annotated data format includes text, start and end positions, and aspect information, which is well-suited for tasks requiring detailed positional information, such as aspect-based sentiment analysis. An example of the transformed data format is presented in Table 3.

Table 2. Sample annotated data from online reviews.

ID	Aspect Terms	A-Start	A-End	Opinion Terms	O-Start	O-End	Categories	Polarities	Text
1	Breakfast	1	2	Plentiful	3	4	Dining	Positive	The breakfast is plentiful
2	Attitude	1	2	Friendly	4	5	Service	Positive	The attitude is very friendly
3	Equipment	1	2	Outdated	4	5	Facilities	Negative	The equipment look rather outdated
4	Breakfast	1	2	Good	4	5	Dining	Positive	The breakfast is also good
5	Cleaning lady	1	3	Attentive	5	6	Service	Positive	The cleaning lady was very attentive
6	Service staff	1	3	Enthusiastic	5	6	Service	Positive	The service staff are very enthusiastic
.....									

Table 3. Example of AOCP annotation data converted to structured annotation format.

Example 1: Sequence Labeling Format	Example 2: Annotated Data Format
<pre>"text": ["The", " Service", "of", "Dream", "Back", "to", "the", "Tang", "Dynasty", "is", "also", "good"], "labels": ["O", "O", "O", "O", "B-Service", "I-Service", "B-Positive", "I-Positive", "I-Positive"]</pre>	<pre>"id": 3639, "text": ["The", "scenery", "is", "beautiful"], "start": [0, 0, 1, 0], "end": [0, 0, 0, 1], "aspect": "scenery"</pre>

4.2. Aspect-Based Sentiment Analysis of Hotels Based on Deep Learning

4.2.1. Model Training for Aspect-Based Sentiment Analysis of Hotels

To conduct a more detailed sentiment analysis of hotel online reviews, we employed the BERT-BiLSTM-CRF model, which supports multi-task joint learning and integrates various sub-tasks of aspect-based sentiment analysis. This model uses contextual semantic understanding to perform nuanced sentiment analysis of hotel reviews. Using the AOCP-annotated dataset, we conducted a joint analysis of NER and RE, enabling the extraction of review categories along with their sentiment insights, as well as aspect-based aspect and sentiment extraction. The model was trained over six epochs with a learning rate of 3×10^{-5} , utilizing the Adam optimizer and a batch size of 64.

Table 4 presents the evaluation results of the NER task of the BERT-BiLSTM-CRF model across six categories. While some categories, such as dining and environment, exhibited relatively lower recall and precision, the overall performance indicates that the model has strong entity recognition capabilities. Notably, the model achieved F1-scores of 0.84 and 0.82 in the family experience and convenience categories, demonstrating excellent precision and recall. These results suggest that the model is capable of effectively recognizing entities within these categories and providing a reliable foundation for subsequent tasks.

Table 4. BERT-BiLSTM-CRF model NER training results.

Categories	Precision	Recall	F1-Score
Family Experience	0.83	0.84	0.84
Convenience	0.88	0.78	0.82
Service	0.68	0.70	0.69
Environment	0.69	0.62	0.65
Facilities	0.62	0.68	0.65
Dining	0.76	0.42	0.54

The performance of the BERT-BiLSTM-CRF model for RE was evaluated based on training loss and classification accuracy. As shown in Figure 8, the model's loss values decreased from 1.31 to 0.02 throughout the training process, indicating successful convergence and gradual improvement in the model's ability to predict relationships. After training, the model was evaluated on a test set consisting of 971 instances, of which 808 were correctly classified, resulting in an overall accuracy of 83.21%. This accuracy suggests that the BERT-BiLSTM-CRF model is able to effectively capture and predict the relationships between entities in the dataset, achieving a reasonable level of performance for the task.

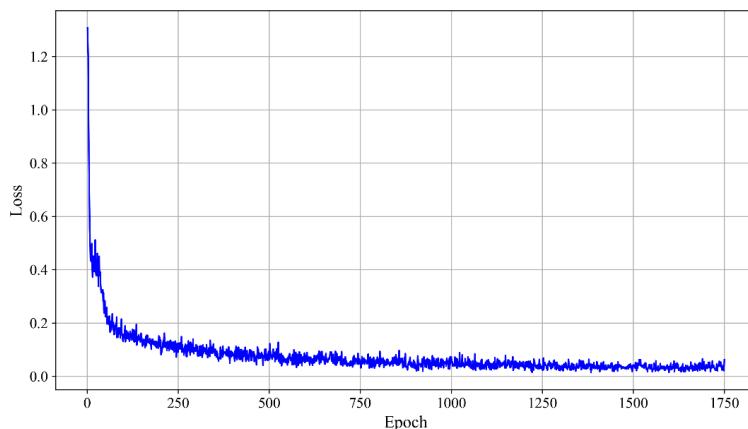


Figure 8. BERT-BiLSTM-CRF model RE training results.

4.2.2. Results for Aspect-Based Sentiment Analysis of Hotels

By utilizing the trained BERT-BiLSTM-CRF model, we conducted an aspect-based sentiment analysis on hotel review data, extracting sentiment tuples consisting of aspect terms, opinion terms, categories, and polarities, as detailed in Table 5.

Table 5. Selected results of aspect-based sentiment analysis using the BERT-BiLSTM-CRF.

Text	Category	Relation
Service is friendly	'Service': [('Service', 0, 1)], 'Positive': [('Friendly', 2, 3)]	[('Service', 'Friendly', 'Positive')]
The bed is very comfortable	'Facilities': [('Bed', 1, 2)], 'Positive': [('Very comfortable', 4, 5)]	[('Bed', 'Very comfortable', 'Positive')]
Parking is also very convenient	'Convenience': [('Parking', 0, 1)], 'Positive': [('Convenient', 4, 5)]	[('Parking', 'Convenient', 'Positive')]
The hotel's breakfast has a wide variety	'Dining': [('Breakfast', 2, 3)], 'Positive': [('Wide variety', 5, 7)]	[('Breakfast', 'Wide variety', 'Positive')]
The hotel's environment is excellent	'Environment': [('Environment', 2, 3)], 'Positive': [('Excellent', 4, 5)]	[('Environment', 'Excellent', 'Positive')]
Hygiene is clean	'Environment': [('Hygiene', 0, 1)], 'Positive': [('Clean', 2, 3)]	[('Hygiene', 'Clean', 'Positive')]
The taste is good	'Dining': [('Taste', 0, 1)], 'Positive': [('Good', 3, 4)]	[('Taste', 'Good', 'Positive')]
...

4.3. Aspect-Based Visualization

4.3.1. Overall Sentiment Visualization of Hotels

Based on the results of deep learning-based sentiment analysis, the satisfaction and importance of each category and its associated aspect terms were quantified. Satisfaction is defined as the proportion of positive reviews for a specific aspect relative to the total number of reviews for that aspect, while importance is determined by the proportion of reviews mentioning that aspect relative to the total number of reviews. The results of the satisfaction and importance analysis across various categories are presented in Figure 9.

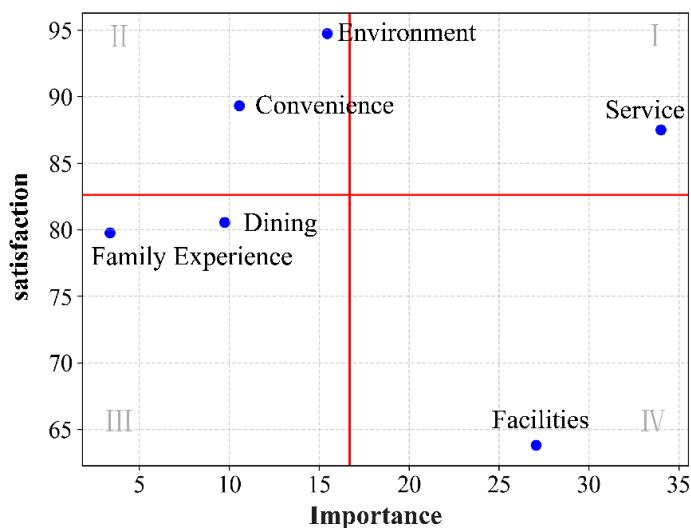


Figure 9. IPA chart of various categories for five-star hotels in Henan Province. Note: Quadrant thresholds are determined by the mean scores of satisfaction and importance (indicated by red lines). I = Maintain current efforts; II = Potential overemphasis; III = Low priority; IV = Focus attention here.

The service category received the highest satisfaction score of 87.5, indicating that respondents were generally satisfied with the quality of service. Additionally, it also exhibited the highest score of importance, suggesting that service quality is a key factor influencing overall satisfaction.

In comparison, the facilities and environmental categories showed varying levels of satisfaction and importance. The facilities category received a moderate satisfaction score of 63.82, with a relatively low importance score of 27.06. This implies that while facilities are important to satisfaction, they are less critical compared to service in shaping overall satisfaction.

The environment category, however, achieved the highest satisfaction score of 94.75, but had a relatively low importance score of 15.45. This indicates that while respondents were highly satisfied with the environment, they did not perceive it as a major driver of their overall experience. The convenience category received a satisfaction score of 89.32 and an importance score of 10.55, indicating that while convenience is a valued aspect of the experience, it plays a less significant role compared to service or facilities.

Similarly, the dining category, with a satisfaction score of 80.56 and an importance score of 9.74, showed that while food and beverage quality is important, it is less important than service and facilities. Lastly, the family experience category received the lowest satisfaction score of 79.78, with an even lower importance score of 3.36. This indicates that while family-related services are somewhat valued, they have a minimal impact on the overall experience.

These results suggest that while service and environment are the most critical categories influencing satisfaction, the importance of facilities, convenience, and dining should not be overlooked. Future improvements in service quality and facility offerings, with a particular focus on enhancing the overall environment, may further enhance customer satisfaction.

4.3.2. Fine-Grained Sentiment Visualization of Hotels

Using the BERT-BiLSTM-CRF model, the aspect-based sentiment analysis identified a total of 826 aspect terms. Aspect terms with a frequency greater than 100 were presented in Figure 10.

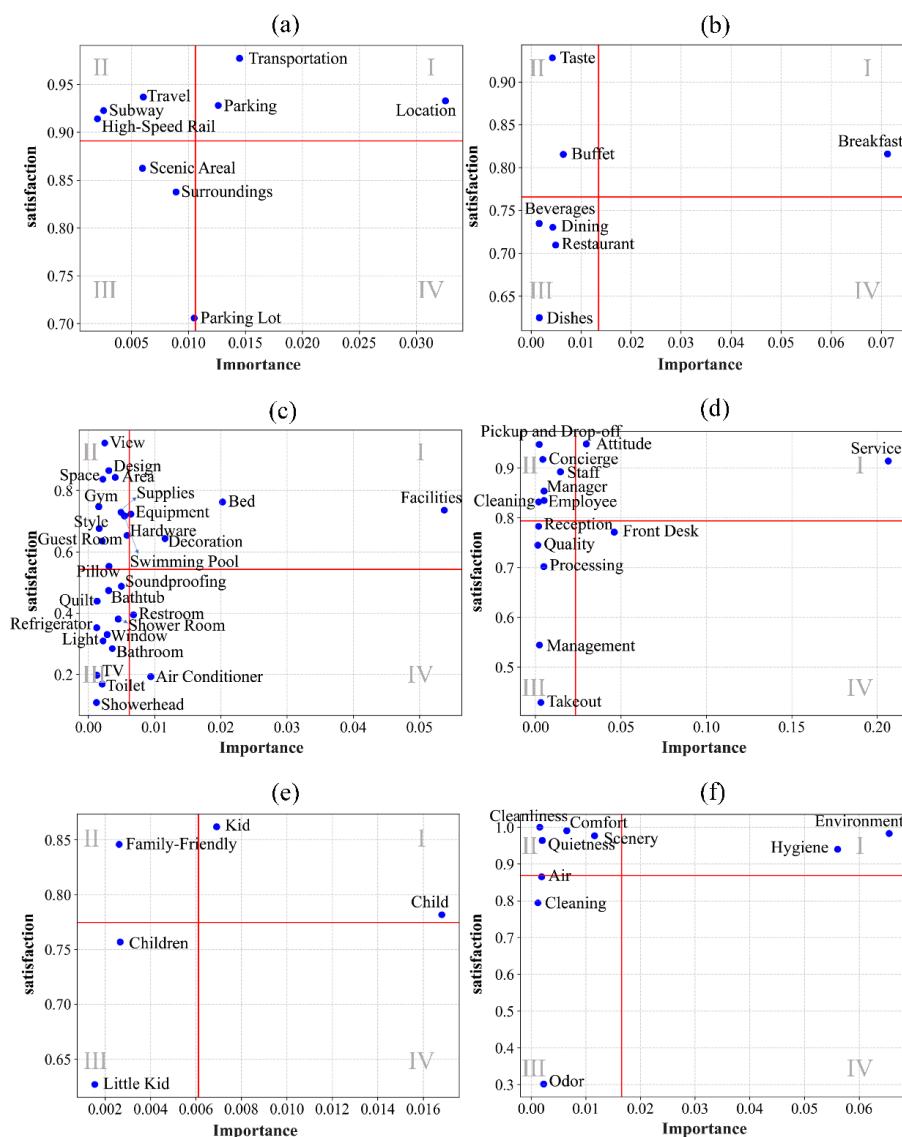


Figure 10. IPA chart of aspect terms across six categories for five-star hotels in Henan Province. Note: The figure presents IPA charts for six categories: (a) convenience, (b) dining, (c) facilities, (d) service, (e) family experiences, and (f) environment. Quadrant thresholds are determined by the mean scores of satisfaction and importance (indicated by red lines). I = maintain current efforts; II = potential overemphasis; III = low priority; IV = focus attention here.

In categories of convenience, customers generally expressed high satisfaction with the convenience of transportation. The aspect terms “transportation” and “location” received particularly high scores. However, “parking lot” had a lower score, reflecting the need for improvements in parking-related facilities.

In categories of dining, “taste” received a relatively high score, especially for breakfast and buffet offerings. However, the reviews for “dishes” and “restaurant” were lower, indicating notable variability in the overall dining experience.

In categories of facilities, customers were generally satisfied with the “view” and “design”. However, dissatisfaction was evident with basic infrastructure aspects such as “air conditioning” and “showerhead”, signaling an urgent need for improvement in these areas.

In categories of service, customers expressed high satisfaction with “attitude” and “pickup and drop-off”. However, areas such as “takeout” and “management” emerged as notable weaknesses in the overall service experience.

In categories of family experiences, “children’s activity facilities” received high recognition, but unmet customer expectations remain regarding the comprehensiveness and detailed design of children-related services.

In categories of environment, “cleanliness” and “comfort” received exceptionally high satisfaction scores, demonstrating the hotel’s strong performance in creating a pleasant environment. However, “odor” was a significant issue, highlighting an area that requires immediate attention from management.

The aspect-based sentiment analysis reveals that Henan’s five-star hotels generally perform well in key areas such as transportation convenience, service attitude, cleanliness, and location, with high customer satisfaction in these aspects. However, there are areas needing improvement, particularly outdated infrastructure such as air conditioning and bathrooms, as well as parking facilities. While the dining experience is praised for taste, concerns about the variety and uniqueness of offerings, especially at breakfast, were noted. Family-friendly services also require enhancement, particularly in the design of children’s facilities.

Furthermore, while cleanliness and comfort receive positive feedback, odor issues remain a significant concern. As shown in Figure 11, the frequency analysis of aspect terms and opinion terms provides a clearer understanding of customer sentiment. These findings suggest that improving infrastructure, dining variety, and family experience will be essential to boost customer satisfaction. Additionally, service quality, especially during peak periods, and better management efficiency were highlighted as areas for improvement. Addressing these issues will help create a more satisfying and comprehensive customer experience.



Figure 11. Word cloud of aspect terms and opinion terms. Note: (a) positive sentiments; (b) negative sentiments.

5. Discussion

This study explored the factors influencing customer satisfaction in luxury hotels in Henan, China, through aspect-based sentiment analysis. By integrating the LDA topic model with advanced deep learning techniques, we analyzed online review data to identify specific aspects of the customer experience and their associated sentiment.

5.1. Theoretical Implications

In recent years, the rapid development of natural language processing technologies has drawn increasing attention to sentiment analysis research. Scholars emphasize the importance of aspect-level sentiment analysis in tourism studies, particularly the application of deep learning techniques in this area [43,44]. However, existing research predominantly focused on sentence-level or document-level sentiment analysis, with limited exploration of aspect-level sentiment analysis in Chinese texts within the tourism sector. The application of deep learning methods in this field is still in its infancy. This study contributes theoretically by proposing an innovative analytical framework that combines the LDA topic model with the BERT-BiLSTM-CRF model. This framework can automatically extract dimensions

of customer satisfaction from large-scale online review data and conduct fine-grained sentiment analysis, addressing methodological gaps in existing research.

Traditional customer satisfaction models, such as the Expectancy-Disconfirmation Model and the SERVQUAL model [20,21,45], often rely on structured survey data. Notably, our framework, based on aspect-level sentiment analysis, excels at analyzing unstructured data like online customer reviews, which are abundant in modern digital environments. Driven by artificial intelligence (AI), our framework operates with higher automation and requires minimal human intervention, making it well-suited to today's big data landscape. Additionally, sentiment analysis provides a more nuanced understanding by capturing not only the emotional tone of customer feedback but also the specific reasons behind their sentiments. This fine-grained approach allows for the identification of detailed dimensions of customer satisfaction and emotional responses, offering deeper insights into customer experiences.

Based on 29,334 Chinese online reviews, the findings of this study demonstrate high reliability and generalizability, offering new perspectives for scholars and advancing the application of deep learning techniques in Chinese aspect-level sentiment analysis. Additionally, through the LDA model, we identified six dimensions of customer satisfaction in luxury hotels in Henan: service, facilities, environment, dining, convenience, and family experience. The dimensions of service, facilities, environment, and dining align with a previous study [43]. Meanwhile, the newly identified dimensions of convenience and family experience highlight the unique competitive advantages of luxury hotels in Henan, enriching the theoretical framework of hotel customer satisfaction dimensions.

5.2. Practical Implications

This study offers valuable insights for hotel management practices by analyzing online review data to identify key dimensions influencing customer satisfaction. This approach can help hotel managers and investors in understanding customer needs and optimizing resource allocation. The findings revealed that service attributes remain paramount in the hotel industry, which aligned with the conclusions drawn by Guo, et al. [46]. Facilities were identified as the dimension with the lowest satisfaction, consistent with findings of Nash, et al. [47]. The majority of negative sentiments primarily stemmed from complaints about facilities, particularly noise issues in luxury hotels.

Based on aspect-level sentiment analysis, we further identified specific aspects within each dimension that significantly impact customer satisfaction. For instance, in the service dimension, front desk efficiency and staff attitude played an important role in shaping customer experiences [48]. Issues with air conditioning, bedding, and showers in the facilities dimension, as well as odors in the environment dimension, greatly affected customer satisfaction. This finding further reinforced the observation by Guo, Barnes and Jia [46] that luxury hotel guests maintain exceptionally high expectations for both infrastructure and environmental quality. Additionally, the diversity and taste of dining options emerged as key determinants of customer satisfaction. Quan and Wang [49] also emphasized the importance of food in tourism. Our study further highlighted the demand for diverse dining experiences in luxury hotels, as customers are increasingly eager to explore new culinary offerings. These granular findings provide hotel managers with more targeted directions for service improvement. Moreover, our reliance on online customer reviews as a data source ensures that our findings are grounded in authentic customer experiences, thereby mitigating the biases often associated with traditional survey methods and enhancing the overall objectivity and reliability of our conclusions. With China's hotel market accounting for nearly 60% of hotel supply in the Asia-Pacific region [50], and the rapid growth of online hotel reviews, sentiment analysis plays an increasingly

crucial role in opinion mining. Furthermore, our study indicated that family experience is becoming a significant factor influencing hotel customer satisfaction. According to Mafengwo's statistics, family travel accounted for 55% of tourism in 2024, surpassing other types of travel. This suggests that hotel managers should prioritize the needs of family travelers by providing more diverse family-friendly services and facilities. By integrating the linguistic habits of Chinese hotel consumers, our study established a solid foundation for future research using sentiment analysis to accurately interpret the emotions expressed in Chinese reviews.

5.3. Managerial Recommendations

Based on the study's findings, we propose the following managerial recommendations. First, luxury hotels should prioritize updating and upgrading their facilities, as outdated facilities were identified as a primary source of customer dissatisfaction. Routine maintenance and timely upgrades, particularly for in-room bathrooms, air conditioning, and shower equipment, are essential. Additionally, hotels should consider the application of smart facilities, such as smart toilets, curtains, and robots, to meet the demands of new-generation consumers. Enhancing the quality of dining services is also crucial, with a focus on innovation and diversity. Besides offering high-quality dishes, hotels should consider incorporating local culinary specialties to create distinctive dining experiences. Furthermore, optimizing parking services is vital, as the convenience of parking significantly impacts customer satisfaction with the rise in self-driving tours. Hotels should provide sufficient parking spaces and optimize the efficiency of parking management processes to enhance guest convenience and satisfaction. Strengthening the management of food delivery services is also important, with the potential introduction of delivery robots to facilitate customer access to deliveries and enhance the accommodation experience. Simultaneously, attention should be given to the needs of family travelers by increasing family-friendly facilities and services, such as children's play areas and parent-child activities, to attract more family guests. Improving the hotel environment, enhancing hygiene management, eliminating odors, and creating a quiet and comfortable accommodation environment are equally critical. Lastly, hotels should utilize tools like aspect-based sentiment analysis to continuously monitor customer feedback, promptly identify issues, and implement improvements. By embracing these sustainable practices, hotels can enhance customer satisfaction while contributing to a more environmentally responsible tourism sector.

5.4. Limitations and Future Research

The primary limitations of this study arise from its reliance on a single data source and the restricted time frame of the analysis. As our research is based exclusively on online review data, it may be subject to sample selection bias, potentially overlooking key aspects of customers' offline experiences. Additionally, the study does not consider potential differences across regions, hotel tiers, or brands, which limits the generalizability of the findings. Future research could address these constraints by incorporating additional data sources, examining a broader range of variables, and testing the applicability of these conclusions across more diverse settings. Such efforts would yield more comprehensive insights and offer stronger guidance for sustainable strategic development within the industry.

6. Conclusions

This paper proposed a novel research framework combining the LDA model and the BERT-BiLSTM-CRF model for conducting topic mining and aspect-based sentiment in online hotel reviews. It was applied to the case study of Henan Province, China, with a focus on reviews from 2022 to 2024. The LDA model was used to identify key customer

concerns, while the BERT-BiLSTM-CRF model was used to quantify customer sentiment across multiple dimensions. As a result, this study provides actionable insights into customer needs and improvement areas for the hotel sector, offering guidance for the sustainable development of the industry.

The LDA model identified six critical dimensions shaping customer evaluations of five-star hotels in Henan Province: service, facilities, environment, dining, convenience, and family experience. Notably, “family experience” emerged as a significant new theme, reflecting the growing demand for family-oriented services and presenting a unique opportunity to integrate sustainable practices.

The BERT-BiLSTM-CRF model revealed aspect-based sentiment insights, showing that customers highly valued the environment and convenience, particularly cleanliness and hotel accessibility. However, dining and facilities were identified as key areas necessitating substantial enhancement, with common issues including breakfast quality, poor sound-proofing, and outdated facilities such as air conditioning and bathrooms. While family experiences received some positive feedback, the findings underscore the importance of enhancing both the quality and functionality of family-oriented facilities to better meet customer expectations. These findings highlight the value of our proposed method in informing sustainable development strategies for the hotel sector.

Author Contributions: Conceptualization, T.P. and D.Y.; methodology, T.P. and D.Y.; software, T.P.; validation, T.P., D.Y., and J.L.; formal analysis, D.Y. and J.L.; investigation, D.Y.; resources, T.P. and D.Y.; data curation, T.P. and L.H.; writing—original draft preparation, T.P.; writing—review and editing, D.Y. and J.L.; visualization, T.P.; supervision, H.L. and L.H.; project administration, D.Y. and J.L.; funding acquisition, D.Y. and J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Postdoctoral Research Program of Henan Province (No. 202103033), General Program of Humanities and Social Sciences Research in Henan Universities (No. 2025-ZZJH-150), National Social Science Foundation of China under Grant (No. 20BJY200), and Intangible Cultural Heritage Research Project of Culture and Tourism Department of Henan Province (No. 24HNFY-LX225).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are not contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. De Pelsmacker, P.; Van Tilburg, S.; Holthof, C. Digital marketing strategies, online reviews and hotel performance. *Int. J. Hosp. Manag.* **2018**, *72*, 47–55. [[CrossRef](#)]
2. Chu, Q.H.; Zhang, Z.Q.; Wu, T.J.; Zhang, Z.L. Interaction between online retail platforms’ private label brand introduction and manufacturers’ channel selection. *J. Retail. Consum. Serv.* **2025**, *84*, 104208. [[CrossRef](#)]
3. Berezina, K.; Bilgihan, A.; Cobanoglu, C.; Okumus, F. Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *J. Hosp. Mark. Manag.* **2016**, *25*, 1–24. [[CrossRef](#)]
4. Zhang, D.; Niu, B. Leveraging online reviews for hotel demand forecasting: A deep learning approach. *Inf. Process. Manag.* **2024**, *61*, 103527. [[CrossRef](#)]
5. Wu, T.J.; Zhang, R.X. Exploring the impacts of intention towards human-robot collaboration on frontline hotel employees’ positive behavior: An integrative model. *Int. J. Hosp. Manag.* **2024**, *123*, 103912. [[CrossRef](#)]
6. Wang, P.; Hou, Y. The Effects of Hotel Employees’ Attitude Toward the Use of AI on Customer Orientation: The Role of Usage Attitudes and Proactive Personality. *Behav. Sci.* **2025**, *15*, 127. [[CrossRef](#)]
7. Zhang, W.; Li, X.; Deng, Y.; Bing, L.; Lam, W. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Trans. Knowl. Data Eng.* **2022**, *35*, 11019–11038. [[CrossRef](#)]
8. Hu, N.; Zhang, T.; Gao, B.; Bose, I. What do hotel customers complain about? Text analysis using structural topic model. *Tour. Manag.* **2019**, *72*, 417–426. [[CrossRef](#)]

9. Wankhade, M.; Rao, A.C.S.; Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. *Artif. Intell. Rev.* **2022**, *55*, 5731–5780. [[CrossRef](#)]
10. Tan, K.L.; Lee, C.P.; Lim, K.M. A survey of sentiment analysis: Approaches, datasets, and future research. *Appl. Sci.* **2023**, *13*, 4550. [[CrossRef](#)]
11. Alqaryouti, O.; Siyam, N.; Monem, A.A.; Shaalan, K. Aspect-based sentiment analysis using smart government review data. *Appl. Comput. Inform.* **2020**, *20*, 142–161. [[CrossRef](#)]
12. Do, H.H.; Prasad, P.W.; Maag, A.; Alsadoon, A. Deep learning for aspect-based sentiment analysis: A comparative review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [[CrossRef](#)]
13. Nazir, A.; Rao, Y.; Wu, L.; Sun, L. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Trans. Affect. Comput.* **2020**, *13*, 845–863. [[CrossRef](#)]
14. Chen, Y.; Zhang, H.; Liu, R.; Ye, Z.; Lin, J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowl. Based Syst.* **2019**, *163*, 1–13. [[CrossRef](#)]
15. Buenano-Fernandez, D.; Gonzalez, M.; Gil, D.; Luján-Mora, S. Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access* **2020**, *8*, 35318–35330. [[CrossRef](#)]
16. Meng, F.; Yang, S.; Wang, J.; Xia, L.; Liu, H. Creating knowledge graph of electric power equipment faults based on BERT-BiLSTM-CRF model. *J. Electr. Eng. Technol.* **2022**, *17*, 2507–2516. [[CrossRef](#)]
17. Khoi, N.H.; Le, A.N.H. Is coolness important to luxury hotel brand management? The linking and moderating mechanisms between coolness and customer brand engagement. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 2425–2449. [[CrossRef](#)]
18. Liu, W.; Xue, Y.; Shang, C. Spatial distribution analysis and driving factors of traditional villages in Henan province: A comprehensive approach via geospatial techniques and statistical models. *Herit. Sci.* **2023**, *11*, 185. [[CrossRef](#)]
19. Khalayleh, M.; Al-Hawary, S. The impact of digital content of marketing mix on marketing performance: An experimental study at five-star hotels in Jordan. *Int. J. Data Netw. Sci.* **2022**, *6*, 1023–1032. [[CrossRef](#)]
20. Oliver, R.L. A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Mark. Res.* **1980**, *17*, 460–469. [[CrossRef](#)]
21. Parasuraman, A.P.; Zeithaml, V.; Berry, L. SERVQUAL: A multiple—Item Scale for measuring consumer perceptions of service quality. *J. Retail.* **1988**, *64*, 12.
22. Fornell, C.; Johnson, M.D.; Anderson, E.W.; Cha, J.S.; Bryant, B.E. The American customer satisfaction index: Nature, purpose, and findings. *J. Mark.* **1996**, *60*, 7–18. [[CrossRef](#)]
23. Jain, P.K.; Pamula, R.; Srivastava, G. A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Comput. Sci. Rev.* **2021**, *41*, 100413. [[CrossRef](#)]
24. Cetin, G.; Walls, A. Understanding the Customer Experiences from the Perspective of Guests and Hotel Managers: Empirical Findings from Luxury Hotels in Istanbul, Turkey. *J. Hosp. Mark. Manag.* **2016**, *25*, 395–424. [[CrossRef](#)]
25. Padma, P.; Ahn, J. Guest satisfaction & dissatisfaction in luxury hotels: An application of big data. *Int. J. Hosp. Manag.* **2020**, *84*, 102318.
26. Ismail, T.A.T.; Zahari, M.S.M.; Hanafiah, M.H.; Balasubramanian, K. Customer Brand Personality, Dining Experience, and Satisfaction at Luxury Hotel Restaurants. *J. Tour. Serv.* **2022**, *13*, 26–42. [[CrossRef](#)]
27. Shin, H.H.; Jeong, M. Redefining luxury service with technology implementation: The impact of technology on guest satisfaction and loyalty in a luxury hotel. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 1491–1514. [[CrossRef](#)]
28. Chan, N.L.; Guillet, B.D. Investigation of Social Media Marketing: How Does the Hotel Industry in Hong Kong Perform in Marketing on Social Media Websites? *J. Travel Tour. Mark.* **2011**, *28*, 345–368. [[CrossRef](#)]
29. Giglio, S.; Pantano, E.; Bilotta, E.; Melewar, T.C. Branding luxury hotels: Evidence from the analysis of consumers' "big" visual data on TripAdvisor. *J. Bus. Res.* **2020**, *119*, 495–501. [[CrossRef](#)]
30. Pantano, E.; Giglio, S.; Dennis, C. Making sense of consumers' tweets Sentiment outcomes for fast fashion retailers through Big Data analytics. *Int. J. Retail Distrib. Manag.* **2019**, *47*, 915–927. [[CrossRef](#)]
31. Thakur, R. Customer engagement and online reviews. *J. Retail. Consum. Serv.* **2018**, *41*, 48–59. [[CrossRef](#)]
32. Zhao, X.; Wang, L.; Guo, X.; Law, R. The influence of online reviews to online hotel booking intentions. *Int. J. Contemp. Hosp. Manag.* **2015**, *27*, 1343–1364. [[CrossRef](#)]
33. Chan, S.W.; Chong, M.W. Sentiment analysis in financial texts. *Decis. Support Syst.* **2017**, *94*, 53–64. [[CrossRef](#)]
34. Liu, H.; Chatterjee, I.; Zhou, M.; Lu, X.S.; Abusorrah, A. Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 1358–1375. [[CrossRef](#)]
35. Bello, A.; Ng, S.-C.; Leung, M.F. A BERT framework to sentiment analysis of tweets. *Sensors* **2023**, *23*, 506. [[CrossRef](#)]
36. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
37. Osmani, A.; Mohasefi, J.B.; Gharehchopogh, F.S. Enriched Latent Dirichlet Allocation for Sentiment Analysis. *Expert Syst.* **2020**, *37*, e12527. [[CrossRef](#)]

38. Liu, J.; Hu, S.K.; Mehraliyev, F.; Liu, H.L. Text classification in tourism and hospitality—A deep learning perspective. *Int. J. Contemp. Hosp. Manag.* **2023**, *35*, 4177–4190. [[CrossRef](#)]
39. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K.; Assoc Computat, L. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the North-American-Chapter of the Association-for-Computational-Linguistics—Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
40. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
41. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)]
42. Srivastav, A.; Singh, S. Proposed Model for Context Topic Identification of English and Hindi News Article Through LDA Approach with NLP Technique. *J. Inst. Eng. India B Electr. Electron. Telecommun. Comput. Eng.* **2022**, *103*, 591–597. [[CrossRef](#)]
43. Luo, J.Q.; Huang, S.S.; Wang, R.W. A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *J. Hosp. Mark. Manag.* **2021**, *30*, 71–95. [[CrossRef](#)]
44. Ameur, A.; Hamdi, S.; Ben Yahia, S. Sentiment Analysis for Hotel Reviews: A Systematic Literature Review. *ACM Comput. Surv.* **2024**, *56*, 1–38. [[CrossRef](#)]
45. Chauhan, G.S.; Nahta, R.; Meena, Y.K.; Gopalani, D. Aspect based sentiment analysis using deep learning approaches: A survey. *Comput. Sci. Rev.* **2023**, *49*, 100576. [[CrossRef](#)]
46. Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [[CrossRef](#)]
47. Nash, R.; Thyne, M.; Davies, S. An investigation into customer satisfaction levels in the budget accommodation sector in Scotland: A case study of backpacker tourists and the Scottish Youth Hostels Association. *Tour. Manag.* **2006**, *27*, 525–532. [[CrossRef](#)]
48. Kandampully, J.; Zhang, T.T.; Jaakkola, E. Customer experience management in hospitality A literature synthesis, new understanding and research agenda. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 21–56. [[CrossRef](#)]
49. Quan, S.; Wang, N. Towards a structural model of the tourist experience: An illustration from food experiences in tourism. *Tour. Manag.* **2004**, *25*, 297–305. [[CrossRef](#)]
50. Yang, Z.S.; Cai, J.M. Do regional factors matter? Determinants of hotel industry performance in China. *Tour. Manag.* **2016**, *52*, 242–253. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.