Data Science and Management 7 (2024) 228–237

Contents lists available at ScienceDirect

Data Science and Management

journal homepage: www.keaipublishing.com/en/journals/data-science-and-management

Research article

Survival strategies for family-run homestays: analyzing user reviews through text mining

Jay Krishnan [a], Biplab Bhattacharjee [b], Maheshwar Pratap [a], Janardan Krishna Yadav [b], Moinak Maiti [c,*]

[a] Amrita School of Business, Amrita Vishwa Vidyapeetham, Coimbatore, 641112, India
[b] Jindal Global Business School, O.P. Jindal Global University, Sonipat, 131001, India
[c] Department of Finance, School of Economics and Finance, University of the Witwatersrand, Johannesburg, 2050, South Africa

ARTICLE INFO

Keywords:
Text mining
Tourism
Homestays
Topic modeling
Predictive modeling

ABSTRACT

Online booking of homestays through e-travel portals is based on the virtual brand and perception, which are largely affected by user-generated electronic word-of-mouth (eWOM). With the objective of mining actionable insights from eWOM, this study conducted opinion mining for homestays located in four thematic areas of Kerala. Accordingly, various techniques have been deployed, such as sentiment and emotional analyses, topic modeling, and clustering methods. Key themes revealed from topic modeling were breakfast, facilities provided, ambience, bathroom, cleanliness, hospitality exhibited, and satisfaction with the host. A lasso logistic regression-based predictive binary text classification model (with 97.6% accuracy) for homestay recommendations was developed. Our findings and predictive model have implications for managers and homestay owners in devising appropriate marketing strategies and improving their overall guest experience.

1. Introduction

Homestays are gaining ground as alternative accommodation options for tourists, providing them with a more localized experiences (Kulshreshtha and Kulshrestha, 2019). Homestays are often run by homeowners, families, or small-time entrepreneurs. With e-commerce gaining popularity, online booking platforms and travel websites are rapidly becoming direct interfaces between travelers and hospitality service providers for all accommodation categories (Agag and El-Masry, 2017; Llach et al., 2013). As per HolidayIQ report, in Indian top-50 emerging tourism destinations, 13% are homestays among the total booked accommodations. The same report stated that 17% of homestays are in South India and 72% are in North India (Kulshreshtha and Kulshrestha, 2019). According to Statistica.com 2023 reports, 67% of visitors could book homestays through third-party online booking portals during the post-COVID period. Apart from e-booking on travel-booking portals, travelers also share their experiences with property or services via ratings and reviews to guide other travelers on their travel decisions (Hwang et al., 2018).

Visitors often share their travel experiences with a particular place and property on social media (e.g., YouTube channels, Instagram, Facebook), travel review sites (e.g., TripAdvisor, Google reviews), online booking sites, and travel blogs. This information is known as electronic word-of-mouth (eWOM) (Hwang et al., 2018; Racherla et al., 2013). eWOM are the "voice of the customer" that offer actionable insights into their likes and dislikes about a service, and sometimes, even recommendations for improvement. To generate leads and strengthen their brands in online booking spaces, homestay owners must understand customer expectations. Survey-based studies or analyses of the eWOM posted on online booking portals, travel blogs, and social media portals can reveal interesting insights for homestay owners to align their offerings with customer needs (Molina-Azorín et al., 2015).

Previous consumer behavior research has leveraged eWOM as data to unearth diverse perspectives. Research has focused on the helpfulness of reviews (Magnini et al., 2011), their impact on customer satisfaction, sales, and e-booking (Yang et al., 2011). Previous studies have performed opinion mining on eWOM affiliated with diverse sectors of the hospitality industry such as theme parks (Zhang et al., 2021), lodges, restaurants

Peer review under responsibility of Xi'an Jiaotong University.
* Corresponding author.
E-mail addresses: maitisoft@gmail.com, moinak.maiti@wits.ac.za (M. Maiti).

https://doi.org/10.1016/j.dsm.2024.03.003
Received 8 October 2023; Received in revised form 4 March 2024; Accepted 7 March 2024
Available online 8 March 2024

(Gan et al., 2016; Jia, 2018.), hotels (He et al., 2017), and cruises (Buzova et al., 2018), among others. A considerable body of work has been dedicated to analyzing eWOM in the context of hotels and peer-to-peer (P2P) shared accommodation booking platforms such as Airbnb (Bao et al., 2022; Brochado et al., 2017; Cheng and Jin, 2019; Li et al., 2019; Magnini et al., 2011; Sthapit and Jiménez-Barreto, 2018; Yu et al., 2020). However, literature on family-managed homestays is scarce. This constitutes a major research gap.

Homestays are emerging as popular alternative accommodation options owing to the growing P2P economy and the post-pandemic push for authentic homestays. The popularity of a homestay as an alternative accommodation option depends on several factors provided by owners, such as personalized services, a homely atmosphere, home-cooked food, cultural immersion, and authentic local experiences (Agyeiwaah, 2013; Wang, 2007). Although their authentic cultural experience is crucial to homestay bookings (Wang, 2007), price is also an important factor (Hsu and Lin, 2011). Other factors also play a role, such as quiet local neighborhoods (Tussyadiah and Pesonen, 2016), the pollution level of the selected destination (Agyeiwaah, 2013), relaxation and leisure activities (Hsu and Lin, 2011), and scenic surroundings.

In the context of e-bookings for homestays, the individual and overall average rankings of a particular homestay are of particular interest to people while booking. These two rankings provide quantitative information on the quality and performance of homestays in terms of their amenities and services (Yong and Hassan, 2019). Although numerical ranking scores are crucial in booking decisions, detailed reviews provide qualitative dimensions of services and amenities provided by an accommodation (Litvin and Sobel, 2019). In the absence of an actual view of the property in question, these reviews significantly influence prospective customers' booking decisions (Litvin and Sobel, 2019). Therefore, a deeper analysis of online customer reviews can provide actionable insights for homestay owners to continuously improve their service offerings and attract new customers for sustainable growth.

A vast majority of studies on the factors guiding homestay bookings have collected data based on surveys. Moreover, a few eWOM-based studies have been reported based on Airbnb surveys (Bao et al., 2022; Brochado et al., 2017; Cheng and Jin, 2019; Li et al., 2019; Lyu et al., 2019; Sthapit and Jimenez-Barreto, 2018; Yu et al., 2020). Airbnb, an alternative accommodation booking platform, enables visitors to book private homestays with local families. Guests can book a room or the entire property depending on whether the homestay offers an entire apartment, cottage, hostel, or hotel. Previous studies have considered all properties listed on the platform, and none are specific to the family homestay category. Therefore, these studies do not specifically capture the essence of family homestays. From the extant literature, it can be inferred that a substantial volume of work has been reported on eWOM analysis of data sourced from lodges, restaurants, hotels, tourist destinations, and P2P shared accommodation-booking platforms. Though, from a perspective of a "pure authentic homestay", wherein guests stay with the family of the host and gets a real homelike experience, eWOM studies are scarce. This constitutes a major research gap. Furthermore, eWOM studies on homestays in the Indian context are rare in existing literature.

As an alternative accommodation option to hotels, guesthouses, and resorts, family-managed homestays offer a very different and, in most cases, authentic local experience. Thus, data mining of eWOM specific to homestays offers insights that are different from those of professionally managed hotels. Most studies on consumer behavior regarding homestays have adopted survey-based methods and ignored opinion mining of eWOM data. Furthermore, while studies have considered Twitter, TripAdvisor, Yelp, Booking.com, and other travel websites, e-booking portals in the Indian context have largely been ignored, with the exception of Thapa et al.'s (2018) study. Therefore, our study differs in terms of advanced modeling and comprehensive analysis.

Text analytics has been widely applied to online tourism reviews using various analytical methods, such as sentiment analysis (Chang et al., 2019), trend analysis (Chang et al., 2019), text mining (Boo and Busser, 2018), topic modeling and cluster analysis (Bassolas et al., 2016; Boo and Busser, 2018), and predictive analytics (Geetha et al., 2017; Ghose et al., 2012). Opinion mining of eWOM using text analysis methods can offer several useful actionable insights for homestay owners, such as service attributes that hold relevance for tourists and their order of priority (Molina-Azorín et al., 2015). Such insights can aid hospitality service providers in aligning their offerings with customer demand, building virtual brands, and improving their service quality measures.

This study focuses on opinion mining of the eWOM of homestay accommodations in four distinct locations: Kochi, Alappuzha, Trivandrum, and Wayanad in the Southern Indian state of Kerala. Kerala is a renowned tourist destination and attracts a wide range of travelers, including travel groups, foreign tourists, honeymooners, couples, sole travelers, and families (Thimm, 2017). This study employed a text analytical workflow consisting of data preprocessing, sentiment analysis, emotional analysis, text clustering, and topic modeling. Additionally, text classification of homestay recommendations was also attempted. Particularly, this study seeks to identify the frequently occurring themes mentioned in online reviews regarding family-managed homestays, the negative qualities associated with homestays, the emotional connection evoked in these experiential stays at the selected destinations, whether the aspects highlighted in these online reviews for a select destination are similar in nature, whether perceptions about a destination influence homestay experiences, and whether one model can predict the relationship between eWOM attributes and future book-recommendation decisions.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the literature, followed by Section 3 on the methodology. Section 4 elaborates on the results. Finally, we conclude the paper with policy implications.

2. Literature review

2.1. Opinion mining and hospitality sector

Opinion mining has been used extensively in qualitative inquiries in the context of tourism. For example, it has been employed to understand how the quality of online reviews affects consumer attitudes and choice behaviors when booking a property (Filieri and Mcleay, 2014; Schuckert et al., 2015). Extant literature is of the view that online reviews (big data) provide deep insights into customer needs and are, therefore, of great use in helping owners implement suitable promotional strategies (Özer et al., 2022). A large volume of user-generated data is easily available in the form of online reviews. However, mining them to extract insights remains a challenge. This is primarily because the data are unstructured, making it difficult to capture, analyze, interpret, and derive actionable insights (Han et al., 2016; He et al., 2017). Past studies have performed such exercises on eWOM derived from diverse sectors of the hospitality industry such as theme parks (Zhang et al., 2021), lodges, restaurants (Guo et al., 2017; Jia, 2018), hotels (He et al., 2017), and cruises (Buzova et al., 2018), among others. These studies have employed different types of datasets and techniques for opinion mining.

2.2. Text analytics in opinion mining

Text analytics have been widely applied to online reviews on tourism using various analytical methods, such as sentiment analysis (Chang et al., 2019), trend analysis (Chang et al., 2019; Li et al., 2015a), text mining (Boo and Busser, 2018), topic modeling and cluster analysis (Bassolas et al., 2016; Boo and Busser, 2018), and predictive analytics (Geetha et al., 2017; Ghose et al., 2012). The focus of these works has been on the "eWOM helpfulness" to prospective customers, competitive analysis of different hotels and destinations, and customer satisfaction. Earlier studies have mostly been conducted in the context of hotels and hospitality services, followed by tourist destinations and food, restaurants, and beverage sectors. Most studies have sourced user-generated data from popular sites such as TripAdvisor, Yelp, Twitter, and Booking.com. Some were sourced from content communities and discussion forums.

2.3. Sentiment analysis and eWOM

eWOM comprises unstructured data, which may be in textual, image, or video form (Bhattacharjee et al., 2017). Systematic extraction and mining of this information can provide a wealth of insights for practitioners in the hospitality sector. Sentiment analysis is a way to gain insight into textual data. In text-analytics paradigms, sentiments refer to judgments, attitudes, or thoughts stimulated by personal experiences with a product or service (Fang and Zhan, 2015). Sentiment analysis is a machine-driven automated workflow that categorizes data based on their polarity. The workflow includes several tasks, the most prominent of which are sentiment classification and opinion extraction.

2.4. Topic modeling of online reviews

Topic modeling falls under the category of probabilistic methods. Topic modeling utilizes statistical and machine learning techniques to uncover hidden semantic structures and abstract patterns (Blei et al., 2003). In the tourism and hospitality sectors, this method has been used to analyze the satisfaction level of hotel visitors (Guo et al., 2017), conduct a comparative examination of online review platforms (Xiang et al., 2017), and study consumers' perceptions of services and products offered by various hotels (Xu et al., 2017).

2.5. Predictive analytics and eWOM

Predictive analytics apply a range of machine learning and regression models (Gandomi and Haider, 2015). This method has been applied in the hospitality sector in several use cases, such as demand estimation in hotels using user-generated content derived from several social media platforms (Ghose et al., 2012), predicting hotel customer ratings due to customer sentiment polarity changes (Geetha et al., 2017), exploring user reviews and rating prediction accuracies (Rossetti et al., 2016), and creating decision support systems for identifying satisfaction in restaurants (Zhang et al., 2017).

In the extant literature on the tourism domain, unsupervised and supervised machine learning methods have also found extensive applications. The following methods are found in the extant literature: decision tree, support vector machine, K-nearest neighbors, Naïve Bayes, gradient boosting and neural network, random forest, and logistic regression (Khorsand et al., 2020; O'Mahony and Smyth, 2010). Text clustering has been very popular in tourism literature under the umbrella of unsupervised machine learning methods. Some use cases of text clustering include analyzing eWOM data from groups of visitors at different tourist sites (Bassolas et al., 2016), identifying tourism hotspots (Li et al., 2015b), analyzing online reviews of both budget and premium accommodations (Geetha et al., 2017), and detecting emergencies (Ding et al., 2016). All the aforementioned studies deployed hierarchical, spatial clustering, and co-occurrence methods as applicable.

3. Data and methodology

3.1. Data

Secondary data were sourced from multiple Indian online travel booking portals to derive actionable insights that would be useful for homestay owners to streamline their services and grow their businesses. Data were sourced from four tourist locations in Kerala: Alappuzha (Central Kerala), Trivandrum (Southern Kerala), Wayanad (Northern Kerala), and Kochi (Central Kerala). Each of these destinations provides unique experiences based on its location and tourism avenues. Homestays were popular among tourists in all four cities. Therefore, they formed the ideal setting for this study.

Online reviews and ratings of family-run homestays in Kochi, Alappuzha, Wayanad, and Trivandrum were collected from travel booking sites (e.g., Booking.com, Goibibo.com, Makemytrip.com, and Tripadvisor.com). Only verified reviews of registered users who had booked homestays in these cities were considered for the study to ensure that the experiences and feedback were genuine, which also helped in avoiding negative feedback posted deliberately by competition. For data collection, Python 3.4 and its associated web crawling models were employed. Selenium module in Python served as the web driver and Beautiful Soup was used to extract web data from HTML. Downloading of pre-existing data does not qualify as research on human participants and so ethical clearance for the study was not required. Using the web-crawling method the three features of the reviews were extracted, namely, heading, text, and ratings. This study collected 5,056 textual reviews of homestays listed in the four destinations. Fig. 1 depicts the analytical workflow of the study. The pre-processing of data involved removing English stop words and punctuations, converting text into lower case, eliminating numbers, and word stemming. This pre-processed corpus was then subject to further analysis. In the exploratory phase of analysis, term frequency is examined, and world cloud visualization is performed.

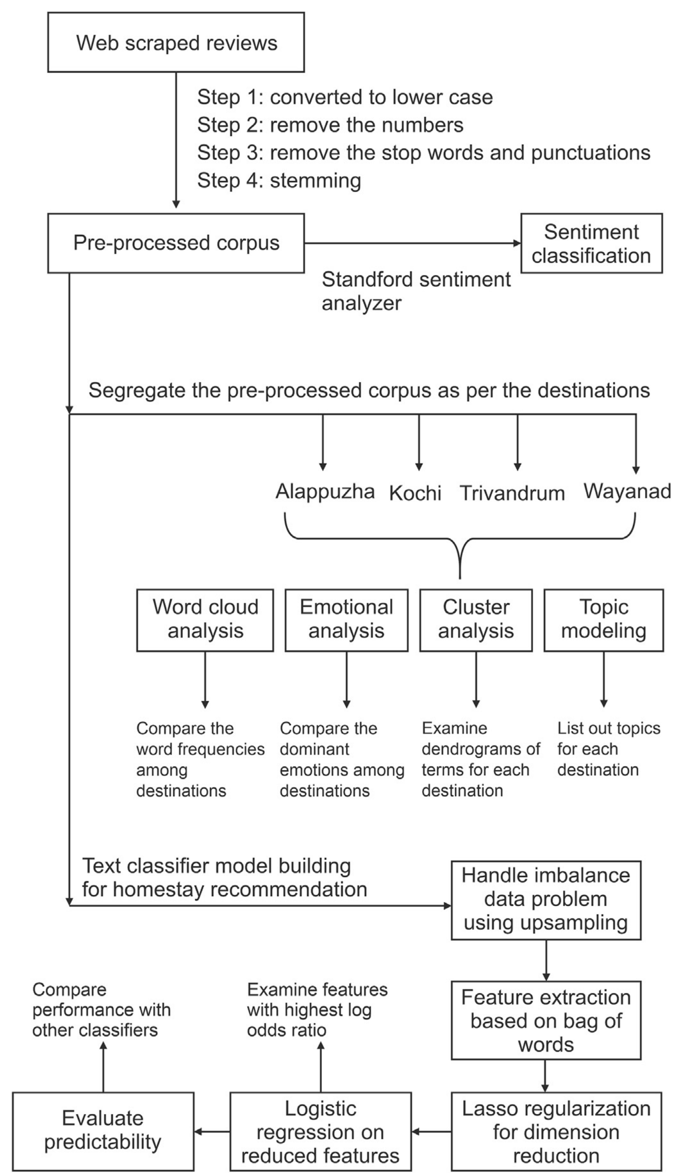

Fig. 1. Analytical workflow of the study.

3.2. Methodology

3.2.1. Cluster analysis

This is performed on the data to investigate inter-term correlations and how frequently terms jointly occur. In this study, to build hierarchical clusters, we employed the Euclidean distance matrix derived from interterm correlations. A complete linkage-based hierarchical cluster method was applied (James et al., 2013). Furthermore, these hierarchical clusters were visualized, and a detailed examination of subgroup formation was conducted.

3.2.2. Sentiment and emotional analysis

A sentiment can be a view, opinion, thought, or attitude rooted in emotions. eWOM sentiment is the level of satisfaction expressed by online reviewers and can be positive, negative, or neutral (Aakash and Gupta Aggarwal, 2020). In hotels and homestays, positive sentiments reflect customer satisfaction with a product or service and the associated sentiments are enthusiastic, happy, or excited. Negative sentiments represent dissatisfaction with the service, and the associated sentiments include annoyance and frustration. Neutral sentiments convey satisfaction but with no expressed sentiments. In this study, sentiment scoring was achieved using a machine learning approach in which supervised classification models trained on a previously annotated corpus of text were used (Ortigosa et al., 2014).

Furthermore, the Stanford sentiment analyzer was deployed to compute each review's sentiment score. This analyzer uses the Stanford Sentiment Treebank to compute sentiment scores (Socher et al., 2013). For multi-label classification, the analyses used a recursive neural tensor network, which is a particular type of deep learning model. Using this approach, five category classifications were performed: 5 = very positive; 4 = positive; 3 = neutral; 2 = negative; and 1 = very negative. Accommodation reviews aptly capture the emotional connect they achieved regarding a property. According to Shaw (2007), emotions are antecedents that affect customer experience in terms of satisfaction, retention, and loyalty. Therefore, homestays must focus on the emotional aspects of their services to attract new clients and retain existing ones (Claeys and Roozen, 2012). Emotionally satisfied customers are more likely to exhibit repeat purchase tendencies, display less price sensitivity, and share positive WOM (Zorfas and Leemon, 2016). This study employs the syuzhet package in "R" to conduct an emotional analysis of the reviews of each destination under study (Jockers, 2015). This method extracts emotions from each review and categorizes them into eight groups: surprise, joy, anticipation, sadness, anger, fear, disgust, and trust.

3.2.3. Topic modeling

This study conducted a comparative analysis of Latent Dirichlet Allocation (LDA) vs. Latent Semantic Analysis (LSA) based on the UMass and CV scores on our dataset, taking the topic numbers as 10, 15, and 20. Topics with higher CV and lower UMass scores were considered superior in any relative comparison. In terms of these scores, LDA performed better (higher CV and lower UMass scores) than LSA for all four datasets (Table 1). Therefore, we employed LDA to conduct topic modeling. The LDA model can represent a given corpus in generative probabilistic form (Tong and Zhang, 2016). The model automatically selects topics from a collection of unlabeled documents, implying that learning is unsupervised. The selected topics may not be in a highly interpretable form; therefore, coherence metrics are recommended to measure the semantic similarity of topic words and distinguish the good from the bad. The greater the value of the average pairwise similarity among words, the more coherent the topic (Mimno et al., 2011). LDA encompasses two coherence measurements based on how humans judge topic quality (Stevens et al., 2012), the UMass measures (Blei and John, 2009) and UCI (Newman et al., 2010). This study employed a UCI-based coherence measure to assess the outcomes of topic models. An LDA was used in this study. Topic modeling was performed for each of the four sets of textual reviews.

Table 1
Comparative analysis of LDA vs. LSA based on UMass and coherence value (CV) scores.

Number of topics	Coherence UMass score (LDA)	Coherence Umass score (LSA)	Coherence CV score (LDA)	Coherence CV score (LSA)
Alappuzha				
10	−3.380	−2.594	0.301	0.293
15	−3.541	−2.319	0.285	0.288
20	−3.593	−2.477	0.299	0.254
Kochi				
10	−5.796	−4.883	0.275	0.247
15	−5.489	−4.278	0.261	0.256
20	−5.13505	−4.118	0.237	0.222
Trivandrum				
10	−3.943	−3.617	0.273	0.268
15	−3.962	−3.096	0.273	0.252
20	−3.965	−3.799	0.273	0.231
Wayanad				
10	−1.044	−1.028	0.403	0.353
15	−1.032	−1.019	0.403	0.350
20	−1.026	−1.012	0.403	0.391

Note: LDA: latent dirichlet allocation; LSA: latent semantic analysis.

3.2.4. Predictive modeling

In addition to sentiment, emotional, and clustering analyses and topic modeling, predictive modeling was performed using a text-based supervised classification model to predict consumer behavior when recommending a homestay for future e-bookings. Manual labeling of the entire review data is conducted, and the reviews are categorized into two classes: (1) "recommended for future bookings" and (2) "not recommended for future bookings". This manual labeling helps create a binary classification model to recommend homestays based on feature selection. This provides a set of aspects that predict whether a visitor will recommend a homestay. Ridge-based logistic regression was employed to benchmark the lasso logistic regression model. The text classifier is represented by the following function (see Eq. (1)):

$$y = f(x) \tag{1}$$

A training set D was used for the model learning task, as shown in Eq. (2).

$$D = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)\} \tag{2}$$

In the case of the text classification exercise, one can represent the vectors using the mathematical formulation shown in Eq. (3):

$$x_i = \left[x_{i,1}, \ldots . x_{i,j}, \ldots \ldots . . x_{i,d}\right]^T \tag{3}$$

The vectors comprise transformed term frequencies. The notations $y_i \in \{1, 0\}$ are representative of the class labels and they encode the two binary classes here, i.e., recommended for booking (1) or not recommended for booking (0).

In text classifier model building, many features are required that have an adverse impact on models, such as logistic regression. This study employed lasso, a regularization technique, to reduce the number of features to be fed into the model (Hastie et al., 2017). Extant studies have shown that in text-classification exercises, when lasso-based logistic regression is compared with ridge-based models, the former outperforms the latter (Genkin et al., 2005). To perform the minimization task using the lasso regularization method, the following equation is used:

$$l_{lasso}(\beta) = l(\beta) + \lambda \sum_{j=1}^{d} |\beta_j| \tag{4}$$

In Eq. (4), the hyperparameter controlling the degree of regularization is represented by the term λ. The post-training model was validated using 10-fold cross-validation. The performance of the model was compared

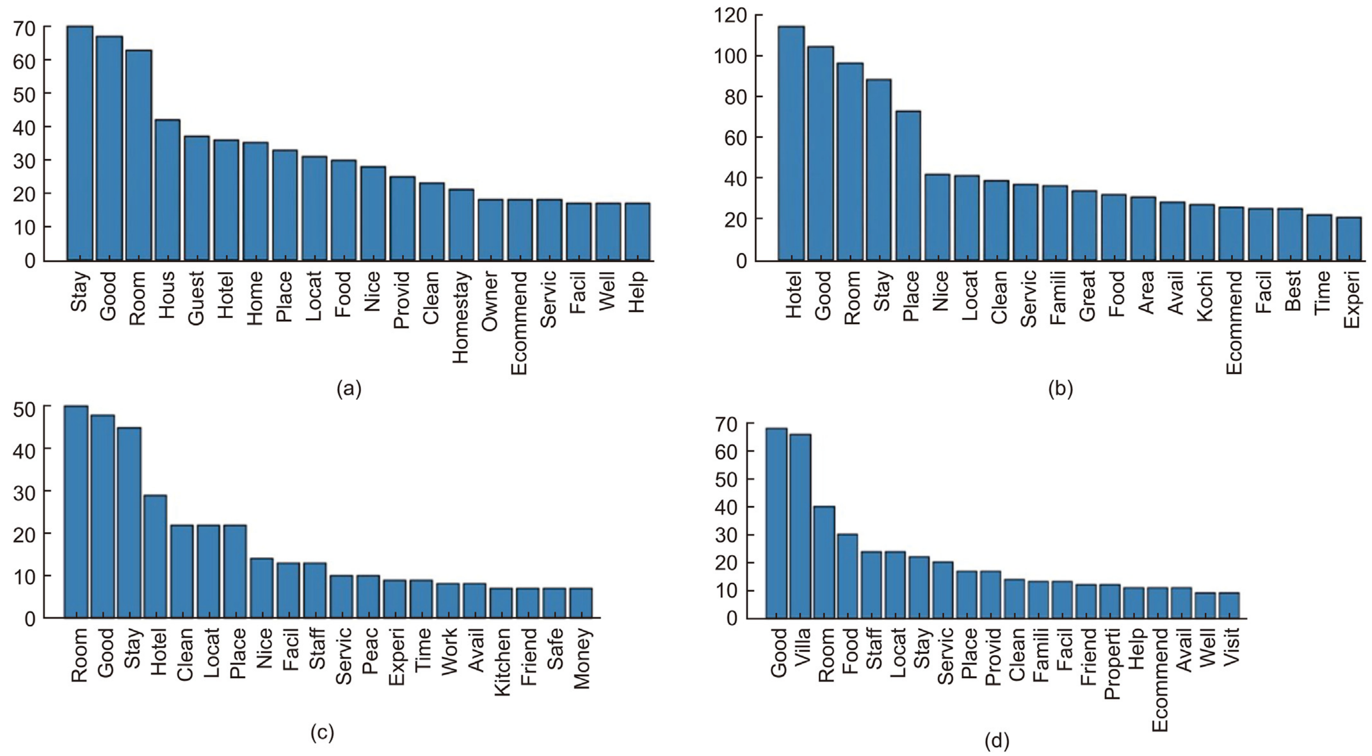

Fig. 2. Top twenty terms for homestays in Alappuzha (a), Kochi (b), Trivandrum (c), and Wayanad (d).

with three other benchmark classification models used in tourism research. These are the decision tree, ridge logistic regression, and Naïve Bayes. For comparison, four evaluation measures (accuracy, precision, recall, and F1-scores) were used. Furthermore, we examine the lasso logistic regression coefficients and compute the odds ratios.

4. Results and discussion

4.1. Exploratory analysis

This study addresses several research questions regarding the viability of homestays as an alternative accommodation option based on online customer reviews. The first question sought to identify words frequently used by users. The most commonly occurring words are "stay", "room", and "good" (Fig. 2(a)) for homestays in Alappuzha. Other words finding extensive use are "guest", "locat", "hotel", "home", and "place". Fig. 3(a) shows some other prominent words, such as "houseboat", "food", "hospitality", "shikara", and "location". The word "bamboo" is frequently used in bamboo huts or rooms. These words are representative of Alappuzha, a popular tourist destination known for its scenic beauty, coconut trees, house boating opportunities, lakes, and authentic local cuisine.

Fig. 2(b) shows that the terms "good" and "hotel" exhibit the highest frequency of use. This can be attributed to the fact that Kochi has more of an urban landscape, so visitors are more inclined toward homestays offering amenities like a hotel as shown in Fig. 3(b), that is, "location", "staff", "service", "value", and "clean". Fig. 2(c) shows the commonly used terms for homestays in Trivandrum. The commonly used words are "good" and "room". The terms "bathrooms" and "kitchen" indicate customer demand for cleanliness and good quality facilities (Fig. 3(c)). Wayanad (Fig. 2(d)) also includes the frequently occurring words of the three other destinations, along with additions such as "service", "facilities", "staff", "location", and "quality" (Fig. 3(d)). The use of the term "villa" is reflective of the fact that tourists prefer this accommodation option (Fig. 3(d)), as Wayanad is an extremely scenic destination.

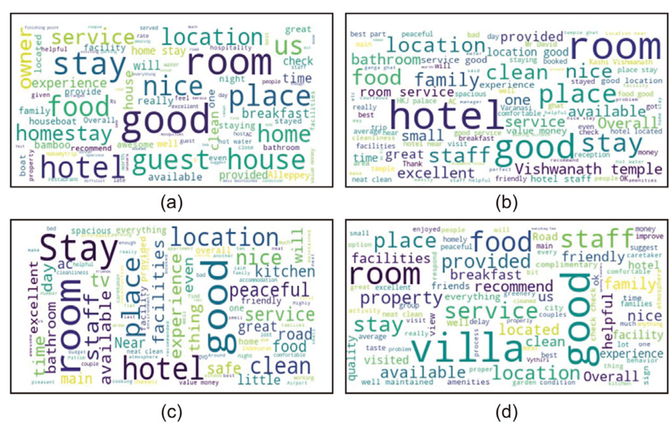

Fig. 3. Word cloud based on the reviews posted about homestays in Alappuzha (a), Kochi (b), Trivandrum (c), and Wayanad (d).

4.2. Clustering analysis

Word cloud visualizations and term frequency examinations helped us compare and contrast the eWOM belonging to the four selected destinations. Nevertheless, examining only the simple term frequency dissimilarity is not adequate for extracting sufficient actionable insights into homestay accommodations. Therefore, we employed clustering analysis to gain more granular insights from these online customer reviews. Figs. 4(a)–(d) illustrate the dendrograms for homestays at each destination. These dendrograms show an inversely proportional relationship between the cluster height and term correlations. The words in the clusters include "nice", "recommend", "good", "excellent", and "great". Additionally, the distance of the terms from the different adjectives reveals the level of satisfaction or dissatisfaction with the services offered at homestays.

For Trivandrum (Fig. 4(a)), the word “clean” is closely placed with “place”, and “nice” with “staff” and “facility”. Likewise, “good” and “stay” are also closely clustered. In the hierarchical cluster of Wayanad (Fig. 4(b)), “villa” is clustered with “good”, “room”, and “provided”, allowing for the inference that the visitors are more inclined toward booking villas in Wayanad. Words such as “visit”, “travel”, “check in”, and “checkout” are in proximity to “nice”, testifying to the flawless check-in and check-out services provided by the homestays in Wayanad. Further, “staff” and “help” were closely clustered, indicating that the visitors perceived the homestay staff in favorable light.

In case of Alappuzha (Fig. 4(c)), the terms “awesome”, “help”, and “great” are closely clustered with the term’s “family”, “recommend”, “visit”, and “experience”, indicating that the visitors had a satisfying experience in the homestays and that they found the family host helpful and welcoming. Based on their positive experiences, they recommended homestays to other prospective visitors. Moreover, the term “good” is closely clustered with the term “stays”, indicating a satisfactory experience with the property. Similarly, “service” is clustered with “bamboo”, “water”, and “breakfast”, pointing to the services considered important for reviewers staying in bamboo huts. For Kochi (Fig. 4(d)), the term “nice” appears in proximity to “area”, indicating the preference for strategically located homestays. The terms “love”, “service”, and “excel” were found to be closely clustered, indicating that visitors were satisfied with the services and had rated them highly in their reviews. We also found “recommend” to be closely associated with “host”, “staff”, “kitchen”, “hospitality”, “facilities”, and “location”, indicating that the homestay was recommended for these services and qualities.

4.3. Sentiment and emotional analysis

Negative attributes are as important as positive ones. They find reflection in terms, such as anger, disappointment, regret, or worry (Mattila and Ro, 2008). The second research question aimed to unearth the “negative aspects highlighted in the online reviews of homestays”. Negative sentiments reflect difficulty in locating homestays, not-value for money, unclean toilets and bathrooms, unreasonable price structure, ill-maintained infrastructure, mosquitoes, dirt, stink, and pollution. Positive sentiments reflected cleanliness, safety, price, location, and hospitality. Visitors provide positive reviews when they experience satisfactory and memorable homestays.

The third research question aims to understand customers’ emotions. Emotional analysis of the properties show that “trust” is the main emotional perception of the visitors irrespective of the destinations. Trust is created by positive WOM, that is, recommendations from other visitors and homestay owners providing impeccable hospitality and personal attention, and ensuring the safety and security of their guests. The trust factor increased in the following sequence: Alappuzha, Kochi, Trivandrum, and Wayanad. The next most common emotion on the list was joy. This indicates that homestay owners are welcomed by their guests, provide excellent service with a smile, are helpful, provide authentic local experiences and recommendations, serve delicious food, are scenically located, and go an extra mile to personalize their visitors’ experiences. Anticipation has emerged as a crucial emotion that varies on a case-to-case basis and is mostly neutral in nature. Our reviews did not reveal negative emotions, such as disgust, anger, or sadness. Our emotional analysis shows that anger, fear, and disgust have similar values in all four locations (Figs. 5(a)–(d)); hence, our results can be generalized for the entire state of Kerala. In summary, our results indicate that visitors are highly satisfied with the location, services, and experiences offered by homestays in these four locations in Kerala.

4.4. Topic modeling

The fourth research question seeks to identify the “topics that the homestay customer emphasizes more in the posted reviews” based on the CV score and UMass as shown in Table 2.

These results highlight several key findings. First, similar patterns can be observed in eWOM topics belonging to different destinations, implying that cleanliness and satisfaction with homestay amenities, hosts, and staff are crucial aspects of services, irrespective of the destination. Second, there are qualities unique to each destination. For instance, in Alappuzha, there is a preference for boathouses and backwaters. Location and venue are crucial for any booking destination (Vu et al., 2019). Third, there are certain aspects that are unique to homestays and have no relevance in professionally managed hotels, such as satisfaction with homestay owners and families, and the personalized services and experience they provide. Based on the topic models listed, it can be inferred that tourists base their reviews on the

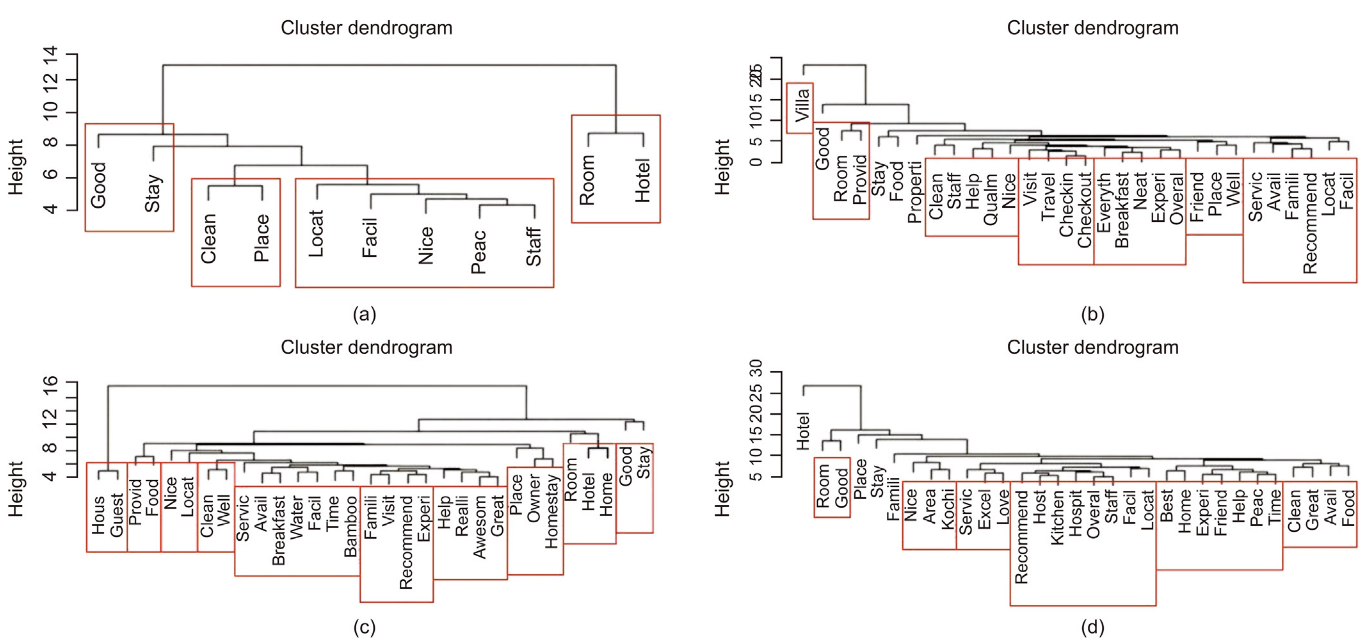

Fig. 4. Dendrograms representing cluster of terms in online reviews of homestays in Trivandrum (a), Wayanad (b), Alappuzha (c), and Kochi (d).

following categories:

Satisfaction with the homestay accommodation: Visitors list their experiences and satisfaction with a homestay based on the room they were provided for their stay. For instance, bamboo rooms are appreciated by some and not by others. It is important to keep in mind that rooms in homestays are part of the owner's house and not separate, as is the case with hotels and lodges.

Homestay facilities: This is an important factor when booking a homestay. The keywords identified in this study included breakfast, water, telephonic services, free mineral water, towels, room facilities, Wi-Fi, intercom, fan, and air conditioning.

Homestay atmosphere: This refers to the interactions of guests with homestay owners in terms of making them feel part of the family and guiding them with their travel plans, itineraries, and so on.

Breakfast served at homestays: Breakfasts are inclusive or separately paid services. Traditional Kerala cuisine dominates the breakfast menu and is generally appreciated by visitors. However, some visitors were dissatisfied with the limited menu and recommended that more items be included for a variety of purposes.

Bathroom: Visitors expressed both satisfaction and dissatisfaction with bathrooms. Negative reviews indicated a lack of cleanliness and the unavailability of hot water, mugs, buckets, and dustbins. Bathroom feedback is crucial in eWOM in hotels and has been studied extensively in emerging economies such as Indonesia.

Hospitality offered by the homestays: This refers to the warmth and hospitality shown by homestay owners and their families. Online reviewers greatly appreciated welcoming guests with traditional coconut milk, inquiring about their journey, and making them feel at home.

Satisfaction with the host and family: Tourists residing in homestays look for localized, authentic experiences that are incomplete without the personal touches provided by homestay owners and their families. The more cordial the relationship, the more satisfied the visitors. This is reflected in the keywords used in the reviews, such as family and owner. This is not found in the eWOM analysis of hotels or other types, in which personal touch is missing. The role of the host is also highlighted in the P2P accommodation listed on Airbnb (Bao et al., 2022; Brochado et al., 2017; Cheng and Jin, 2019; Li et al., 2019; Sthapit and Jimenez-Barreto, 2018; Yu et al., 2020).

Cleanliness in and around homestays: Cleanliness is another crucial factor governing repeated and new bookings. It encompasses all amenities provided by the homestay including bathrooms, washbasins, towels, outdoor premises, food, kitchen, bed linens, pan stains, and

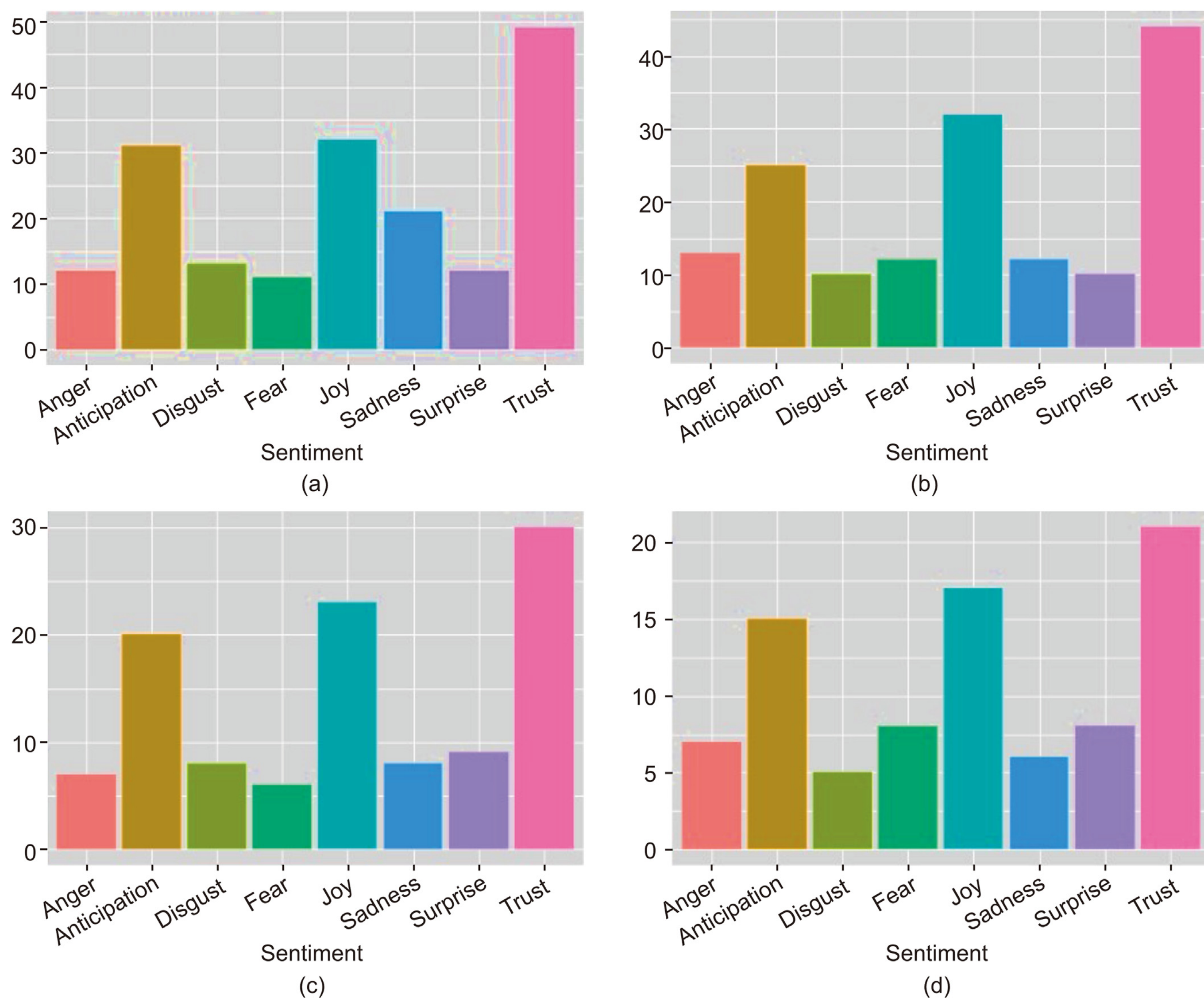

Fig. 5. Emotional analyses of reviews posted about homestays in Alappuzha (a), Kochi (b), Trivandrum (c), and Wayanad (d).

Table 2
Topic models of online reviews for homestays in each of the four destinations.

Destination	Topic number	Keyword	Topic extracted from keyword	Percentage of relevance of tokens from total
Alappuzha (CV score: (0.2751808811742122, UMass: −3.6270931749305575)	1a	[(0, '0.184*"like" + 0.154*"water" + 0.080*"homestay" + 0.072*"have" + 0.066*"that" + 0.063*"will" + 0.059*"facility" + 0.059*"from" + 0.050*"house" + 0.046*"also"')	Boathouse and backwaters	17.4
	2a	(1, '0.143*"well" + 0.137*"really" + 0.134*"clean" + 0.090*"awesome" + 0.089*"breakfast" + 0.058*"provided" + 0.049*"great" + 0.045*"owner" + 0.042*"also" + 0.033*"time"'),	Breakfast	16.8
	3a	(2, '0.199*"service" + 0.191*"location" + 0.107*"facility" + 0.105*"that" + 0.101*"available" + 0.061*"great" + 0.058*"from" + 0.050*"also" + 0.037*"well" + 0.023*"owner"')	Services, facilities and location	16.8
	4a	(3, '0.183*"experience" + 0.152*"homestay" + 0.114*"bamboo" + 0.092*"family" + 0.082*"from" + 0.075*"located" + 0.069*"that" + 0.061*"clean" + 0.039*"owner" + 0.030*"great"')	Satisfaction with the host family	16.7
	5a	(4, '0.108*"have" + 0.103*"there" + 0.103*"would" + 0.077*"only" + 0.074*"time" + 0.072*"owner" + 0.054*"like" + 0.053*"will" + 0.041*"breakfast" + 0.039*"location"')	Location	16.2
	6a	(5, '0.322*"house" + 0.272*"guest" + 0.062*"provided" + 0.055*"there" + 0.040*"service" + 0.036*"facility" + 0.032*"family" + 0.027*"would" + 0.026*"located" + 0.018*"clean"')]	Cleanliness	16.2
Kochi (CV-score: 0.3889462700557723, UMass: −4.915156307074308)	1b	[(0, '0.074*"area" + 0.067*"with" + 0.049*"kochi" + 0.041*"kitchen" + 0.040*"nice" + 0.039*"host" + 0.033*"were" + 0.032*"even" + 0.032*"really" + 0.031*"they"')	Satisfaction with homestay	24.6
	2b	(1, '0.081*"have" + 0.066*"home" + 0.063*"they" + 0.048*"from" + 0.039*"also" + 0.039*"nice" + 0.037*"kochi" + 0.036*"family" + 0.035*"like" + 0.034*"best"')	Satisfaction with host	20.5
	3b	(2, '0.112*"service" + 0.087*"small" + 0.085*"great" + 0.074*"bathroom" + 0.066*"food" + 0.061*"nice" + 0.053*"available" + 0.052*"peaceful" + 0.051*"staff" + 0.044*"check"')	Facilities	19.2
	4b	(3, '0.087*"best" + 0.081*"service" + 0.066*"will" + 0.057*"time" + 0.049*"cochin" + 0.045*"family" + 0.044*"excellent" + 0.043*"hospitality" + 0.034*"amazing" + 0.031*"staff"')	Hospitality	18.7
	5b	(4, '0.065*"were" + 0.051*"facility" + 0.045*"provided" + 0.040*"food" + 0.040*"family" + 0.040*"recommend" + 0.039*"would" + 0.038*"experience" + 0.038*"that" + 0.037*"available"')]	Recommend	17.0
Trivandrum (CV score: 0.2696643499045678, UMass: −3.9533231891034264)	1c	[(0, '0.498*"facility" + 0.426*"peaceful" + 0.016*"hotel" + 0.015*"nice" + 0.015*"were" + 0.015*"staff" + 0.015*"with"')	Satisfaction with facilities provided	35.8
	2c	(1, '0.470*"hotel" + 0.266*"with" + 0.221*"nice" + 0.026*"facility" + 0.006*"staff" + 0.006*"were" + 0.006*"peaceful"')	Satisfaction with the homestay	32.9
	3c	(2, '0.500*"staff" + 0.402*"were" + 0.035*"nice" + 0.025*"hotel" + 0.014*"facility" + 0.013*"peaceful" + 0.011*"with"')]	Satisfaction with staff	31.3
Wayanad (CV score: 0.4035255871168412, UMass −1.0396094761397452)	1d	[(0, '0.562*"facility" + 0.408*"there" + 0.015*"villa" + 0.015*"clean"')	Satisfaction with facilities provided	34.7
	2d	(1, '0.915*"villa" + 0.071*"there" + 0.007*"clean" + 0.007*"facility"')	Satisfaction with the homestay	33.0
	3d	(2, '0.914*"clean" + 0.029*"there" + 0.029*"villa" + 0.028*"facility"')]	Cleanliness of the homestay	32.3

cigarette butts that find frequent mention in the reviews.

Topics specific to unique destinations: Destination-specific factors such as boathouses, backwaters, bamboo huts, and geographical proximity to well-known tourist attractions and temples have also been mentioned in several reviews.

The fifth research question explored the similarities and disparities in online homestay reviews for the four thematic destinations. The comparative analysis of the sentiments, term frequencies, hierarchical clusters, word clouds, and topic models between destinations shows both similarities and dissimilarities.

4.5. Predictive modeling analysis

The sixth research question aims to build a predictive model for reservations based on customer reviews.

To address this question, a 10-fold cross-validation was used. An odds value above one indicates a positive relationship, implying that homestays are very likely to be recommended (Table 3). Thus, the ranked features provide interesting insights into what works and does not include homestay visitors.

Table 4 presents the evaluation performance of the four classifiers (lasso logistic regression, ridge logistic regression, decision tree, and Naïve Bayes classifier). It is evident from Table 4 that the lasso logistic regression model outperforms other benchmark models, with reported accuracy and F1-scores of 97.6% and 97.5%, respectively.

5. Conclusions

This study conducts a systematic analysis of the sentiments, themes, emotions, and key decision attributes that motivate e-bookings for family-run homestays at proximate and competing destinations in Kerala. Our findings add to the hospitality literature by highlighting the service attributes that family-based homestay travelers normally focus on. Opinion mining of online customer reviews from a specific e-booking portal for homestays at four destinations in Kerala highlighted several key points. First, the exploratory analysis shows that homestay visitors in Kochi have relatively better sentiments than those in the other three destinations. The online reviews show several references to homestays in proximity as the "good" and "great". Second, positive sentiments are much higher than negative sentiments in terms of the attributes of homestay accommodation. This highly positive sentiment is in line with the consistent growth in demand for homestays in Kerala. Third, Wayanad showed the highest positive-to-negative ratio, followed by

Table 3
Odds ratio of the features for logistic regression coefficients.

Sl. No.	Feature	Odds ratio	Sl. No.	Feature	Odds ratio
1	High_recommend	26.42370931	17	Good_place	1.802278627
2	Want	4.749385041	18	Citi	1.632726701
3	High	4.538722248	19	Walk	1.474418936
4	View	4.197801039	20	Nice	1.4732526
5	Bus	3.844180592	21	Lot	1.39793459
6	Fine	3.509769903	22	Price	1.377447374
7	Facil_good	3.136310191	23	Food	1.37301699
8	Coupl	2.347318118	24	Full	1.356583295
9	Nearby	2.06757773	25	Amen	1.303117622
10	Provid	1.998770026	26	Comfort	1.294425648
11	Realli	1.987281181	27	Staff	1.156531502
12	Family	1.959732562	28	Nice_room	1.081058155
13	Property	1.913614588	29	Like	1.079334521
14	Cook	1.817368678	30	Facil	1.079315421
15	Home	1.815699699	31	Villa	1.045309907
16	Travel	1.803718902	32	Hotel	1.020759421

Table 4
Performance evaluation of the binary text classifiers.

Classifier	Accuracy	F1-score	Precision	Recall
Lasso logistic regression	97.60	97.50	98.30	96.80
Ridge logistic regression	86.81	92.85	86.66	1.00
Decision tree	95.80	95.66	97.80	93.77
Naïve Bayes	96.50	96.44	97.00	96.00

Trivandrum, Kochi, and Alappuzha. Fourth, negative sentiments were low and were mostly triggered by problems faced by guests during their stay. Negative sentiments were mostly related to mosquitoes, unhygienic toilets, dirty surroundings, noise, and lack of in-house services. Fifth, topic modeling revealed the major themes of discussion in online reviews. These themes revolved around homestay accommodation, breakfast, facilities provided, ambience, bathrooms, cleanliness, and hospitality, as shown by the hosts. Finally, our study revealed that each homestay destination differs in relation to climate, tourist attractions, terrains, food habits, and lifestyles and so on.

Our findings have managerial implications for e-booking sites and homestay owners. First, online reviewers can become unpaid brand ambassadors for homestays if they enjoy satisfactory accommodation. However, dissatisfied customers may also take the e-booking site to list a property with which they have a negative experience. Homestay owners must focus on the customer experience to evoke positive eWOM. Second, e-booking portals must filter homestays with continuously poor feedback from their listed properties. Visitors expect complete value for money and those looking for homestays may consider competition in the form of OYO rooms, a popular hotel brand in India, and other budget hotels. Apart from value, money owners must differentiate their services from unique experiences that traditional hotels cannot provide. Additionally, owners benefit from obtaining the necessary certifications from accreditation agencies to build a strong virtual brand on e-booking sites. Third, in case of negative reviews, owners must respond promptly to specify the corrective measures they have taken to address the issue. This will go a long way to winning back customer trust (Chan and Guillet, 2011; Du et al., 2015; Wei et al., 2013). New positive reviews would also gradually outweigh older negative reviews (Sparks and Browning, 2011). Fourth, text analysis shows that “location identification” and “reachability” are crucial to homestay bookings. Homestays are usually located in residential areas that can be difficult for visitors to locate and reach. Homestay owners must address these issues by providing clear directions or adding maps to their listings. Further, they should also provide and keep the contact details updated on their social media pages so that visitors have no problems locating their homestays. Fifth, visitors look for complementary facilities such as mineral water, strong Wi-Fi, and mobile connectivity. According to O'Mahony and Smyth (2010), free services attract superior ratings from e-booking websites.

This study had a few limitations. First, the data were confined to four destinations in Kerala; therefore, there was homogeneity in the sample. This could restrict the generalizability of the findings to homestays in other regions. Second, the data were sourced from an Indian booking website and many of the reviewers were Indian tourists. A comparative analysis could be performed by sourcing reviews from foreign tourists. This could also be considered in future studies.

CRediT author statement

Jay Krishnanjaykrishnan: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Biplab Bhattacharjee:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maheshwar Pratap:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Janardan Krishna Yadav:** Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Moinak Maiti:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

References

Aakash, A., Gupta Aggarwal, A., 2020. Assessment of Hotel Performance and guest satisfaction through eWOM: big data for better insights. Int. J. Hospit. Tourism Adm. 23 (2), 317–346.

Agag, G.M., El-Masry, A.A., 2017. Why do consumers trust online travel websites? Drivers and outcomes of consumer trust toward online travel websites. J. Trav. Res. 56 (3), 347–369.

Agyeiwaah, E., 2013. Volunteer tourists' motivations for choosing homestay in the Kumasi Metropolis of Ghana. Afr. J. Hosp. Tour. Leis. 2 (3), 1–11.

Bao, Y., Ma, E., La, L., et al., 2022. Examining the airbnb accommodation experience in Hangzhou through the lens of the experience economy model. J. Vacat. Mark. 28 (1), 95–116.

Bassolas, A., Lenormand, M., Tugores, A., et al., 2016. Touristic site attractiveness seen through Twitter. EPJ Data Sci. 5 (1), 12.

Bhattacharjee, B., Sridhar, A., Dutta, A., 2017. Identifying the causal relationship between social media content of a Bollywood movie and its box-office success - a text mining approach. Int. J. Bus. Inf. Syst. 24 (3), 344.

Blei, D., John, L, 2009. Topic Models. In: Srivastava, A.N., Sahami, M. (Eds.), Text Mining: Classification, Clustering, and Applications. Chapman and Hall/CRC, New York, pp. 365–452.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan.), 993–1022.

Boo, S., Busser, J.A., 2018. Meeting planners' online reviews of destination hotels: a twofold content analysis approach. Tourism Manag. 66 (Jun.), 287–301.

Brochado, A., Troilo, M., Shah, A., 2017. Airbnb customer experience: Evidence of convergence across three countries. Ann. Tourism Res. 63 (Mar.), 210–212.

Buzova, D., Sanz-Blas, S., Cervera-Taulet, A., 2018. Does culture affect sentiments expressed in cruise tours' eWOM? Serv. Ind. Jpn. 39 (2), 154–173.

Chan, N.L., Guillet, B.D., 2011. Investigation of social media marketing: how does the hotel industry in Hong Kong perform in marketing on social media websites? J. Travel Tour. Mark 28 (4), 345–368.

Chang, Y.-C., Ku, C.-H., Chen, C.-H., 2019. Social media analytics: extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. Int. J. Inf. Manag. 48 (Oct.), 263–279.

Cheng, M., Jin, X., 2019. What do airbnb users care about? an analysis of online review comments. Int. J. Hospit. Manag. (Part A) 76 (Jan.), 58–70.

Claeys, C., Roozen, I., 2012. Do emotional and functional customer experiences influence customer satisfaction, recommendation, and loyalty? Working Papers. Available at: https://ideas.repec.org/p/hub/wpecon/201202.html.

Ding, S., Gong, S., Li, H., 2016. A new method to detect bursty events from micro-blog posts based on bursty topic words and agglomerative hierarchical clustering algorithm. New Tech. Lib. Info. Serv. 32 (7–8), 12–20.

Du, X., Zhao, Z., Cui, X., 2015. The effect of review valence, new product types and regulatory focus on new product online review usefulness. Acta Psychol.Sin 47 (4), 555–568.

Fang, X., Zhan, J., 2015. Sentiment analysis using product review data. J. Big Data 2 (1), 5.

Filieri, R., McLeay, F., 2014. E-WOM and accommodation: an analysis of the factors that influence travelers' adoption of information from online reviews. J. Trav. Res. 53 (1), 44–57.

Gan, Q., Ferns, B.H., Yu, Y., et al., 2016. A text mining and multidimensional sentiment analysis of online restaurant reviews. J. Qual. Assur. Hosp. Tour 18 (4), 465–492.

Gandomi, A., Haider, M., 2015. Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manag. 35 (2), 137–144.

Geetha, M., Singha, P., Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels - an empirical analysis. Tourism Manag. 61 (Aug.), 43–54.

Genkin, A., Lewis, D.D., Madigan, D., 2005. Sparse logistic regression for text Categorization. DIMACS working group. Available at: http://archive.dimacs.rutgers.edu/Research/MMS/loglasso-v3a.pdf.

Ghose, A., Ipeirotis, P.G., Li, B., 2012. Designing ranking systems for hotels on travel search engines by mining user-generated and crowd sourced content. Mark. Sci. 31 (3), 493–520.

Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. Tourism Manag. 59 (Apr.), 467–483.

Han, H.J., Mankad, S., Gavirneni, N., et al., 2016. What guests really think of your hotel: text analytics of online customer reviews. Cornell Hosp. Q. 16 (2), 3–17.

Hastie, T., Tibshirani, R., Tibshirani, R.J., 2017. Extended comparisons of best subset selection, forward stepwise selection, and the lasso. arXiv preprint arXiv:1707.08692. Available at: https://arxiv.org/pdf/1707.08692.pdf.

He, W., Tian, X., Tao, R., et al., 2017. Application of social media analytics: a case of analyzing online hotel reviews. Online Inf. Rev. 41 (7), 921–935.

Hsu, S.L., Lin, Y.M., 2011. Factors underlying college students' choice homestay accommodation while travelling. World Trans. Eng. Technol. Educ. 9 (3), 196–202.

Hwang, J., Park, S., Woo, M., 2018. Understanding user experiences of online travel review websites for hotel booking behaviours: an investigation of a dual motivation theory. Asia Pac. J. Tourism Res. 23 (4), 359–372.

James, G., Witten, D., Hastie, T., et al., 2013, first ed. An Introduction to Statistical Learning. Springer, New York.

Jia, S.S., 2018. Behind the ratings: text mining of restaurant customers' online reviews. Int. J. Mark. Res. 60 (6), 561–572.

Jockers, M.L., 2015. Syuzhet: extract sentiment and Plot Arcs from text. Github. Available at: https://github.com/mjockers/syuzhet.

Khorsand, R., Rafiee, M., Kayvanfar, V., 2020. Insights into TripAdvisor's online reviews: The case of Tehran's hotels. Tour. Manag. Perspect. 34 (Apr.), 100673.

Kulshreshtha, S., Kulshrestha, R., 2019. The emerging importance of "homestays" in the Indian hospitality sector. Worldw. Hosp. Tour.Themes 11 (4), 458–466.

Li, J., Chu, Y.J., Zhao, Z.B., 2015a. Tibetan tourism hotspots: Co-word cluster analysis of English blogs. Tour. Trib. 30 (3), 35–43.

Li, J., Hudson, S., So, K.K.F, 2019. Exploring the customer experience with Airbnb. Int. J. Cult. Tour. Hosp. Res. 13 (4), 410–429.

Li, Y.R., Lin, Y.C., Tsai, P.H., et al., 2015b. Traveller-generated contents for destination image formation: Mainland China Travellers to Taiwan as a case study. J. Trav. Tourism Market. 32 (5), 518–533.

Litvin, S.W., Sobel, R.N., 2019. Organic versus Solicited hotel TripAdvisor reviews: Measuring their respective characteristics. Cornell Hosp. Q. 60 (4), 370–377.

Llach, J., Marimon, F., del Mar Alonso-Almeida, M., et al., 2013. Determinants of online booking loyalties for the purchasing of airline tickets. Tourism Manag. 35 (Apr.), 23–31.

Lyu, J., Li, M., Law, R., 2019. Experiencing P2P accommodations: Anecdotes from Chinese customers. Int. J. Hosp. Manag. 77 (Jan.), 323–332.

Magnini, V.P., Crotts, J.C., Zehrer, A., 2011. Understanding customer delight. J. Trav. Res. 50 (5), 535–545.

Mattila, A.S., Ro, H., 2008. Discrete negative emotions and customer dissatisfaction responses in a casual restaurant setting. J. Hosp. Tour. Res. 32 (1), 89–107.

Mimno, D., Wallach, H., Talley, E., et al., 2011. Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, ACL, pp. 262–272.

Molina-Azorín, J.F., Tarí, J.J., Pereira-Moliner, J., et al., 2015. The effects of quality and environmental management on competitive advantage: a mixed methods study in the hotel industry. Tourism Manag. 50 (Oct.), 41–54.

Newman, D., Noh, Y., Talley, E., et al., 2010. Evaluating topic models for digital libraries. In: Proceedings of the 10th Annual Joint Conference on Digital Libraries - ACM, vol. 10, pp. 215–224.

O'Mahony, M.P., Smyth, B., 2010. A classification-based review recommender. Knowl. Base Syst. 23 (4), 323–329.

Ortigosa, A., Martín, J.M., Carro, R.M., 2014. Sentiment analysis in Facebook and its application to e-learning. Comput. Hum. Behav. 31 (Feb.), 527–541.

Özer, M., Küçüksakarya, S., Maiti, M., 2022. Nexus of tourism demand, economic growth, and external competitiveness in leading tourist destination countries. Tourism Manag. Perspect. 42 (Apr.), 100965.

Racherla, P., Connolly, D.J., Christodoulidou, N., 2013. What determines consumers' ratings of service providers? An exploratory study of online traveler reviews. J. Hospit. Market. Manag. 22 (2), 135–161.

Rossetti, M., Stella, F., Zanker, M., 2016. Analyzing user reviews in tourism with topic models. Inf. Technol. Tourism 16 (1), 5–21.

Schuckert, M., Liu, X., Law, R., 2015. Hospitality and tourism online reviews: recent trends and future directions. J. Trav. Tourism Market. 32 (5), 608–621.

Shaw, C., 2007. The DNA of customer experience: How emotions drive value (Vol. 1). Palgrave Macmillan, New York.

Socher, R., Perelygin, A., Wu, J. Y, et al., 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642.

Sparks, B.A., Browning, V., 2011. The impact of online reviews on hotel booking intentions and perception of trust. Tourism Manag. 32 (6), 1310–1323.

Stevens, K., Kegelmeyer, P., Andrzejewski, D., et al., 2012. Exploring topic coherence over many models and many topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 952–961.

Sthapit, E., Jiménez-Barreto, J., 2018. Exploring tourists' memorable hospitality experiences: an Airbnb perspective. Tourism Manag. Perspect. 28 (Oct.), 83–92.

Thapa, B., Anicar, D.M., Malini, D.H., 2018. Unraveling tourists" preferred homestay Attributes from online reviews: a sentiment analysis approach. Int. J. Pure Appl. Math. 119 (15), 1567–1585.

Thimm, T., 2017. The Kerala tourism model - an Indian state on the road to sustainable development. Sustain. Dev. 25 (1), 77–91.

Tong, Z., Zhang, H., 2016. A Text Mining Research Based on LDA Topic Modelling. Computer Science & Information Technology (CS & IT), pp. 201–210. Available at: https://www.airccj.org/CSCP/vol6/csit65316.pdf.

Tussyadiah, I.P., Pesonen, J., 2016. Impacts of peer-to-peer accommodation use on travel patterns. J. Trav. Res. 55 (8), 1022–1040.

Vu, H.Q., Li, G., Law, R., 2019. Discovering implicit activity preferences in travel itineraries by topic modeling. Tourism Manag. 75 (Dec.), 435–446.

Wang, Y., 2007. Customized authenticity begins at home. Ann. Tourism Res. 34 (3), 789–804.

Wei, W., Miao, L., Huang, Z.J., 2013. Customer engagement behaviors and hotel responses. Int. J. Hospit. Manag. 33 (Jun.), 316–330.

Xiang, Z., Du, Q., Ma, Y., et al., 2017. A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. Tourism Manag. 58 (Feb.), 51–65.

Xu, X., Wang, X., Li, Y., Haghighi, M., 2017. Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. Int. J. Inf. Manag. 37 (6), 673–683.

Yang, C.-C., Jou, Y.-T., Cheng, L.-Y., 2011. Using integrated quality assessment for hotel service quality. Qual. Quantity 45 (2), 349–364.

Yong, K., Hassan, R.A., 2019. The relationships between social media marketing and entrepreneurial success: a conceptual study on homestay business in Sabah, Malaysia. Rev. Integr. Bus. Econ. Res. 8 (1), 25–33.

Yu, M., Cheng, M., Yu, Z., et al., 2020. Investigating Airbnb listings? amenities relative to hotels. Curr. Issues Tour. 25 (19), 3168–3185.

Zhang, H.Y, Ji, P., Wang, J.Q, et al., 2017. A novel decision support model for satisfactory restaurants utilizing social information: a case study of TripAdvisor.com. Tourism Manag. 59 (Apr.), 281–297.

Zhang, T., Li, B., Hua, N., 2021. Chinese cultural theme parks: text mining and sentiment analysis. J. Tourism Cult. Change 20 (1–2), 37–57.

Zorfas, A., Leemon, D., 2016. An emotional connection matters more than customer satisfaction. Harv. Bus. Rev. Available at: https://hbr.org/2016/08/an-emotional-connection-matters-more-than-customer-satisfaction.