Assessing Hotel Attribute and Facilities to Online Hotel Popularity: Data Mining from Google

Muhammad Izharuddin[1](✉) and Jeng-Chung Chen[2]

[1] University of Surabaya, Surabaya, Indonesia
izharuddin@staff.ubaya.ac.id
[2] National Cheng Kung University, Tainan, Taiwan

Abstract. Today's online consumer reviews are one of the most important elements for hotel marketing. This study examines how hotel facilities, prices, and ratings can affect online reviews of hotel consumers. Using a data mining approach from 1,248 hotels in major cities in Indonesia, this paper estimates the trigger factors for online consumer reviews, including ratings, stars, price, and other facilities (Parking, Wi-Fi, AC, Pool, breakfast, and spa). 2.5 million hotel reviews from the Google platform were collected and analyzed using regression. Empirical evidence shows the relative influence of facilities, absolute rating, stars, and price factors on the popularity of online hotels. These findings provide a managerial basis for enhancing a hotel's online presence across multiple platforms by strategically leveraging review factors of importance.

Keywords: Hotel · Facilities · Popularity · Data Mining

1 Introduction

Consumer reviews have increased dramatically on social media networks due to technological advancements over the past few decades. Many buyers now use the reviews of other customers who have purchased various goods and services as part of their decision-making process [1]. Consumer feedback is particularly significant for learning about the experience of products like hotel rooms because their quality is frequently unknown until after purchase [2]. Pine et.al. [3] shows that online reviews provide reliable information for lowering the risk associated with experiential purchases. Online reviews allow customers to read about other customers' experiences with products and services without actually using them. Consumer evaluations that customers have already purchased represent a specific evaluation level and experiential items feedback, offering potential customers a crucial resource to aid in their decision-making and help them choose the item that best meets their preferences. As a result, approximately 50% of customers use online review sites to learn more about their online travel purchases. About three-quarters of consumers now consider online consumer reviews when making trip arrangements [4]. Online customer reviews have taken the place of offline word-of-mouth (WOM) and other business-to-consumer (B2C) and WOM communication quality of service from service providers as a major information source for consumers [5].

W. R. Murhadi et al. (Eds.): INSYMA 2023, AEBMR 256, pp. 385–392, 2024.
https://doi.org/10.2991/978-94-6463-244-6_57

Online customer reviews have become one of the most significant factors for marketers in recent years due to the widespread use of these reviews by consumers; thus, substantial research has been done on online customer reviews [6]. There is a growing corpus of scholarly material examining how the success of offline businesses is impacted by online consumer reviews [4, 7].

The economic benefit of online consumer evaluations has emerged as one of the most crucial study areas. However, the empirical literature still needs to catch up, which is what drives our investigation. First, aspects that online consumer reviews have on business performance have been covered in prior literature, but they have typically been examined in isolation and fragments without taking into account multiple aspects of online consumer review factors (e.g.,Sun [8]).

Second, prior research on social media has mainly focused on commodities with short life cycles or one-time purchases, such as software, books, and movies [1, 9, 10]. Information goods, however, stand out because they have short product life cycles and adhere to a general exponential trend (Moe & Fader, 2001). Studies exploring social media exposure for less glamorous consumers and commercial products like hotels are uncommon.

By experimentally analyzing the impact of service facility elements and hotel qualities on online hotel customer reviews, this research tries to close this knowledge gap. Based on a distinct data set of 2.5 million reviews from 1,248 hotels in Indonesia's major cities, our analysis (i.e., Jakarta, Surabaya, Bandung, Bali, and Yogyakarta). Our data set is distinctive in two ways. First, by using this new data, we can investigate the direct effects of consumer review factors within a single independent platform while also avoiding the confounding effects of the various platforms used in the majority of earlier studies [11]. Second, critical phenomena at the individual, dyadic hotel level have largely been ignored in prior media research, mainly concentrating on aggregate market-level effects of consumer review factors. Our data collection enables us to analyze individual hotel-level disaggregation to determine the impact of consumer review elements. In order to quantify the effectiveness of amenity features and hotel attributes on online consumer reviews while controlling for variance resulting from differences among hotels, we used our data set to use a combination of econometric models (fixed effect and random effect). This study clarifies the factors that affect internet reviews. It backs a holistic approach to social media marketing that focuses on the strategic application of online customer review variables to influence hotel company success.

2 Literature Review

It has been demonstrated that the recent rise in user-generated online reviews has a major impact on consumers' purchasing decisions. As a way for customers to learn more, give opinions, make purchasing choices, and for businesses to gather consumer data and provide recommendations, it is customary to incorporate these user-generated evaluations alongside product descriptions.

Consumer opinions at least make the brand visible and include it in the options available to them [12]. Peer consumers frequently adopt the ideas of earlier reviewers due to peer pressure to conform, claims the social contagion theory [13]. Customers strive to

persuade their fellow customers to see their point of view by posting recommendations and views about goods and services on social media platforms to influence their purchasing decisions [14]. Consumer feedback thus reinforces the notion that many other customers either purchase or do not purchase the same goods or service, which can help consumers justify their purchasing choices. Prior studies have validated the prevalence of social contagion in online customer evaluations. For instance, [7] demonstrates that customer feedback from the past greatly affects future consumer behavior. [15] discovers in an experimental environment that consumers who expressed their comments tend to change their product assessments negatively after reading unfavorable reviews from other reviewers. It suggests that previously examined hotel features and amenities impact customer posting behavior. Consequently, we suggest the following:

Hypothesis: Various types of hotel services and attributes have an impact on the quantity of online hotel consumer reviews.

3 Methodology

3.1 Data and Measurement

The first step for data mining is data collection. This step contains the selection of hotels for research. A Google Maps search for hotels in five major cities of Jakarta, Bali, Surabaya, Bandung, and Yogyakarta in Indonesia was conducted. The city was chosen because it has the largest number of hotels in Indonesia. The author first searched for the key term "Hotel [name of city]" to collect all available data. The selection of these keywords was intended to simplify searches. The search used Python scripts and the Google Maps API. In total, there were 1,952 hotels on the initial list that were successfully screened. From this list, the data was then selected and further filtered. This stage is data preparation, resulting in 1,248 hotel samples. Hotels were filtered using two criteria: blank reviews and reviews with less than 100 reviews. There were two reasons for using the 100-review limit for hotels: (1) to ensure representation of the hotel and (2) to account for the size of the data set by excluding less relevant reviews. The data available at each hotel in the final sample was collected using a Python script and sorted by relevance. In total, this study collected around 2.5 million relevant reviews in December 2022 and January 2023.

The data analysis phase involves codification. In data regarding prices, the symbol Rp. (rupiah) was omitted and replaced with a numeric one. The next codification was regarding hotel's services. If a hotel has a parking service, it is coded 1; if it does not, it is coded 0. It applies to other services: Wi-Fi, air conditioning, swimming pool, breakfast, spa, airport shuttle, and fitness center. We also calculated the total facility. Hotels that have all these facilities are counted as 8. Hotels that do not have one facility are counted as 7, and so on.

Google, a popular consumer review website where online opinion and review distribution had an immediate and widespread impact, served as the study's context [16]. Along with offering map services, Google serves as a platform for regular consumers to voice their thoughts about the caliber of service providers and read referrals from other customers [17]. In the meantime, hotel management should proactively encourage

their customers to share information about their items online by regularly posting information about their products and services [13]. Our sample includes big and medium hotels, in addition to independent and network hotels, in contrast to most prior studies that concentrate on large hotels or hotel chains.

3.2 Multiple Regression

Multiple regression is a statistical tool that can test hypotheses about the relationship between the dependent variable and continuous predictor variables [18]. Multiple regression modeling can help answer research questions by examining the relationship between attributes and hotel facilities on the popularity of online hotel reviews. The author used SPSS software to perform multiple regression analyses. This study used the approach of using the number of reviews as the dependent variable of hotel popularity. The variables of hotel facilities and attributes were used to predict hotel popularity. This approach allows the hotel to see the effect of service attributes on the level of online popularity of the hotel.

3.3 Model Specifications

Reviews are from individual customers for every hotel in our collection. Relevant econometric models must be used to compensate for individual heterogeneity using this data structure. In order to assess the data and account for non-observable characteristics among hotels, we employed three models (hotel culture, management strategy). The models that arise are:

$$\begin{aligned} TOTALREVIEWS = & \beta 0 + \beta 1RATING + \beta 2STARS + \beta 3PRICE + \beta 4PARKING \\ & + \beta 5WI - FI + \beta 6AC + \beta 7SWIMMINGPOOL + \beta 8BREAKFAST \\ & + \beta 9SPA + \beta 10AIRPORTPICKUP + \beta 11RESTAURANT \\ & + \beta 12FITNESSCENTER + \varepsilon \end{aligned} \tag{1}$$

$$TOTALREVIEWS = \beta 0 + \beta 1RATING + \beta 2STARS + \beta 3PRICE + \beta 4FACILITIES + \varepsilon \tag{2}$$

$$\begin{aligned} TOTALREVIEWS = & \beta 0 + \beta 1PARKING + \beta 2WI - FI + \beta 3AC \\ & + \beta 4SWIMMINGPOOL + \beta 5BREAKFAST + \beta 6SPA \\ & + \beta 7AIRPORTPICKUP + \beta 8RESTAURANT + \beta 9FITNESSCENTER + \varepsilon \end{aligned} \tag{3}$$

Where: NUMBER OF REVIEWS is the number of reviews on the hotel. RATING is the average rating that customers give on the platform. PRICE is the average price for the rooms that the hotel offers. PARKING is a vehicle parking facility from the hotel for customers. WI-FI is an internet connection facility from the hotel for customers. AC is the air conditioning facility of the hotel for customers. SWIMMING POOL is a swimming infrastructure facility from the hotel for customers. BREAKFAST is a free meal from the hotel for customers. SPA is a body care facility from the hotel for

customers. AIRPORT PICKUP is a shuttle service for hotel customers. RESTAURANT is a facility that provides food from hotels to customers. FITNESS CENTER is a hotel's body health facility for customers. FACILITIES are the aggregate facilities of the hotel for customers.

4 Results and Findings

We estimate the number of reviews in the hotel popularity model and present the estimation results in Table 1. By estimating the three models, we examine the effect of attribute and facility factors on the number of customer reviews while calculating these variables in detail.

Table 1 shows the regression results. In the beta column, the coefficient and t-statistics model (1) show that the hotel rating attribute, hotel star attribute, and hotel price attribute impact the number of online hotel reviews. On the price attribute, the impact is negative. In addition, hotel facilities, including Parking, Wi-Fi, AC, Swimming Pool, Breakfast, and Spa, also impact the number of online hotel reviews. Meanwhile, airport pick-up facilities, restaurants, and fitness centers have no impact on the hotel's online popularity. The regression model (1), which includes detailed attributes and facilities, can estimate the popularity of hotels by 29%. The column beta coefficient and t-statistics model (2) show that the hotel rating attribute, hotel star attribute, hotel price attribute, and hotel

Table 1. Regression results

	Model 1	**Model 2**	**Model 3**
	Beta	**Beta**	**Beta**
Ratings	1143***	1275***	
Star	733***	1135***	
Price	–0.000177***	–0.00019***	
Parking	1322**		1251*
Wi-Fi	1669***		1106***
Air Conditioning	1547**		1059***
Swimming pool	1526**		1506***
Breakfast	1251*		1003*
Spa	2572***		2814***
Airport Pickup	706		258.84017
Restaurant	963		384.92585
Fitness centre	313		437.64042
Facility		442**	
R^2	0.29202688	0.249151962	0.235191636
Durbin-Watson	1.605344877	1.611363976	1.493431963

aggregate facilities impact the number of online hotel reviews. On the price attribute, the impact is also negative. The regression model (2), which includes the total attributes and facilities, can estimate the hotel's popularity by 25%. The beta coefficient column and the t-statistics model (3) show that hotel facilities, including parking, Wi-Fi, air conditioning, swimming pool, breakfast, and spa, impact the number of online hotel reviews. Meanwhile, airport pick-up facilities, restaurants, and fitness centers have no impact on the hotel's online popularity. The regression model (3), which includes the detailed facilities, can estimate the hotel's popularity by 24%.

5 Conclusions and Implications

Our analysis shows that hotel facilities which include Parking, Wi-Fi, AC, Swimming Pool, Breakfast, and Spa, as well as hotel attributes consisting of rating, stars, and price have an impact on the number of online hotel reviews, partially supporting the hypothesis. Specifically, that influence on online reviews can last for at least two quarters. Additionally, customers frequently express a wide range of opinions about the caliber of hotel services and goods. However, consistency of opinion affects how well-liked the hotel is.

These findings have significant ramifications for managerial research and practice. We specifically demonstrate how enhancing a hotel's features and traits in terms of the volume of guest reviews can boost a hotel's popularity. This link reinforces the significance of the quality of the product and hospitality. Despite the emergence of digital networking in the hospitality sector, the secret to hotels' popularity continues to be their exceptional service quality. In order to elicit good feedback from clients immediately following their departure or throughout their stay, hotel businesses should constantly improve the quality of their services. Hotel managers can also offer incentives to entice guests to leave reviews and recognize loyal reviewers.

6 Limitations and Future Studies

There are restrictions to this study. Although we have only begun to explore several potential routes for future research into the popularity of hotels, our study may serve as inspiration for others. First, we are convinced of the accuracy of our dataset set because it represents genuine consumer review activity on social media channels for five important Indonesian tourist destinations. However, the hotel lodging market in Indonesia does not necessarily reflect that of the country's other five major cities. Collecting a greater representative group by academics is advocated to support estimations of the factor impact of online customer evaluations. Future studies that combine actual behavioral data gathering with cutting-edge data analytics methods to maximize strategic vision in the hospitality sector are something we look forward to.

References

1. J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," *J. Mark. Res.*, vol. 43, no. 3, pp. 345–354, 2006.
2. B. J. Pine, J. H. Gilmore, and others, "Welcome to the experience economy," 1998.
3. U. Gretzel and K. H. Yoo, "Use and impact of online travel reviews," *Inf. Commun. Technol. Tour. 2008*, pp. 35–46, 2008.
4. K. L. Xie, Z. Zhang, and Z. Zhang, "The business value of online consumer reviews and management response to hotel performance," *Int. J. Hosp. Manag.*, vol. 43, pp. 1–12, 2014.
5. D. Lewis and D. Bridger, *Soul of the new consumer: Authenticity-what we buy and why in the new economy*. Nicholas Brealey, 2011.
6. P. Cunningham, B. Smyth, G. Wu, and D. Greene, "Does TripAdvisor makes hotels better?," 2010.
7. W. W. Moe and M. Trusov, "The value of social dynamics in online product ratings forums," *J. Mark. Res.*, vol. 48, no. 3, pp. 444–456, 2011.
8. M. Sun, "How does the variance of product ratings matter?," *Manage. Sci.*, vol. 58, no. 4, pp. 696–707, 2012.
9. W. Duan, B. Gu, and A. B. Whinston, "The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry," *J. Retail.*, vol. 84, no. 2, pp. 233–242, 2008.
10. Y. Liu, "Word of mouth for movies: Its dynamics and impact on box office revenue," *J. Mark.*, vol. 70, no. 3, pp. 74–89, 2006.
11. A. T. Stephen and J. Galak, "The effects of traditional and social earned media on sales: A study of a microlending marketplace," *J. Mark. Res.*, vol. 49, no. 5, pp. 624–639, 2012.
12. C. R. Clark, U. Doraszelski, and M. Draganska, "The effect of advertising on brand awareness and perceived quality: An empirical investigation using panel data," *Qme*, vol. 7, no. 2, pp. 207–236, 2009.
13. L. Zhang, B. Pan, W. Smith, and X. R. Li, "An exploratory study of travelers' use of online reviews and recommendations," *Inf. Technol. Tour.*, vol. 11, no. 2, pp. 157–167, 2009.
14. S.-C. Chu and Y. Kim, "Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites," *Int. J. Advert.*, vol. 30, no. 1, pp. 47–75, 2011.
15. A. E. Schlosser, "Posting versus lurking: Communicating in a multiple audience context," *J. Consum. Res.*, vol. 32, no. 2, pp. 260–265, 2005.
16. S. W. Litvin, R. E. Goldsmith, and B. Pan, "Electronic word-of-mouth in hospitality and tourism management," *Tour. Manag.*, vol. 29, no. 3, pp. 458–468, 2008.
17. I. Jeacle and C. Carter, "In TripAdvisor we trust: Rankings, calculative regimes and abstract systems," *Accounting, Organ. Soc.*, vol. 36, no. 4–5, pp. 293–309, 2011.
18. C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002.