

Analyzing Online Customer Reviews for the Hotel Classification in Vietnam

Ha Thi Thu NGUYEN¹, Tuan Minh TRAN², Giang Binh NGUYEN³

Received: April 20, 2021 Revised: July 04, 2021 Accepted: July 15, 2021

Abstract

The classification standards for hotels in Vietnam are different from many other hotel classification standards in the world. This study aims to analyze customer reviews on the TripAdvisor website to develop a new algorithm for hotel rating that is independent of Vietnam's hotel classification standards. This method can be applied to individual hotels, or hotels of a region or the whole country, while online booking sites only rate individual hotels. Data was crawled from TripAdvisor with 22,287 reviews of 5 cities in Vietnam. This study used a statistical model to analyze the review dataset and build an algorithm to rate hotels according to aspects or hotel overall. The results have less rating deviation when compared to the TripAdvisor system. This study also supports hotel managers to regularly update the status of their hotels using data from customer reviews, from which, managers can strategize long-term solutions to improve the quality of the hotel in all aspects and attract more travelers to Vietnam. Moreover, this method can be developed into an automatic system to rate hotels and update the status of service quality more quickly, thus, saving time and costs.

Keywords: Hotel Management, Customer's Review, Vietnamese Hotel, Aspect Rating, Overall Rating

JEL Classification Code: C10, C40, C82, M15, P21, P25

1. Introduction

Hotel ratings are often used to classify hotels according to their quality. From the initial purpose of informing travelers on basic facilities that can be expected, the objectives of hotel rating have expanded into a focus on the hotel experience as a whole (Le et al., 2020). Currently, hotel rating has become one of the important factors to promote a hotel brand to affirm its position. There are a number of hotel classification organizations in the world. Each organization develops its own criteria for hotel classification. In recent years, the strong development of online hotel booking systems shows that the trend of

tourism has changed, the quality and service also changed according to the opinions and evaluations of guests instead of the criteria developed by organizations. Many hotel classification organizations update their criteria according to guest preferences and needs when mining guest opinion (Rhee & Yang 2015; Xue et al., 2017).

The hotel industry is a competitive market where hotels are continuously trying to improve their service quality to satisfy customers because higher customer satisfaction will lead to customer loyalty and increase the return rate of guests. Online booking systems help to promote the hotel's brand and reputation better than traditional business but it increases competition among hotels (Kim et al., 2021). Ratings by hotel classification organizations are not of much significance because travelers now tend to trust the reviews of travelers who have experienced the same service. Therefore, today, the opinions/reviews of travelers have more weight than any other standard. When selecting hotels, travelers rely on the number of positive reviews. Travelers selecting hotels base their decisions on online reviews. TripAdvisor is a hotel rating system based on the reviews of travelers/customers who have stayed in that hotel and experienced their service. As for the TripAdvisor rating system, a review of a hotel furnishes information on the quality

¹First Author and Corresponding Author. Lecturer, Department of Ecommerce, Electric Power University, Vietnam [Postal Address: 235 Hoang Quoc Viet St., Bac Tu Liem Dist., Hanoi City, 10000, Vietnam] Email: hanntt@epu.edu.vn

²Associate Professor of Economics, Vietnam Academy of Social Sciences, Vietnam. Email: minh Tuanvass@gmail.com

³Lecturer, Vietnam Academy of Social Sciences, Vietnam. Email: ng.binh.giang@gmail.com

of accommodation, restaurant, and on-site spots worth visit. TripAdvisor uses an online scoring method on the website system for hotels across the globe the world. TripAdvisor star rating is based on third-party users with respect to the quality of services – staff quality, quality of the amenities available. It provides full data such as hotel overall rating, list of reviews, hotel aspects rating, and hotel ranking. Online hotel rating has now become a new criterion that hotel managers pursue because of its usefulness and convenience (Wang et al., 2019; Le et al., 2020):

- The standard rating is based on the traveler’s opinion.
- The price is equivalent among the same level hotels.
- Measuring the quality of hotel service based on travelers’ reviews will help to better understand travelers’ need.
- Help travelers make decisions faster when they book online.

The first behavior of a traveler when booking online is to look at the number of stars assigned to the hotel, and second, they tend to read the reviews related to that hotel (Masiero & Nicolau 2016; El-Said 2020; Goeltom et al. 2020). Star ratings are indicators used to classify hotels according to their quality. For customers, they can lead the decision-making process; for staff, they can help to define best practices and guidelines, as well as helping to attract customers in a competitive market. Online hotel rating systems such as TripAdvisor rate only individual hotels (De Pelsmacker et al., 2018). Therefore, the tourism managers do not have a general overview of the hotels of a region or the whole country.

Tourism has become one of the major sectors within the economy of Vietnam in recent years. Vietnam is the only Southeast Asian country among 10 world’s fastest-growing travel destinations. The classification standard system of Vietnamese hotels is approved by the Ministry of Science and Technology since 2015, however, the classification standards are not in line with today’s online travel trends. In this study, a new Vietnamese hotel rating method is proposed, which will help managers in the Vietnam Ministerial and Department of Culture, Sport and Tourism have a detailed rating system of Vietnamese hotels based on all the aspects of the hotel, and overall, can be applied to hotels in a region or for the whole of Vietnam, respectively (<https://vietnamtourism.gov.vn>).

2. Literature Review

2.1. Hotel Classification System

Hotel classification plays an important role in developing the tourism and hospitality industry. Hotel classification

aims to achieve the following main objectives: (1) Improving the quality of guest service, and ensuring the interests of tourists so as to attract tourists; (2) As a standard for building hotels, (3) The standard of providing service to ensure the consistency of service quality between types of hotels in each country and between the nations; (4) Evaluating and measuring the quality of the hotel; (5) Determining service prices and have appropriate pricing policies for each hotel segment within the global market (Abrate et al., 2021; Sánchez-Lozano et al., 2021).

There are many different hotel classification standards in the world. HOTREC is the umbrella association of hotels, restaurants, pubs, cafes, and similar establishments in Europe, and is the voice of the hospitality industry in Europe. Since 2004, HOTREC and its associations have been working on bringing the hotel classification systems in the various European countries closer to one another. Under the auspices of HOTREC, hotel associations of Austria, Czech Republic, Germany, Hungary, the Netherlands, Sweden, and Switzerland founded the Hotelstars Union as a common standard for rating hotels in Europe (<https://www.hotelstars.eu/criteria/>). Hotels in the UK star ratings are awarded and managed by four Hotel rating boards: The AA, Visit Britain, Visit Scotland, and Visit Wales (<https://www.theaa.com/>). The standard of a 5-star hotel in the UK requires hotels to have among others a spa, fitness center, butler service, valet parking, reception, afternoon tea, and 24/7 room service. Australia is one of the countries with the strictest hotel classification standards in the world. Therefore, Australia now has only thirty-nine 5-star hotels. With more than 200 different criteria to compare, evaluate and score (room size, opening hours, etc.) it is not easy for hotels in Australia to achieve a high star rating. In the US, hotel classification standards focus on quality bedrooms. This is one of the top criteria for a hotel rating. In addition, a 5-star hotel must also have services such as alarm service, luxury spas, golf courses, tennis courts, personal trainers, etc. or even babysitting services.

In Vietnam, hotels are classified by National Standard TCVN 4391: 2009 (Vietnam Directorate for Standards, Metrology, and Quality, <https://tcvn.gov.vn>). Hotels are rated according to location, architecture, equipment and service facilities, service quality and service level, staff, cleanliness, room availability and standards, etc. Hotels from 1 to 5 stars are hotels with high-quality facilities, equipment, and services, meeting the diverse needs of guests’ requirements such as food, accommodation, activities, and entertainment according to international standards.

2.2. Hotel Star Rating from Analyzing Online Review

Data is everywhere and it is revolutionizing the field of statistics radically (Lee 2020; Park & Javed 2020). Undeniably big data has huge potential for many fields

of statistics. The emergence of big data is also changing the working environment of statisticians. Ignoring innovation will push statistical agencies out of the information market (Kim & Yoo 2021). Tourism boards and companies in the tourism sector can benefit from data of this type in many ways. Players in the tourism industry can now make informed decisions based on analytics and number-driven data. They can identify targeted groups of potential customers at every stage in the trip planning process. They can also increase efficiency and the quality of services (Jeon et al., 2019).

Online travelers tend to trust reviews of customer experience more than hotel advertisements. This online review data over the years become big data and is readily available for anyone to tap into. Big data analysis using artificial intelligence tools and models was suggested to match the tasks and challenges of big data (Park & Javed 2020; Luo & Tang 2019; Manes & Tchetchik 2018; Chang et al., 2019). TripAdvisor uses data of online guest reviews to rate hotels. Officially, TripAdvisor follows bubble rating and star rating. TripAdvisor bubble rating is indicative of the summary of user reviews. Thus, it indicates real reviews by real people at a point in time. The TripAdvisor bubble rating is a 1 to 5 scale. TripAdvisor star rating is based on third-party users with respect to the quality of services – staff quality, quality of the amenities available (Al-Natour & Turetken 2020; Bigné et al., 2020).

There are many studies on opinion and aspect ratings for products, but hotel ratings have a few studies mentioned. The rating problem usually has two main approaches, one is based on supervised learning and the other is based on unsupervised (Sokhin et al., 2020; Lai & Hsu 2021). While supervised methods achieve high scores, it is hard to use them in real-world applications due to the lack of labeled datasets. For aspect ratings, unsupervised learning is more effective because travelers often score overall and do not score aspects (Hu et al., 2016). Aspects are only mentioned in reviews, however, customers actually choose products based on their interests, such as color, function, services, etc.

Studies in the field of aspect rating focus on analyzing customer evaluation data using unsupervised learning. Lai and Hsu (2021) proposed aspect-based rating prediction methods, which integrate aspect detection and sentiment analysis to generate user preference and business performance, combined with the results of social behavior analysis to predict the ratings of the businesses that users will be interested in the future. Margaris and Vassilakis (2018) used a methodology of dynamic average computing based on previous ratings; this method can improve the accuracy of results. Sokhin et al. (2020) presented a novel unsupervised neural network with a convolutional multi-attention mechanism, that allows extracting pairs (aspect, term) simultaneously, and demonstrates the effectiveness on the real-world dataset.

Few studies on hotel ratings used unsupervised learning. Akhtar et al. (2017) analyzed hotel reviews and gave information that ratings might overlook. The reviews and metadata are crawled from the website and classified into predefined classes as per some of the common aspects. Then the Topic modeling technique (LDA) is applied to identify hidden information and aspects, followed by sentiment analysis on classified sentences and summarization. Wang et al. (2010) defined and studied a new opinionated text data analysis problem called Latent Aspect Rating Analysis (LARA), which aimed at analyzing opinions expressed about an entity in an online review at the level of topical aspects to discover each individual reviewer's latent opinion on each aspect as well as the relative emphasis on different aspects when forming the overall judgment of the entity.

Sharma et al. (2019) proposed a hotel ranking model based on the aspect ratings accessed from Tripadvisor website. The aspects play the role of criteria consisting of service, cleanliness, value, sleep quality, room, and location. These ratings are classified into positive, neutral, and negative sentiments, which are transformed to Neutrosophic numbers and results in the formation of interval-valued Neutrosophic decision matrix. Also, since the aspect weights are completely unknown, a non-linear programming model called maximizing deviation method is employed. Last, the aspect weights and decision matrix are combined to perform the procedure required for applying technique for order preferences by similarity to ideal solution method for ranking five alternative hotels.

Luo and Tang (2019) modified the Latent Aspect Rating Analysis (LARA) to achieve this research objective. They identified five aspects; communication, experience, location, product/service, and value. Joy and surprise are the primary emotions shown in textual reviews. This study provided an innovative research venue for incorporating both textual reviews and numerical ratings into the assessment. Xue et al. (2017) proposed two topic models which explicitly model aspect ratings as observed variables to improve the performance of aspect rating inference on unrated reviews. The experiment results showed that their approaches outperform the existing methods on the data set crawled from the TripAdvisor website.

In this study, the authors used text analysis techniques to rate hotels according to the hotel's aspect and a deep layer neural network model to rate the hotels overall with real data of Vietnamese hotels that is taken from TripAdvisor. The low error shows that the proposed model is really effective and applied in practice in the star rating of hotels in Vietnam. This research with Vietnamese hotels' real data has implications for hotel management and describes the current situation, hence, can be a suggestion in updating the current Vietnamese hotel classification standards.

3. Research Methods and Materials

3.1. Problem Definition

In this study, both overall rating and aspect rating methods are proposed. Figure 1 below depicts the data collection process and hotel star rating based on big data analysis.

In this study, a number of definitions are used as follows:

- Definition 1. Reviews of guests
Reviews of guests are comments about the hotel that are written by guests who have experience, and is expressed as:

$$\mathfrak{R} = \{r_1, r_2, \dots, r_n\} \tag{1}$$

- Definition 2. Aspects of a hotel
Aspects of a hotel are the attributes of the hotel, and the services provided to guests by the hotel, and is expressed as:

$$\mathcal{A} = \{a_1, a_2, \dots, a_m\} \tag{2}$$

- Definition 3. Rating of hotel
Rating is the classification of the hotel into five different star levels, starting from 1 star to 5 stars.

- Definition 4. Overall rating
The overall rating is a measure of the overall quality of the hotel’s services and is assigned a score from 1star to 5 stars.

- Definition 5. Aspect rating
Aspect rating is a measure of the quality of each service in a hotel and is assigned a score from 1star to 5 stars. For example, Room is rated 3 stars, Location is rated 2 stars

- Definition 6. Region rating
Region rating is a measure of the quality of each service or overall service of all hotels in a region and classified according to the star level from 1–5. For example, Hotels in Danang city with an overall rating of 3.7 stars and hotels in Nhatrang city with an overall rating of 4.2 stars.

- Definition 7. Hotel ranking
Hotel ranking is its rank in the list of hotels in Vietnam or a certain region. For example, Lotte Hanoi Hotel is number #6 in the list of hotels in the Hanoi capital.

- Definition 8. Sentiment words
Sentiment words are the set of adverbs and adjectives that are extracted from reviews (\mathfrak{R}).

$$W = \{w_i, i = \overline{1-k}, w_i \text{ is adverb or adjective}\} \tag{3}$$

- Definition 9. Aspect grouping
Aspect grouping is a set of words with the same meaning.

$$a_1 = \{a_{11}, a_{12}, \dots, a_{1m}\} \tag{4}$$

For example:
Aspect price = {value, price}
Aspect location = {location, place}

3.2. Aspect Rating Based on Big Data Analysis

Hotel rating by aspects is difficult because most travelers only give stars for overall quality/level and write reviews. A few travelers make detailed reviews by aspects or give star ratings for different aspects. Therefore, when star rating by aspect, it is more appropriate to use rating-based scoring methods than supervised learning methods. This research is conducted in two phases, six steps.

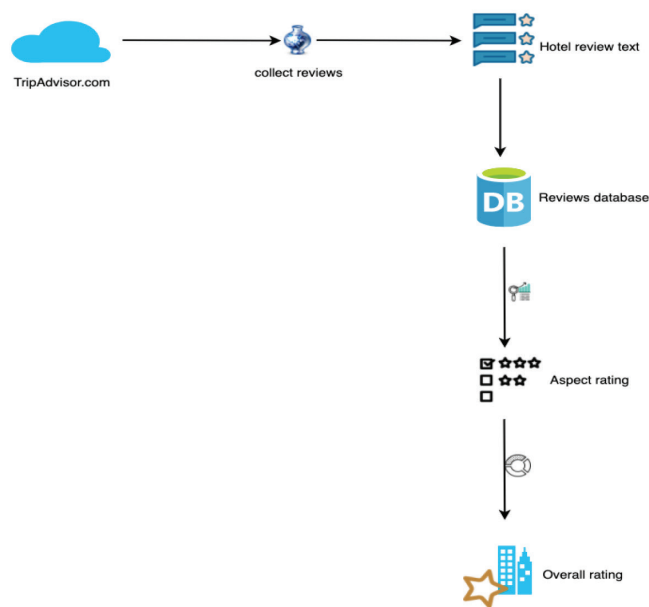


Figure 1: Hotel Star Rating Process

- Phase 1: Calculating sentiment value of adjectives and adverbs
 - Step 1: Building a dataset
Reviews are collected and pre-processed, and then segmented into sentences
 - Step 2: Classifying sentences
Sentences are classified into positive or negative sentences
 - Step 3: Calculating the weight of sentiment words
Using sentiment lexicon for extracting adjectives and adverbs, score them by the equation:

$$\text{Weight}(w_i) = \text{pos}(s_i) + \text{neg}(s_i)$$

In which:

$$\text{pos}(s_i) = \frac{\text{number of occurrences of } s_i \text{ in the positive sentences}}{\text{total of sentences in dataset}} \quad (5)$$

$$\text{neg}(s_i) = \frac{\text{number of occurrences of } s_i \text{ in the negative sentences}}{\text{total of sentences in dataset}} \quad (6)$$

- Phase 2: Aspect rating
 - Step 4: Aspect identification:
Using aspect grouping to extract aspects from a dataset.
 - Step 5: Calculating the weight of aspects:
Weight of an aspect is the product of the total sum of co-occurrence of aspect a_i with each sentiment word w_j and weight of sentiment word w_j .

$$\text{Weight}(a_i) = \sum_{j=1}^m w_j \times \text{co-occur}(w_j, a_i) \quad (7)$$

In which:

Weight(a_i): Weigh of aspect i^{th}
 w_j : Weight of sentiment word j^{th}
 co-occur(w_j, a_i): Number of co-occurrence of aspect i^{th} and sentiment word j^{th}

- Step 6: Aspect rating
Aspects are rated by the equation below:

$$\text{Rating}(a_i) = \frac{\text{Value}(a_i)}{\text{arg}_{\max}(\text{Value}(a))} \times 5 \quad (8)$$

Rating(a_i) is fixed by 1 when the result ≤ 1 .

Example:

Suppose that we have a value matrix, one dimension is aspects and the other is sentiment words in the sentiment set.

Aspect \ Sentiwords	Room	Location	Services	Staff
Great	10.8	4.98	4.02	7.44
Large	6.6	0.4	0	0
Near	4.8	42.24	0	0
....
Bad	2.04	0.69	2.91	1.98

The Value of each cell is the value calculated by:

$$\text{weight}(\text{great}) \times \text{co-occur}(\text{great}, \text{room})$$

And the weight of the aspect ‘room’ is calculated as:

$$\begin{aligned} \text{Weight}(\text{room}) = & \text{weight}(\text{great}) \times \text{co-occur}(\text{great}, \text{room}) \\ & + \text{weight}(\text{large}) \times \text{co-occur}(\text{larger}, \text{room}) \times \text{weight}(\text{near}) \\ & \times \text{co-occur}(\text{near}, \text{room}) \times \dots \times \text{weight}(\text{bad}) \times \text{co-occur}(\text{bad}, \text{room}) \end{aligned}$$

Aspect	Weight	Aspect Rating
Room	24.24	2.5
Location	48.31	5
Service	6.93	1
Staff	9.42	1

3.3. Hotel Overall Rating

The overall star rating of the hotel allows travelers to know about the quality of the whole service not only each aspect. The overall star rating of a hotel is calculated as the average of the aspect rating

$$\text{Rating}(\text{hotel}_k) = \frac{\sum_{j=1}^m (\text{rating}(a_j))}{m} \quad (9)$$

In which Rating(hotel_k) is the rating of hotel k and m is the total of aspects of hotel k .

The overall star rating for hotels in a region is calculated as the average of the overall ratings of hotels in this region:

$$\text{Rating}(\text{region}_s) = \frac{\sum_{k=1}^n (\text{rating}(\text{hotel}_k))}{n} \quad (10)$$

In which Rating(region_s) is the rating of region s and n is the number of hotels in region s .

4. Results

4.1. Data

In this study, the data was collected from the world’s most famous travel site, TripAdvisor. TripAdvisor uses an efficient rating algorithm and is always updated in real-time. In addition, TripAdvisor also warns when it detects fraud from hotel managers for increasing Internet rankings. The authors used a tool to crawl guest reviews of Vietnamese hotels from 4 to 5 stars. The dataset is saved as a .csv file with the information: reviewer name, review content, date of stay, and the number of review stars. The number of reviews includes 22,287 online reviews taken from the TripAdvisor website with 12 hotels from 4 stars to 5 stars in 04 cities in Vietnam including Hanoi, Ho Chi Minh City, Nha Trang, Danang (Table 1).

The data is pre-processed as follows:

- Travelers wrote reviews but do not give the hotel star rating, then these reviews are removed from the list.
- Travelers wrote reviews with symbols like <, and then gave stars, these cases are also removed from the list.

The data is used for 02 tasks: Overall rating and aspect rating. The data is prepared differently for each purpose.

4.2. Hotel Aspect Rating

Most travelers use star ratings for overall service, so aspect rating is done by using an unsupervised learning method (without labeled data). Because previous studies have not been conducted with Vietnamese hotel data, this study uses some of the aspects that are rated by TripAdvisor such as Location, Cleanliness, Services, and Value. Other aspects are not used for comparison.

The deviation is calculated by the formula:

$$\text{Deviation_aspect (a)} = \frac{\sum_{k=1}^c (\text{rating_a}(\text{hotel}_k) - \text{TRIPrating_a}(\text{hotel}_k))}{c} \times 100\% \quad (11)$$

In which Deviation_aspect(a) is the difference between this study’s rating and TripAdvisor rating of aspect a; rating_a(hotel_k) is the rating of aspect a for hotel k by the study method and TRIPrating_a(hotel_k) is the rating of aspect a for hotel k by TripAdvisor.

Results for 12 hotels in Vietnam with an average deviation of 0.38, when comparing ratings by this study method with the ratings by TripAdvisor (Table 2).

4.3. Hotel Overall Rating

In this study, the authors compare the results of this method with the hotel star rating of Vietnamese hotels in the

Table 1: List of Hotels for Data Crawling

ID Hotel	Name of Hotel	Location	Number of Reviews
1	Inter Continental Hanoi Westlake	Hanoi	2,839
2	Thang Loi Hotel	Hanoi	55
3	BaoSon Hotel	Hanoi	72
4	Rex Hotel	Ho Chi Minh	3,067
5	Hotel Nikko Saigon	Ho Chi Minh	4,488
6	Sofitel Saigon Plaza	Ho Chi Minh	3,011
7	Muong Thanh Luxury Da Nang Hotel	Danang	1,028
8	Intercontinental Danang	Danang	2,370
9	Sheraton Nha Trang Hotel and Spa	Nha Trang	2,312
10	Vinpearl Resort & Spa Nha Trang Bay	Nha Trang	1,756
11	Muong Thanh Holiday Hue Hotel	Hue	671
12	Huong Giang Hotel Resort & Spa	Hue	618
Total			22,287

database of the website <http://vietnamhotel.org.vn> and the hotel star rating of the TripAdvisor site. The simple deviation comparison uses the following formula

$$\text{Deviation (hotel in VN)} = \frac{\sum_{k=1}^c (\text{standard_RATED}(\text{hotel}_k) - \text{RATED}(\text{hotel}_k))}{c} \times 100\% \quad (12)$$

Standard_RATED(hotel_k) is the national standard rating of hotel *k*. *c* is the total number of hotels. The resulting deviation for 12 hotels is 56.4%. This value shows that online guests rated lower than Vietnamese standards. The difference is about 0.5 stars. The table below shows the difference between the actual rating and the rating on the TripAdvisor site and the analysis method of the study for a number of hotels in Vietnam (Table 3).

The results on the table show that the rating according to this method and the rating on the TripAdvisor site has less difference than the difference between the national standard rating and the rating on the TripAdvisor site.

6. Discussion and Conclusion

The trend of booking tours is growing and the number of travelers booking through online channels is increasing rapidly, showing the fierce competition between hotels and tourist destinations. Vietnam is a country rated by tourism organizations as the 3rd most attractive destination for

tourists in Southeast Asia and the top 10 most attractive destinations in the world. Tourism has become one of the major sectors within the economy of Vietnam in recent years, with a direct GDP contribution of over nine percent in 2019. In the same year, Vietnam also welcomed a record-high number of international visitor arrivals. The contribution of accommodation services to the tourism industry accounts for more than 70%, which shows the importance of the tourism sector which includes the hotel industry for the country’s economic development.

To continue to attract tourists, especially international tourists, the quality of services needs to be upgraded in line with guest requirements and global standards. Hotel classification organizations have considered online ratings as a suitable criterion for updating rating standards over time in order to rank hotels accordingly. However, it is also not possible to update faster than online rating because these systems work with huge amounts of data and always adjust ratings in real-time. Therefore, hotel classification organizations must always update the rating criteria as well as update the current status of hotel ratings in the latest way.

For Vietnam, hotel management focuses mainly on quality management in hotels and hotel promotion to attract tourists. Therefore, the hotel rating is meant to affirm the quality of the hotel and the quality commitment that guests can enjoy when using the hotel services. Moreover, the rating system indicates the prestige and reputation of hotels as well as the hotel classification system in Vietnam. Therefore, online star ratings for Vietnamese hotels can help hotel managers:

- Adjust criteria in the national standard system to match and harmonize with the user’s online star rating standard and asymptote (get closer and closer to) with the world’s set of star rating standards to gradually unify by global standards.
- Reduce unnecessary procedures, reduce processing time and reduce the cost of star level updating for hotels by Vietnam hotel classification organization. Gradually approach the one-stop electronic administrative process and return results more quickly.

Table 2: Comparing Aspect Rating Results with TripAdvisor

Aspect	Deviation
Location	0.32
Cleanliness	0.47
Service	0.525
Value	0.2
Average of all aspect	0.37875

Table 3: The Result of the Difference Between the Standard Rating, the Study Method, and the TripAdvisor Rating

Name of Hotel	Rated by Vietnam National Administration of Tourism	Rated by Our Method	Rated by TripAdvisor
Fortuna Hotel	4	4.2	4
Hanoi Lotte Hotel	5	4.7	5
BaoSon Hotel	4	3.6	3.5
Intercontinental NhaTrang	5	4.2	5
Rex Hotel	5	4.8	4

- It is possible to build an automatic system for hotel classification by the process applied in this study. This way, hotels can easily manage and strictly control the quality of service of the hotel.
- Hotel classification according to the world's general standards is also the basis for building a price suitable for Vietnamese hotels with international standards.
- Understand the real needs and opinions of tourists as well as the actual quality of the hotel so that hotels can improve and enhance their quality to attract more tourists, and contribute to the country's economic development.

This study has presented two methods to rate hotels in Vietnam including overall rating and aspect rating. Both methods have low errors and can be applied in rating Vietnamese hotels according to the trend of big data analysis of guest reviews as other countries in the world follow.

References

- Abrate, G., Quinton, S., & Pera, R. (2021). The relationship between price paid and hotel review ratings: Expectancy-disconfirmation or placebo effect? *Tourism Management*, 85(2), 104314. <https://doi.org/10.1016/j.tourman.2021.104314>
- Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). Aspect-based sentiment-oriented summarization of hotel reviews. *Procedia Computer Science*, 115, 563–571. <https://doi.org/10.1016/j.procs.2017.09.115>
- Al-Natour, S., & Turetken, O. (2020). A comparative assessment of sentiment analysis and star ratings for consumer reviews. *International Journal of Information Management*, 54(4), 102132. <https://doi.org/10.1016/j.ijinfomgt.2020.102132>
- Bigné, E., William, E., & Soria-Olivas, E. (2020). Similarity and consistency in hotel online ratings across platforms. *Journal of Travel Research*, 59(4), 742–758. <https://doi.org/10.1177/0047287519859705>
- Chang, Y. C., Ku, C. H., & Chen, C. H. (2019). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. *International Journal of Information Management*, 48(April 2017), 263–279. <https://doi.org/10.1016/j.ijinfomgt.2017.11.001>
- De Pelsmacker, P., van Tilburg, S., & Holthof, C. (2018). Digital marketing strategies, online reviews, and hotel performance. *International Journal of Hospitality Management*, 72(1), 47–55. <https://doi.org/10.1016/j.ijhm.2018.01.003>
- El-Said, O. A. (2020). Impact of online reviews on hotel booking intention: The moderating role of brand image, star category, and price. *Tourism Management Perspectives*, 33(10), 100604. <https://doi.org/10.1016/j.tmp.2019.100604>
- Goeltom, V. A. H., Kristiana, Y., Juliana, J., Bernato, I., & Pramono, R. (2020). The effect of service quality and value of five-star hotel services on behavioral intentions with the role of consumer satisfaction as mediator. *Journal of Asian Finance, Economics, and Business*, 7(11), 967–976. <https://doi.org/10.13106/jafeb.2020.vol7.no11.967>
- Hu, H. W., Chen, Y. L., & Hsu, P. T. (2016). A novel approach to rate and summarize online reviews according to user-specified aspects. *Journal of Electronic Commerce Research*, 17(2), 132–152. http://www.jecr.org/sites/default/files/17_2Paper3.pdf
- Jeon, S. W., Lee, H. J., Lee, H., & Cho, S. (2019). *Graph-based aspect extraction and rating classification of customer review data*. New York: Springer International Publishing. https://doi.org/10.1007/978-3-030-18590-9_13
- Kim, M., Lee, S. M., Choi, S., & Kim, S. Y. (2021). Impact of visual information on online consumer review behavior: Evidence from a hotel booking website. *Journal of Retailing and Consumer Services*, 60(2), 102494. <https://doi.org/10.1016/j.jretconser.2021.102494>
- Kim, S. H., & Yoo, B. K. (2021). Topics and sentiment analysis based on reviews of omnichannel retailing. *Journal of Distribution Science*, 19(4), 25–35. <https://doi.org/10.15722/jds.19.4.202104.25>
- Lai, C. H., & Hsu, C. Y. (2021). Rating prediction is based on the combination of review mining and user preference analysis. *Information Systems*, 99, 101742. <https://doi.org/10.1016/j.is.2021.101742>
- Le, Q. H., Nguyen, T. X. T., & Le, T. T. T. (2020). Customer satisfaction in hotel services: A case study of Thanh Hoa Province, Vietnam. *Journal of Asian Finance, Economics, and Business*, 7(10), 919–928. <https://doi.org/10.13106/jafeb.2020.vol7.no10.919>
- Lee, J. W. (2020). Big data strategies for government, society, and policy-making. *Journal of Asian Finance, Economics, and Business*, 7(7), 475–487. <https://doi.org/10.13106/jafeb.2020.vol7.no7.475>
- Luo, Y., & Tang, R. (2019). Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *International Journal of Hospitality Management*, 80(1), 144–154. <https://doi.org/10.1016/j.ijhm.2019.02.008>
- Manes, E., & Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: An empirical investigation of online hotel booking. *Journal of Business Research*, 85(12), 185–196. <https://doi.org/10.1016/j.jbusres.2017.12.019>
- Margaris, D., & Vassilakis, C. (2018). *Enhancing rating prediction quality through improving the accuracy of detection of shifts in rating practices*. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-57932-9_5
- Masiero, L., & Nicolau, J. L. (2016). Choice behavior in online hotel booking. *Tourism Economics*, 22(3), 671–678. <https://doi.org/10.5367/te.2015.0464>
- Park, Y. E., & Javed, Y. (2020). Insights discovery through hidden sentiment in big data: Evidence from Saudi Arabia's financial

- sector. *Journal of Asian Finance, Economics, and Business*, 7(6), 457–464. <https://doi.org/10.13106/jafeb.2020.vol7.no6.457>
- Rhee, H. T., & Yang, S. B. (2015). Does the hotel attribute importance differ by the hotel? Focusing on hotel star classifications and customers' overall ratings. *Computers in Human Behavior*, 50, 576–587. <https://doi.org/10.1016/j.chb.2015.02.069>
- Sánchez-Lozano, G., Pereira, L. N., & Chávez-Miranda, E. (2021). Big data hedonic pricing: Econometric insights into room rates' determinants by hotel category. *Tourism Management*, 85(4). <https://doi.org/10.1016/j.tourman.2021.104308>
- Sharma, H., Tandon, A., & Aggarwal, A. G. (2020). *Ranking hotels based on online hotel attribute ratings using neutrosophic AHP and stochastic dominance*. Singapore: Springer. https://doi.org/10.1007/978-981-15-1420-3_94
- Sokhin, T., Khodorchenko, M., & Butakov, N. (2020). *Unsupervised neural aspect extraction with related terms*. New York: Springer International Publishing. https://doi.org/10.1007/978-3-030-59082-6_6
- Wang, H., Lu, Y., & Zhai, C. (2010). A latent aspect rating analysis on review text data. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington DC, 25–28 July 2010 (pp. 783–792). <https://doi.org/10.1145/1835804.1835903783>
- Wang., Sun, J., & Wen, H. (2019). Tourism seasonality, online user rating and hotel price: A quantitative approach based on the hedonic price model. *International Journal of Hospitality Management*, 79(12), 140–147. <https://doi.org/10.1016/j.ijhm.2019.01.007>
- Xue, W., Li, T., & Rische, N. (2017). Aspect identification and rating inference for hotel reviews. *World Wide Web*, 20(1), 23–37. <https://doi.org/10.1007/s11280-016-0398-9>