www.gi.sanu.ac.rs, www.doiserbia.nb.rs
J. Geogr. Inst. Cvijic. 2024, 74(2), pp. 213–227

Original scientific paper

UDC: 911.3:338.487:659.1
https://doi.org/10.2298/IJGI2402213M

Received: March 2, 2024
Reviewed: June 6, 2024
Accepted: July 15, 2024

CLUSTERING FAMILY-FRIENDLY HOTELS' GUESTS TO DEVELOP TOURISM MARKETING STRATEGIES

Rizka Miladiah Ervianty[1], *Nasa Zata Dina*[2]*, *Sidarta Prassetyo*[1]

[1]Airlangga University, Faculty of Vocational Studies, Department of Business, Surabaya, Indonesia; e-mails: rizka.m.ervianty@vokasi.unair.ac.id; sidarta@vokasi.unair.ac.id
[2]Airlangga University, Faculty of Vocational Studies, Department of Engineering, Surabaya, Indonesia; e-mail: nasazatadina@vokasi.unair.ac.id

Abstract: An increasing number of guests in hotels evaluate the quality by reading online reviews. A deeper analysis of the attitude and behavior of the visitors is conducted to understand the experiences of guests, considering the diverse backgrounds and needs. This study aims to analyze the selection process of family-friendly hotels by guests, using available TripAdvisor online reviews, as well as for hotel management to better understand the comments left by guests and create more organized plans and policies. A model is devised that integrates clustering and Multi-Criteria Decision-Making-VIKOR (MCDM-VIKOR) method to prioritize the attributes of hotels based on the significance within each cluster of guests. Data is collected from online reviews of guests in family-friendly hotels in Indonesia. The features used for ranking preferences are the numerical ratings assigned to four attributes on the platform. These four features included "location", "cleanliness", "service", and "value". The results showed that "cleanliness" evolved as the most critical factor in the majority of segments for selecting family-friendly hotels. To further comprehend the behavioral trends of guests and assist in decision-making, this study proposed a model capable of analyzing online reviews and ratings provided by customers.

Keywords: customer satisfaction; hotels; sustainable tourism development; tourism and economic growth; tourism economics

1. Introduction

Tourism and hospitality sectors are profoundly impacting the existing economic system in Indonesia (Manomaivibool, 2015). Currently, industries tied to tourism account for 9.8% of the world's gross domestic product and approximately 7% of global commerce. The travel and tourism sector in 2022 supported 295 million jobs around the world, representing one in 11 of all the jobs globally and continued to rise to 320 million jobs by the end of 2023 (World Travel and Tourism Council, 2023). Technologies are progressively being integrated into the hospitality sector, resulting in the generation of a significant amount of data and enabling the comprehension of customer behavior (Fernandes et al., 2016; Ivanov et al., 2018). According to

*Corresponding author, e-mail: nasazatadina@vokasi.unair.ac.id

Kandampully et al. (2018), customer behavior is shaped by interactions between staff and customers, as well as the service provided as a product in the tourism and hospitality industries. To understand and comprehend the preferences and behavior of customers, organizing vast amounts of data obtained from online travel evaluations into discrete categories is crucial (Hsu et al., 2012; Jain, 2010). Therefore, extracting useful information from the vast amounts of unstructured data is essential for customers, companies, and policymakers.

Under the present smart tourism system, websites for online reviews have evolved into dynamic platforms for sharing and evaluating business performance (Kim et al., 2016). These platforms connect online bookings and available services in destinations, including attractions, transportation, accommodation, and purchases (Flavián et al., 2020; Gonzalez et al., 2020; Naik et al., 2019; Stankov et al., 2019; Wang et al., 2016). Guests predominantly prefer hotels recommended by others on the platforms, relying heavily on online reviews for decision-making (Nilashi et al., 2018). Online reviews are considered more reliable as the voluntarily supplied information by guests that might offer a more accurate description of the true emotions and experiences (Rhee & Yang, 2015). Numerous studies have focused on the information that may be obtained from these services in the hotel sector (Dina & Juniarta, 2022; Nilashi et al., 2018; Peng et al., 2018; Raguseo et al., 2017). However, less research has been conducted to identify trends in the attitude and behavior of guests regarding the features of hotels, based on a large dataset obtained from platforms regarding online reviews.

There are several types of travelers, for example couple, solo, business, and family travelers. Among those travelers, family travelers are regarded as one of the tourist kinds and make up a significant portion of the traveler population (Schänzel & Yeoman, 2015). The number of family travelers increased because of a multigenerational travel trend (Family Travel Association, 2023). Together, parents, grandparents, and kids are going on vacation to make treasured memories to strengthen family bonding through travelling. In order to provide a smooth and joyful experience for all parties involved, resorts, hotels, and destinations are adjusted to meet the varied demands and interests of families. To offer a perfect environment for families to deepen their relationships and forge lifetime connections, the accommodations are demanded to provide an ideal setting for families. To comprehend and meet the requirements and desires of the selected group of customers, market segmentation evaluation becomes crucial because it can affect all other marketing strategy decisions. Market segmentation, according to Lenti et al. (2021), is the process of dividing a market into discrete groups of consumers with varying needs, traits, or behaviors that may require different products or marketing strategies. The majority of existing studies on market segmentation relied on statistical methods and survey questionnaires to evaluate guests' preferences and behaviors, which is ineffective to identify subjective criteria.

Therefore, this study aims to fill this gap by incorporating clustering to identify market segmentation and at the same time identifying subjective criteria in each segment. Subjective criteria are evaluated to understand guests' preferences regarding hotel features, based on online reviews. This study proposes a model that clusters the guests based on preferences derived from online reviews and then applies multi-criteria decision making to rank the hotel features.

2. Methodology

2.1. Research procedure

The study proposed a method that included several steps, namely: (a) collecting data by scraping reviews directly from the TripAdvisor website, (b) preprocessing the textual data, (c) identifying significant features relevant to family-friendly hotels, (d) clustering the data into multiple clusters of guests using the machine learning KNN algorithm, and (e) ranking preferences based on each cluster of travelers using the Multi-Criteria Decision-Making-VIKOR (MCDM-VIKOR) method. Figure 1 shows the flow of the experiment, depicting the sequential process from data collection to analysis and interpretation.

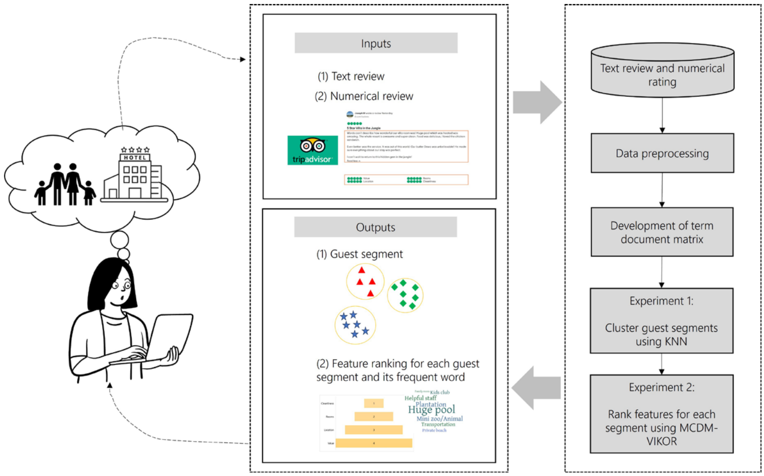

Figure 1. Proposed model.

2.2. Data collection

The proposed model is evaluated using a real data set. For this purpose, the data have been scraped from the TripAdvisor website. It is recognized as the largest platform with an extensive repository of reviews (Choi & Mai, 2018). TripAdvisor has high visibility and a comprehensive mechanism for customer evaluations (Farah & Ramadan, 2020), offering distinct advantages in terms of both quantity and quality of the reviews. These reviews contain unstructured and incoherent data that have to be transformed into machine learning language in order to become an input for our proposed model. The structure of data is organized for every customer's review from specific individuals that have stayed in the hotel previously.

Bali island is selected as the case of this study because it has been well known as one of the main international tourist destinations in Indonesia. In the first quarter of 2024, 45.08% international tourist visited Bali, while the rest (54.92%) visit other tourist destination in Indonesia (Badan Pusat Statistik Provinsi Bali, 2024). The tourism industry in Bali is growing gradually as a result of the growing number of visitors each year. As the tourist sector grows, so do the different businesses that support it. These include numerous hotels, villas, restaurants, travel agencies, art stores, and other businesses of a similar nature that are dispersed over Bali. More than two million tourists visited Bali in 2023–2024, and more than 300,000 foreigners travel to the island every month looking for lodging, hotels, and rental villas (Badan Pusat Statistik Provinsi Bali, 2024). Hence, Bali is a great start to study hotel market segmentation from a variety of different backgrounds.

The selected hotel locations are spread throughout Bali and hotel selection based on the reputation as family-friendly hotel that caters to family travelers. Ten selected hotels must be five-star hotels which have at least 2,000 reviews on TripAdvisor webpage. A higher number of reviews indicates a sign of a hotel's popularity. As a result of greater number of reviews, the influence of extreme evaluations is reduced, and a center trend may be found with more reviews (Jenq, 2019). We scraped 1,000 reviews from each hotel and the timeframe of reviews was limited to the past two years, from 2021 to 2023. It resulted in a total of 10,000 reviews obtained. The data scraped for each review include text review and star rating of each predefined feature as seen in Figure 2.

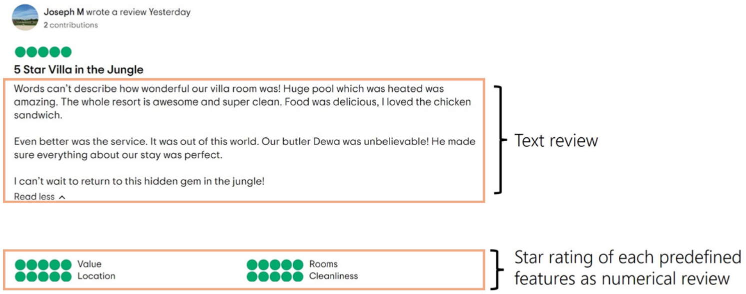

Figure 2. An example of a TripAdvisor review.

2.3. Data preprocessing

Data preprocessing is essential to convert raw data into a refined dataset. The conversion included checking for missing values, noisy data, and other inconsistencies before executing the information into the algorithm. The preprocessing phase in this study comprised tokenization, removal of stop words and special characters, transformation to lowercase, and stemming. An example of a customer review for each data preprocessing stage is described in Table 1.

Table 1. Description and example of data preprocessing

Steps	Description	Example
Tokenization	A step where the text data needs to be broken down into smaller units, such as words, for analysis. Tokenization divides text into meaningful units (Alawadh et al., 2023; Uysal & Gunal, 2014).	Before tokenization: "The whole Resort is awesome and super clean. I loved Chicken Sandwich." After tokenization: "The", "whole", "Resort", "is", "awesome", "and", "super", "clean", ".", "I", "loved", "Chicken", "Sandwich", "."
Stop word and special character removal	A stop word is a common word like "the", "is", and "and" that occur frequently, but convey little semantic meaning. Removing stop words can improve the efficiency of text analysis by reducing noise. In addition, special character removal is a step where all special character and punctuation marks (such as periods, commas, exclamation marks, emojis, etc.) are removed from the text to simplify it and focus on the words themselves (Uysal & Gunal, 2014).	Before removal: "The", "whole", "Resort", "is", "awesome", "and", "super", "clean", ".", "I", "loved", "Chicken", "Sandwich", "." After removal: "Resort", "awesome", "super", "clean", "loved", "Chicken", "Sandwich"

Table 1. Description and example of data preprocessing (*continued*)

Steps	Description	Example
Lower case transformation	A step where all letters in the text are converted to lowercase. This step is implemented so that the algorithm does not treat the same words differently in different situations (HaCohen-Kerner et al., 2020).	Before lower case transformation: "Resort", "awesome", "super", "clean", "loved", "Chicken", "Sandwich" After lower case transformation: "resort", "awesome", "super", "clean", "loved", "chicken", "sandwich"
Stemming	Stemming involves removing suffixes from words to obtain their base form while lemmatization involves converting words to their morphological base form (Dani et al., 2017; Lee et al., 2018).	Before stemming: "resort", "awesome", "super", "clean", "loved", "chicken", "sandwich" After stemming: "resort", "awesome", "super", "clean", "love", "chicken", "sandwich"

2.4. Development of term-document matrix

Following the completion of data preprocessing, the term-document matrix is developed by computing the term frequency-inverse document frequency (TF-IDF) from the text reviews (Nurcahyawati & Mustaffa, 2023). The importance of a term within a document relative to a collection of files is measured by TF-IDF, a widely used statistical method in natural language processing (Wan et al., 2024). The formula for TF-IDF facilitated the computation of these importance scores, as depicted in Equations 1–3, where *t* stands for "term" and *d* stands for "document":

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of } t \text{ in } d} \tag{1}$$

$$IDF(t) = \log \frac{N}{1 + df} \tag{2}$$

$$TF\text{-}IDF(t, d) = TF(t, d) \cdot IDF(t) \tag{3}$$

2.5. Experiment 1: Clustering guest segments using KNN

In the fields of tourism, business, and management, clustering algorithms are primarily used for: (a) market segmentation, which groups customers and uses targeted advertisements to attract more audience; (b) market basket analysis, which analyzes sales to determine which items customers buy together most frequently; and (c) social network analysis, which uses data to understand browsing behavior and recommends content to prospective customers. There are other clustering algorithms aside from KNN, for instance agglomerative clustering, DBSCAN, K-Means, and spectral clustering. However, in this study we utilized KNN because it has the ability to predict numerical and categorical variables (Kiyak et al., 2023). The other advantages of KNN include ease of interpretation and understanding of the results.

KNN in this study assists in market segmentation by clustering guests/customers with categorical attributes which are customer reviews scraped from TripAdvisor webpage. By identifying patterns and grouping based on star rating from predefined attributes and text customer reviews, KNN enables businesses to target specific customer segments with tailored marketing strategies (El Koufi et al., 2024; Zhang et al., 2020).

The KNN algorithm served as a well-known pattern recognition method, renowned for simplicity, clear interpretability, and fast computation. The underlying premise of the algorithm is that those points with comparable inputs would yield similar results. Initially, points are grouped into clusters in *n*-dimensional space based on shared characteristics, with *n* input parameters. Subsequently, a new point is selected, and the *k* points closest to the chosen points are analyzed to determine the predominant class. Consequently, the new point is allocated to the corresponding cluster (Hu et al., 2020).

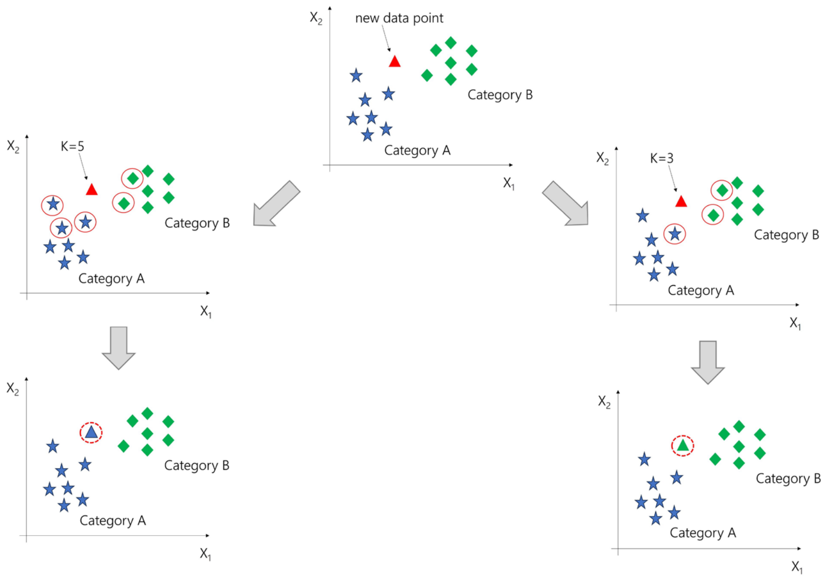

Figure 3. An Example of KNN.

Figure 3 shows an example of KNN where k = 3 and k = 5, showcasing instances of new points and the categorization based on the distances to cluster members. The performance of KNN is significantly influenced by the value of k. Determining the optimal value of k could be achieved through heuristic methods, contributing to the development of an efficient machine-learning model (Galdames, 2008; Mucherino et al., 2009; Papadopoulos & Manolopoulos, 2005). The following steps outlined the procedure followed by KNN:

1. The value of k is selected for training and testing the data;
2. The Euclidean distance between each point in the testing data and all points in the training is calculated. The distance values are recorded in a list and sorted subsequently, resulting in the first k points selected to assign a class to the testing data. The Euclidean distance is computed as shown in Equation 4, where x and y denote the two points between which the distance is calculated:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2} \tag{4}$$

2.6. Experiment 2: Ranking features according to guests' preferences using MCDM-VIKOR

In the second experiment, MCDM-VIKOR is applied after identifying the number of guest segments. MCDM-VIKOR ranks features (value, room, location, and cleanliness) based on guests' preferences for specific hotel features in each segment. MCDM-VIKOR is originally established with a focus on selecting and ranking alternative sets of competing criteria as described by Siregar et al. (2018) and Yazdani and Graeml (2014). Therefore, market segments evaluation and selection can be viewed as a MCDM-VIKOR problem in the presence of quantitative and qualitative criteria. In this study, we used quantitative criteria because the input of MCDM-VIKOR is specifically numerical review data.

The advantages of MCDM-VIKOR are that it can solve discrete decision problems and provide a solution that is the closest to the ideal (Faizi et al., 2020). MCDM-VIKOR method focuses on ranking and selecting from a set of alternatives and determines compromise solutions for a problem with conflicting criteria, which can help the decision makers to reach a final decision (Gao & Wei, 2018). The primary function of MCDM-VIKOR is to establish a multi-criteria ranking index based on a specific metric of proximity to the optimal solution (Opricovic, 2011). MCDM-VIKOR method included a series of steps designed to facilitate MCDM, namely:

1. Decision matrix is constructed by multiplying the attribute's occurrence and polarity value to generate keyword vectors, which are subsequently implemented to build decision matrix, as shown in Equation 5:

$$X = \begin{array}{c} W_1 \\ W_i \\ W_m \end{array} \begin{bmatrix} C_{11} & C_{1j} & C_{1n} \\ C_{i1} & C_{ij} & C_{in} \\ C_{m1} & C_{mj} & C_{mn} \end{bmatrix} \tag{5}$$

where the *i*-th and *j*-th options symbolized W_i and C_{ij} respectively.

2. The attribute occurrence (x) from each criterion is normalized to values between 0 and 1. Equation 6 shows the stated normalization and the weighting method used to determine the weight of each criterion (w_k). This is subsequently derived through Equation 7, where n_k (d) represent the occurrence values of the *k*-th criterion:

$$x^* = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{6}$$

$$w_k = \frac{n_k(d)}{\sum_{k=1}^{m} n_k(d)} \tag{7}$$

3. The positive (f_j^*) and negative (f_j^-) values are calculated using Equation 8:

$$f_j^* = \max_i f_{i,j} \text{ and } f_j^- = \min_i f_{i,j} \tag{8}$$

4. The new weighted decision matrix is estimated using Equation 9 to assign the weight (w_j) for a new decision matrix:

$$\text{new weighted decision matrix} = \frac{w_j(f_j^* - f_{ij})}{f_j^* - f_{ij}} \tag{9}$$

The individual regret (R_i) and group utility (S_i) values are calculated. The S_i and R_i are determined using Equations 10 and 11 respectively, where w_j represents the weight of the criterion:

$$S_i = \sum_{i=1}^{n} \frac{w_j\,(f_j^* - f_{ij})}{f_j^* - f_{ij}^-} \tag{10}$$

$$R_i = \max\left[\frac{w_j\,(f_j^* - f_{ij})}{f_j^* - f_{ij}^-}\right] \tag{11}$$

5. The index value (Q_i) is estimated by calculating Equation 12, where S^* and S^- represent the highest and lowest values, R^* and R^- signify the maximum and minimum values, while R_i and v are the index weight values:

$$Q_i = v\left[\frac{s_i - s^*}{s^- - s^*}\right] + (1 - v)\left[\frac{R_i - R^*}{R^- - R^*}\right] \tag{12}$$

6. The preference is ranked based on the obtained result. A lower quality corresponds to a higher index value (Q_i) and vice versa.

3. Results and discussion

In this section, the results of the proposed model are divided into two parts: (1) clustering guest segments using KNN and (2) ranking features according to guests' preferences using MCDM-VIKOR. The first part highlights hotel features that guests frequently mention. The second part focuses on gaining a deeper understanding of guest preferences and then ranking those hotel features.

3.1. Clustering guest segments using KNN

This study incorporated text reviews as supplementary information along with numerical reviews to enhance the explainability from each segment. Both text and numerical reviews are the input of the proposed model. At the first experiment, it resulted in the formation of eight clusters, which means there are eight distinct guest segments. Figure 4 shows the distinction among those eight segments, for instance, Segment A shows that the guests often mention the "huge pool" and "plantation" in their reviews, while Segment D discusses the availability of "animal/mini zoo" and "plantation" during their stay. Details about the eight guest segments from Segment A to Segment H can be seen in Figure 4. This study further shows the importance of some facilities mentioned by the guests through text review left on TripAdvisor website. Later, this analysis provides valuable insights for hotels actively engaged in family-friendly programs to develop more effective marketing strategies based on the key elements that appeal to each segment's guests.

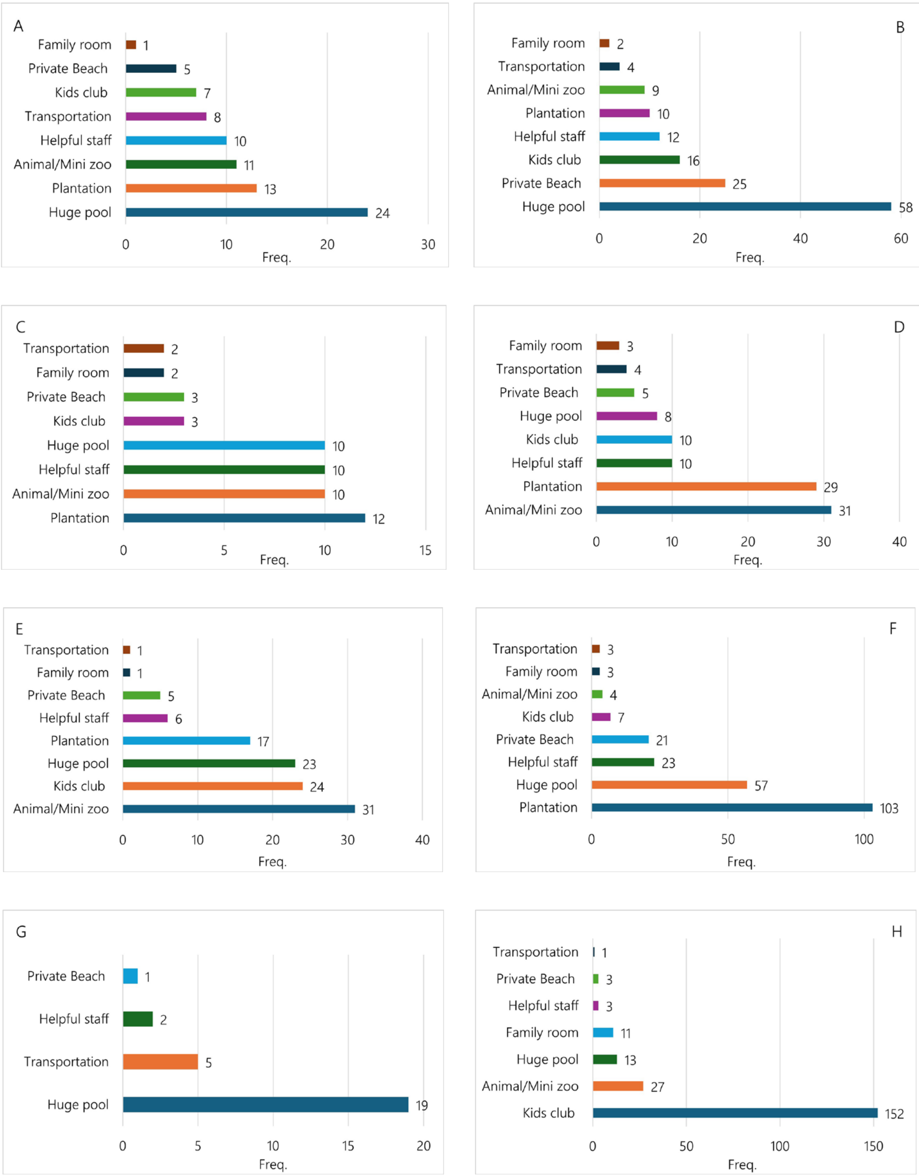

Figure 4. The frequency of keywords in text reviews on each segment.

Figure 5. Ranking of the features for each segment and its frequent words.

Guest segmentation is commonly recommended for marketing purposes, as serving guests with widely disparate wants, tastes, and aspirations can be challenging for businesses operating in a competitive marketplace (Hsu et al., 2012). In the travel and tourism industry, clustering methods are increasingly recognized as essential tools for developing appropriate policies in response to the diverse array of services, goods, and guests' preferences (Frochot & Morrison, 2000). Segmentation strategies have been applied to various aspects of guests' behavior and products/services benefits, including products and services, guests' spending behavior, pricing strategies, and tourism activities (Costa, 2013; Frochot, 2005; Guo et al., 2013; Legohérel & Wong, 2006; Rid et al., 2014).

Most studies use predetermined and fixed clusters incorporating empirical methods (Eusébio et al., 2017; Georgiadis & Tang, 2014; Rid et al., 2014). Despite TripAdvisor being a reputable source of travel information, there has been limited research on patterns within sizable datasets extracted from the websites using clustering methods. This study shows that from text reviews, we can extract important information about guests' opinions during their stay at the hotel and highlight their impression about certain facilities at the hotel. Other than that, the study also successfully clustered the guests based on their text reviews.

3.2. Ranking features based on guests' preferences using MCDM-VIKOR

MCDM-VIKOR is used to rank the features after implementing KNN to cluster the guests into eight segments, facilitating a deeper understanding of guests' preferences per segment. The four features provided on the platform are considered as predefined features because the feature is embedded in the platform without guests having to write a review text. Since these features are available on the platform, the guests convey their satisfaction in the form of numerical reviews or commonly known as star rating. Later, these numerical reviews are used as an input to MCDM-VIKOR. Figure 5 shows the ranking of the features on each segment. It indicates that each segment has its own preference when selecting family-friendly hotels.

In detail, the guests from each segment are described as follows:

- Segment A: Guests in the hotels within this segment prioritized rooms as the second most important attribute for family-friendly hotels after cleanliness. This segment also considered a huge pool, plantation, and animal/mini zoo as supporting facilities, as shown in Figure 5A;
- Segment B: Cleanliness and value are considered the most significant features in the cluster, surpassing location and rooms. Guests rated a huge pool, private beach, and kids club as important facilities, as seen in Figure 5B;
- Segment C: Similar to Segment B, cleanliness and value took precedence in the segment, as shown in Figure 5C. It shows guests suggesting that family-friendly hotels should provide a plantation, a huge pool, and an animal/mini zoo as main attractions;
- Segment D: Cleanliness and rooms received higher ratings than value and location in this cluster, as presented in Figure 5D. It showed that guests from the cluster advocated for mandatory animal/mini zoo and plantation facilities in family-friendly hotels;
- Segment E: This segment appreciates cleanliness and value above rooms and location as key features, as depicted in Figure 5E. Guests also view the availability of animal/mini zoo and kids club as supporting facilities in family-friendly hotels;

- Segment F: Cleanliness and rooms are identified as major features of family-friendly hotels in the cluster, as seen in Figure 5F. Many guests believe that plantations should be available in family-friendly hotels, as depicted in Figure 5F;
- Segment G: Figure 5G shows that this segment highly rates value and cleanliness as major features. This segment is the only segment that ranks value over cleanliness unlike the rest of the segments. Guests expected a huge pool and transportation availability in family-friendly hotels to enhance facilities as shown in Figure 5G; and
- Segment H: Cleanliness and rooms are considered major features, as shown in Figure 5H, with additional importance placed on the availability of the kids club and animal/mini zoo in the hotels.

In general, Figure 5 shows that "cleanliness" is reflected as a significant consideration among other features in seven of eight segments. Also, it implies that most guests that specifically choose family-friendly hotels are aware of the cleaning process of the hotel. Consequently, family-friendly hotels are recommended to prioritize housekeeping tasks and incorporate amenities that guests value.

Surprisingly, "location" is considered the least significant attribute by five out of eight segments as seen in Figure 5, despite the fact that strategic proximity to places of interest, such as city centers or tourist attractions, is a key element for many travelers. Previously, a study by Lin (2020) reported that family guests usually travel by their own cars. It might be the reason why "location" is not a concern for family guests.

Overall, hotel managers can address concerns related to low-ranked features in each segment by considering both short- and long-term strategic objectives. The proposed model has the potential to increase the number of positive reviews, reduce negative reviews from dissatisfied guests, and eventually improve competitive standing within the industry. Furthermore, positive reviews influence potential guests to decide which hotels are suitable for family travelers.

5. Conclusion

To sum up, this study identified eight guest segments based on text reviews using KNN and later guests' preferences of each segment are ranked based on numerical review using the MCDM-VIKOR. Seven of eight segments identified "cleanliness" as the most important attributes/features for hotel guests. It implies that the guests care about "cleanliness" over "value", "location", and "rooms" when deciding which family-friendly hotels to stay at. Additionally, we also adopted a text mining method to identify recurring terms in guests' reviews. The recurring terms from review texts help hotels to understand what other factors guests consider when choosing family-friendly hotels.

The contribution of this paper is to design an improved customer segmentation system of platform based on the KNN and MCDM-VIKOR model. On this basis, KNN is used to further subdivide the guest segments of the platform, and precise marketing strategies of the platform are formulated according to the segmentation results of the guest segments. In addition, this paper studies marketing strategies of e-commerce platforms from the perspective of precision marketing, and relevant studies have not been involved before.

While contributing to the field, this study also encountered several limitations that needed consideration. The case study focused solely on hotels in Bali, Indonesia, limiting the generalizability of the results. Conducting similar studies in comparable nations could yield more comprehensive and conclusive results regarding the clustering and behavior of guests.

Furthermore, this study only used the MCDM-VIKOR method to rank features of hotels within each cluster. Qualitative and quantitative methods are recommended for future studies to explore the factors influencing guests' preferences within each segment.

The first implication of this study is decision-making method-based market segmentation utilizing KNN and MCDM-VIKOR which are systematic and transparent approaches to decision-making. It helps marketing managers to make more informed and objective choices. Second, the integration of cutting-edge technologies holds promise for further enhancing the effectiveness and efficiency of decision-making processes in the dynamic landscape of tourism, business, and management.

References

Alawadh, H. M., Alabrah, A., Meraj, T., & Rauf, H. T. (2023). Discourse analysis based credibility checks to online reviews using deep learning based discourse markers. *Computer Speech & Language*, *78*, Article 101450. https://doi.org/10.1016/j.csl.2022.101450

Badan Pusat Statistik Provinsi Bali. (2024, June 3). *Perkembangan Pariwisata Provinsi Bali April 2024* [Bali Province Tourism Development Report April 2024]. Badan Pusat Statistik Provinsi Bali. https://bali.bps.go.id/pressrelease/2024/06/03/717894/perkembangan-pariwisata-provinsi-bali-april-2024.html

Choi, Y., & Mai, D. Q. (2018). The Sustainable Role of the E-Trust in the B2C E-Commerce of Vietnam. *Sustainability*, *10*(1), Article 291. https://doi.org/10.3390/su10010291

Costa, J. C. C. (2013). Price formation and market segmentation in seaside accommodations. *International Journal of Hospitality Management*, *33*, 446–455. https://doi.org/10.1016/j.ijhm.2012.11.004

Dani, H., Li, J., & Liu, H. (2017). Sentiment Informed Cyberbullying Detection in Social Media. In M. Ceci, J. Hollmén, Lj. Todorovski, C. Vens, & S. Džeroski (Eds.), *Machine Learning and Knowledge Discovery in Databases* (pp. 52–67). Springer. https://doi.org/10.1007/978-3-319-71249-9_4

Dina, N. Z., & Juniarta, N. (2022). Deriving customers preferences for hotels from unstructured data. *GeoJournal of Tourism and Geosites*, *43*(3), 872–877. https://doi.org/10.30892/gtg.43305-899

El Koufi, N., Belangour, A., & Sadiq, M. (2024). Toward a decision-making system based on artificial intelligence for precision marketing: A case study of Morocco. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1), Article 100250. https://doi.org/10.1016/j.joitmc.2024.100250

Eusébio, C., Carneiro, M. J., Kastenholz, E., Figueiredo, E., & da Silva, D. S. (2017). Who is consuming the countryside? An activity-based segmentation analysis of the domestic rural tourism market in Portugal. *Journal of Hospitality and Tourism Management*, *31*, 197–210. https://doi.org/10.1016/j.jhtm.2016.12.006

Faizi, S., Sałabun, W., Ullah, S., Rashid, T., & Wieckowski, J. (2020). A new method to support decision-making in an uncertain environment based on normalized interval-valued triangular fuzzy numbers and comet technique. *Symmetry*, *12*(4), Article 516. https://doi.org/10.3390/sym12040516

Family Travel Association. (2023). *US Family Travel Survey*. https://familytravel.org/wp-content/uploads/2023/10/FTA-Survey2023-Report.pdf

Farah, M. F., & Ramadan, Z. B. (2020). Viability of Amazon's driven innovations targeting shoppers' impulsiveness. *Journal of Retailing and Consumer Services*, *53*, Article 101973. https://doi.org/10.1016/j.jretconser.2019.101973

Fernandes, S., Belo, A., & Castela, G. (2016). Social network enterprise behaviors and patterns in SMEs: Lessons from a Portuguese local community centered around the tourism industry. *Technology in Society*, *44*, 15–22. https://doi.org/10.1016/j.techsoc.2015.11.004

Flavián, C., Guinaliu, M., & Lu, Y. (2020). Mobile payments adoption – introducing mindfulness to better understand consumer behavior. *International Journal of Bank Marketing*, *38*(7), 1575–1599. https://doi.org/10.1108/IJBM-01-2020-0039

Frochot, I. (2005). A benefit segmentation of tourists in rural areas: a Scottish perspective. *Tourism Management*, *26*(3), 335–346. https://doi.org/10.1016/j.tourman.2003.11.016

Frochot, I., & Morrison, A. M. (2000). Benefit Segmentation: A review of its applications to travel and tourism research. *Journal of Travel & Tourism Marketing*, *9*(4), 21–45. https://doi.org/10.1300/J073v09n04_02

Galdames, P. (2008). *Managing continuous k-nearest neighbor queries in mobile peer-to-peer networks* (Master's thesis). Retrieved from https://dr.lib.iastate.edu/bitstreams/f0ed0bbd-ea26-4018-bcbd-802e82a041f6/download

Gao, H., & Wei, G.-W. (2018). Multiple attribute decision making based on interval-valued Pythagorean uncertain linguistic aggregation operators. *International Journal of Knowledge-based and Intelligent Engineering Systems*, *22*(1), 59–81. https://doi.org/10.3233/KES-180375

Georgiadis, G., & Tang, C. S. (2014). Optimal reservation policies and market segmentation. *International Journal of Production Economics*, *154*, 81–99. https://doi.org/10.1016/j.ijpe.2014.04.021

Gonzalez, R. A., Ferro, R. E., & Liberona, D. (2020). Government and governance in intelligent cities, smart transportation study case in Bogotá Colombia. *Ain Shams Engineering Journal*, *11*(1), 25–34. https://doi.org/10.1016/j.asej.2019.05.002

Guo, X., Ling, L., Yang, C., Li, Z., & Liang, L. (2013). Optimal pricing strategy based on market segmentation for service products using online reservation systems: An application to hotel rooms. *International Journal of Hospitality Management*, *35*, 274–281. https://doi.org/10.1016/j.ijhm.2013.07.001

HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PloS ONE*, *15*(5), Article e0232525. https://doi.org/10.1371/journal.pone.0232525

Hsu, F.-M., Lu, L.-P., & Lin, C.-M. (2012). Segmenting customers by transaction data with concept hierarchy. *Expert Systems with Applications*, *39*(6), 6221–6228. https://doi.org/10.1016/j.eswa.2011.12.005

Hu, J., Peng, H., Wang, J., & Yu, W. (2020). kNN-P: A kNN classifier optimized by P systems. *Theoretical Computer Science*, *817*, 55–65. https://doi.org/10.1016/j.tcs.2020.01.001

Ivanov, S., Webster, C., & Garenko, A. (2018). Young Russian adults' attitudes towards the potential use of robots in hotels. *Technology in Society*, *55*, 24–32. https://doi.org/10.1016/j.techsoc.2018.06.004

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jenq, S.-N. (2019). Exploring the Impact of a Hotel's Rating and Number of Reviews through Online Transactions. *Open Access Library Journal*, *6*(5), Article e5401. https://doi.org/10.4236/oalib.1105401

Kandampully, J., Zhang, T. C., & Jaakkola, E. (2018). Customer experience management in hospitality: A literature synthesis, new understanding and research agenda. *International Journal of Contemporary Hospitality Management*, *30*(1), 21–56. https://doi.org/10.1108/IJCHM-10-2015-0549

Kim, J.-Y., Hlee, S., & Joun, Y. (2016). Green practices of the hotel industry: Analysis through the windows of smart tourism system. *International Journal of Information Management*, *36*(6), 1340–1349. https://doi.org/10.1016/j.ijinfomgt.2016.05.001

Kiyak, E. O., Ghasemkhani, B., & Birant, D. (2023). High-Level K-Nearest Neighbors (HLKNN): A Supervised Machine Learning Model for Classification Analysis. *Electronics*, *12*(18), Article 3828. https://doi.org/10.3390/electronics12183828

Lee, P.-J., Hu, Y.-H., Chen, K., Tarn, J. M., & Cheng, L.-E. (2018). Cyberbullying Detection on Social Network Services. *PACIS 2018 Proceeding*. https://aisel.aisnet.org/pacis2018/61

Legohérel, P., & Wong, K. K. F. (2006). Market Segmentation in the Tourism Industry and Consumers' Spending: What About Direct Expenditures? *Journal of Travel & Tourism Marketing*, *20*(2), 15–30. https://doi.org/10.1300/J073v20n02_02

Lenti, P., Kottmair, S., Stock, S., Shukri, A., & Müller, D. (2021). Predictive modeling to identify potential participants of a disease management program hypertension. *Expert Review of Pharmacoeconomics & Outcomes Research*, *21*(2), 307–314. https://doi.org/10.1080/14737167.2020.1780919

Lin, P. M. C. (2020). Is Airbnb a Good Choice for Family Travel? *Journal of China Tourism Research*, *16*(1), 140–157. https://doi.org/10.1080/19388160.2018.1551167

Manomaivibool, P. (2015). Wasteful tourism in developing economy? A present situation and sustainable scenarios. *Resources, Conservation and Recycling*, *103*, 69–76. https://doi.org/10.1016/j.resconrec.2015.07.020

Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). *Data Mining in Agriculture*. Springer. https://doi.org/10.1007/978-0-387-88615-2

Naik, M. B., Kumar, P., & Majhi, S. (2019). Smart public transportation network expansion and its interaction with the grid. *International Journal of Electrical Power & Energy Systems*, *105*, 365–380. https://doi.org/10.1016/j.ijepes.2018.08.009

Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E., & Alizadeh, A. (2018). Travelers decision making using online review in social network sites: A case on TripAdvisor. *Journal of Computational Science*, *28*, 168–179. https://doi.org/10.1016/j.jocs.2018.09.006

Nurcahyawati, V., & Mustaffa, Z. (2023). Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation. *Journal of Information Systems Engineering and Business Intelligence*, *9*(1), 108–118. https://doi.org/10.20473/jisebi.9.1.108-118

Opricovic, S. (2011). Fuzzy VIKOR with an application to water resources planning. *Expert Systems with Applications*, *38*(10), 12983–12990. https://doi.org/10.1016/j.eswa.2011.04.097

Papadopoulos, A. N., & Manolopoulos, Y. (2005). *Nearest Neighbor Search: A Database Perspective*. Springer. https://doi.org/10.1007/0-387-27544-4

Peng, H.-G., Zhang, H.-Y., & Wang, J.-Q. (2018). Cloud decision support model for selecting hotels on TripAdvisor.com with probabilistic linguistic information. *International Journal of Hospitality Management*, *68*, 124–138. https://doi.org/10.1016/j.ijhm.2017.10.001

Raguseo, E., Neirotti, P., & Paolucci, E. (2017). How small hotels can drive value their way in infomediation. The case of 'Italian hotels vs. OTAs and TripAdvisor'. *Information & Management*, *54*(6), 745–756. https://doi.org/10.1016/j.im.2016.12.002

Rhee, H. T., & Yang, S.-B. (2015). Does hotel attribute importance differ by 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

Rid, W., Ezeuduji, I. O., & Pröbstl-Haider, U. (2014). Segmentation by motivation for rural tourism activities in The Gambia. *Tourism Management*, *40*, 102–116. https://doi.org/10.1016/j.tourman.2013.05.006

Schänzel, H. A., & Yeoman, I. (2015). Trends in family tourism. *Journal of Tourism Futures, 1*(2), 141–147. https://doi.org/10.1108/JTF-12-2014-0006

Siregar, D., Nurdiyanto, H., Sriadhi, S., Suita, D., Khair, U., Rahim, R., Napitupulu, D., Fauzi, A., Hasibuan, A., Mesran, M., & Siahaan, A. P. U. (2018). Multi-Attribute Decision Making with VIKOR Method for Any Purpose Decision. *Journal of Physics: Conference Series*, *1019*, Article 012034. https://doi.org/10.1088/1742-6596/1019/1/012034

Stankov, U., Filimonau, V., & Slivar, I. (2019). Calm ICT design in hotels: A critical review of applications and implications. *International Journal of Hospitality Management*, *82*, 298–307. https://doi.org/10.1016/j.ijhm.2018.10.012

Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, *50*(1), 104–112. https://doi.org/10.1016/j.ipm.2013.08.006

Wan, Q., Xu, X., & Han, J. (2024). A dimensionality reduction method for large-scale group decision-making using TF-IDF feature similarity and information loss entropy. *Applied Soft Computing*, *150*, Article 111039. https://doi.org/10.1016/j.asoc.2023.111039

Wang, Y.-S., Li, H.-T., Li, C.-R., & Zhang, D.-Z. (2016). Factors affecting hotels' adoption of mobile reservation systems: A technology-organization-environment framework. *Tourism Management*, *53*, 163–172. https://doi.org/10.1016/j.tourman.2015.09.021

World Travel and Tourism Council. (2023). Economic impact 2023. The economic impact of global travel & tourism. In N. Jus, J. Mitcham, & C. Tsering (Eds.), *Travel & Tourism Economic Impact 2023* (pp. 1–21). https://assets-global.website-files.com/6329bc97af73223b575983ac/648b1a25bc78c3018ad52edc_EIR2023-World.pdf

Yazdani, M., & Graeml, F. R. (2014). VIKOR and its Applications: A State-of-the-Art Survey. *International Journal of Strategic Decision Sciences*, *5*(2), 56–83. https://doi.org/10.4018/ijsds.2014040105

Zhang, S., Liao, P., Ye, H.-Q., & Zhou, Z. (2020). Multiple resource allocation for precision marketing. *Journal of Physics: Conference Series, 1592*, Article 012034. https://doi.org/10.1088/1742-6596/1592/1/012034