applied sciences

MDPI

Article

Transforming Customer Digital Footprints into Decision Enablers in Hospitality

Achini Adikari [1,*], Su Nguyen [2], Rashmika Nawaratne [1], Daswin De Silva [1] and Damminda Alahakoon [1,*]

[1] Research Centre for Data Analytics and Cognition, La Trobe University, Bundoora, VIC 3086, Australia
[2] Department of Accounting, Information Systems & Supply Chain, RMIT University, Melbourne, VIC 3000, Australia
* Correspondence: a.adikari@latrobe.edu.au (A.A.); d.alahakoon@latrobe.edu.au (D.A.)

Abstract: The proliferation of online hotel review platforms has prompted decision-makers in the hospitality sector to acknowledge the significance of extracting valuable information from this vast source. While contemporary research has primarily focused on extracting sentiment and discussion topics from online reviews, the transformative potential of such insights remains largely untapped. In this paper, we propose an approach that leverages Natural Language Processing (NLP) techniques to convert unstructured textual reviews into a quantifiable and structured representation of emotions and hotel aspects. Building upon this derived representation, we conducted a segmentation analysis to gauge distinct emotion and concern-based profiles of customers, as well as profiles of hotels with similar customer emotions using a self-organizing unsupervised algorithm. We demonstrated the practicality of our approach using 22,450 online reviews collected from 44 hotels. The insights garnered from emotion analysis and review segmentation facilitate the development of targeted customer management strategies and informed decision-making.

Keywords: social media analytics; GuidedLDA; emotion modelling; unsupervised learning; decision support; online reviews

check for updates

Citation: Adikari, A.; Nguyen, S.; Nawaratne, R.; De Silva, D.; Alahakoon, D. Transforming Customer Digital Footprints into Decision Enablers in Hospitality. *Appl. Sci.* **2024**, *14*, 3114. https://doi.org/10.3390/app14073114

Academic Editors: Carlos J. Costa and Manuela Aparicio

Received: 8 March 2024
Revised: 29 March 2024
Accepted: 4 April 2024
Published: 8 April 2024

1. Introduction

The expression of thoughts and experiences by customers has undergone a significant transformation since the advent of the Internet, thereby changing the technological landscape of the tourism and hospitality industry [1]. In this era of e-commerce, consumers readily share their experiences on social media platforms, thus creating a digital trail of their thoughts and emotions, that are projected to a wider audience. Such online interactions, opinions and emotions expressed via social media posts collectively generate a representation of an individual in the digital space. These traces or 'footprints' left by customers on the internet are known as 'digital footprints' [2]. Compared to traditional methods of conducting customer surveys, digital footprints allow organizations to tap into an abundance of customer data encapsulating their opinions and emotions.

To this end, online reviews have become one of the most valuable sources of information in the digital age, providing an opportunity for customers to share their post-purchase experiences [3]. This information is useful for potential customers as well as for organizations to manoeuvre their decision-making [4]. An online review is an unstructured piece of text that encompasses customers' experiences and emotions in an unorganized manner. Given the importance of such information, topic and sentiment analyses using online hotel reviews have gained the attention of researchers and many extant studies have focused on detecting hotel aspects and identifying emotions and sentiments from hotel reviews [5–8]. However, these approaches lack a more in-depth exploration on utilising customer emotions as insights for decision-making. Since an online review cannot be easily quantified due to its textual form, it is essential to transform the expressed emotions and the mentions of hotel aspects into a structured and quantifiable format that can be effectively applied for decision-making and analysis.

Thus, the primary objective of this research is to generate a structured emotion–hotel aspect representation from unstructured online reviews to support decision-making. We adopt Plutchik's emotion model, which offers a more nuanced understanding with its eight emotions compared to broader sentiment categories [9]. Next, we showcase the use of derived emotion representations to conduct a segmentation analysis that illustrates different profiles of customer reviews based on the diverse emotions expressed. Such segmentation is particularly advantageous for the hospitality industry, given its diverse customer base expressing a range of emotions and concerns in reviews [5,10].

To achieve this, we employ Natural Language Processing (NLP) to extract emotions and hotel aspects and the Growing Self-Organizing Maps (GSOMs) algorithm [11], an unsupervised clustering technique, to generate profiles based on similar characteristics in data. Additionally, the GSOM algorithm serves as a visualization tool, aiding in the identification of these distinctive profiles.

To demonstrate the practicality and applicability of using this approach in real-world scenarios, we have used 22,450 online reviews from 44 hotels from the Tripadvisor platform over three years. The rest of this paper is organized as follows. Section 2 discusses the role of online reviews in the tourism industry and outlines different approaches taken for emotion modelling and profiling. Section 3 provides the methodology to develop the emotion profile from online reviews, while Section 4 demonstrates the results of the analysis. The paper concludes with a discussion of the findings and potential implications, followed by future enhancements.

2. Related Work

2.1. Digital Footprints in the Tourism Industry

Customer opinions are now more pronounced as they are broadcasted via social networks to a wider audience. It is stated that eWOM (electronic Word-Of-Mouth) has become dominant among customers when making purchase decisions compared to traditional means of information sharing [12]. Moreover, exposure to online reviews has been shown to improve the average probability of customer bookings, thereby highlighting the crucial role of sales [13]. Inevitably, this has led the hotel industry to consider online reviews as a vital source of information related to customer experience [14].

As online reviews encapsulate customer emotions, the valence of the reviews is deemed a crucial source of information [15]. Valence refers to the degree of positivity expressed in the online review. Sentiment analysis has emerged as a vital analytics task in this space, to derive the positive and negative aspects of online reviews [7,12]. The emotions expressed in online reviews can improve or harm the reputation of the hotel and impact its performance [16]. This can significantly influence the purchase decisions of future customers [17–19]. Therefore, the business value that can be gained by analysing online reviews is significant [20].

2.2. Sentiment Analysis vs. Deep Emotion Modelling

Contemporary research studies have advanced beyond sentiment as basic polarity to detect extended categories of emotion. Unsupervised approaches such as word embeddings have been used recently for emotion detection and to understand the distributed representations of words based on syntactic and semantic word relationships [21–23]. Several studies have published emotion lexicons that combine general expressions for each emotion that can be used to categorize texts [12,24,25]. Moreover, many studies have developed supervised machine learning models to predict the category of emotions on Twitter and online reviews related to tourism and hospitality that depend on previously annotated data [6,26]. In this space, deep learning models and language models have been extensively used to extract sentiments from online hotel reviews [27,28]. However, most of the approaches have focused on only positive and negative sentiments and are thus incapable of exploiting the richer meanings of deeper human emotions ingrained in the reviews. For example, the emotions 'anger' and 'sad' are classified as 'negative sentiments', although they convey vastly different feelings. R. Li et al., in their recent research, outline that sentiments accentuate only the broad category of emotions as positive, negative and neutral; therefore, future studies should pursue creating a more granular form of emotion representation [28]. The identification of the emotions of customers at a granular level from their online reviews is crucial to properly exploit the power of social media [29]. These models rely on previously annotated data to train the machine learning model and lack deeper exploration of the derived emotions which could be used for decision-making.

A significant improvement in the current state of the art using sentiment analysis in online reviews is the capture and modelling of deep emotions with quantifiable levels of intensity. This facilitates underlying (even subtle) variations in opinions and preferences in customer experiences to be represented and highlighted, which will provide deeper insights into the causality of reviews in quantified form.

2.3. Emotion Modelling for Analytics

Due to the significant influence of customers' emotions on their post-purchase behaviour, emotional and affective factors are gaining recognition as key elements in the decision-making processes [30]. Although various studies are attempting to capture topics and emotions from hotel reviews, the number of studies to show the potential of transforming these data into analytics processes is limited. A key analysis that is underexplored in the current literature is a completely data-driven grouping (profiling) of customers based on their emotions and concerns. More traditional approaches focus on clustering customers based on their demographic and survey responses which do not accentuate their experience with the hotel. Several market segmentation methods classify consumers based on criteria such as socioeconomic, demographic and psychographic factors. However, previous research advocates that these segmentation bases are of limited value in exploring consumer behaviour, given that influential factors such as customers' emotions and affect are not utilised [31]. One research study related to psychology and marketing has indicated that emotions could be used as the basis for market segmentation [32]. On this premise, we propose a completely automated, data-driven approach to create profiles of reviews, customers and hotels based on the hotel's digital data.

3. Materials and Methods

This study introduces an approach that harnesses NLP and unsupervised machine learning techniques to develop a structured representation of customer emotions extracted from online reviews. The platform's high-level architecture consists of three core components: data collection and integration, a machine learning-based social media analytics module, and an outcome visualization module. These components seamlessly collaborate to process the data, extract valuable insights and visually present them in an informative manner. The proposed methodology is illustrated in Figure 1 for reference.

3.1. Data Collection and Processing

The proposed methodology is designed to take textual inputs that contain customer opinions. These data could be from multiple social media channels such as online reviews, blogs and posts and comments from social networks such as Twitter, Facebook or Instagram. In this study, we focused on online reviews from Tripadvisor, which is a popular online hotel booking platform. These reviews were used as the main data source due to their descriptive content and rich information on customer experiences. We focused on 4–5-star hotels given the large volume of reviews available compared to lower-rated hotels. The selected hotels were urban hotels in the USA. We used an automated Python programme to extract publicly available data from the Tripadvisor platform, which resulted in 22,450 reviews across three years. The dataset comprised the textual content of the review, review title, review rating, date and time and review ID. Following the data collection, data regarding the customers were redacted for privacy concerns. Afterwards, pre-processing was performed to remove redundant characters, weblinks and symbols. The processed data were then fed into the platform for analysis.

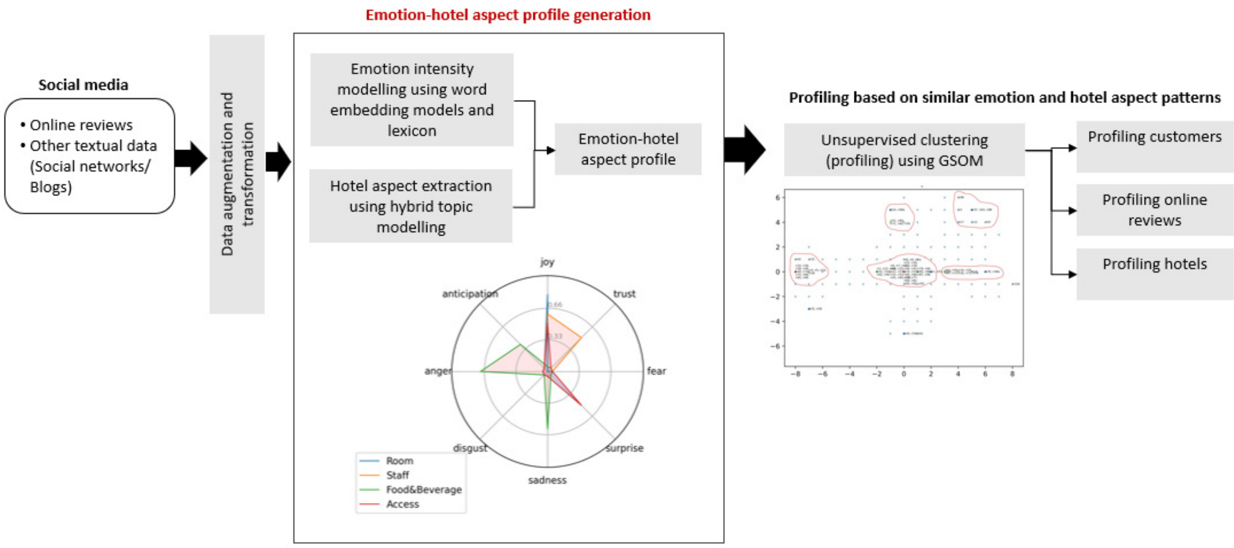


Figure 1. The architecture of the proposed methodology.

3.2. Emotion Modelling

The emotions articulated by customers enable a comprehensive understanding of their overall experience; however, limited attention has been given to the extraction of emotions at a granular level related to different individual hotel aspects in previous studies. A combination of NLP and machine learning techniques were applied to extract different emotional expressions from online reviews for this research.

The emotion modelling was based on the highly cited and endorsed psychological emotion model proposed by Plutchik [33] which demonstrates eight basic emotion states, "Joy", "Anger", "Sad", "Surprise", "Fear", "Anticipation", "Trust" and "Disgust". Initially, a generic list of emotion words was used to build the dictionary with the use of a published emotion-lexicon dictionary [34]. A standard series of text pre-processing steps of stemming, lemmatization and stop word removal was carried out to clean the textual review. Next, a keyword extraction method was developed to identify emotional expressions presented in social media content [35]. For this purpose, regular expression-based text mining was conducted to handle the negations of emotional expressions.

To further enrich the initial emotion terms and to contextually align the emotion expressions with the hospitality industry, we used a customisation process for the emotion lexicon as noisy or incomplete vocabularies could lead to errors in the emotion extraction. For this purpose, a word embedding model was trained using the Word2Vec approach [36], which can capture the context of a word in a document and identify the semantic and syntactic similarity with other words. The Word2Vec model was trained based on 250,000 online reviews related to tourism [37] and was then used to generate similar words for each emotion based on hotel review data. The similar words were generated based on the neighbourhood of the seed word (emotion). The generated terms were reviewed by the authors to determine the applicability. Using this expanded vocabulary, it was possible to capture indirect, contextual emotional expressions which are otherwise challenging to capture from a dictionary-based approach. A few examples of contextual emotion expressions are noted below:

- "Joy" → {'tasty', 'fragrant', 'serenity'};
- "Anger" → {'dirty', 'rude', 'noisy'};
- "Disgust" → {'smelly', 'rotten', 'filthy'}.

The following Table 1 represents an excerpt from the tourism-related emotion dictionary which was created to conduct the emotion analysis.

Table 1. Excerpt from the emotion vocabulary developed for emotion extraction.

Emotion	Examples of Emotional Expressions
Joy	{"happy", "enjoyed", "delighted", "beautiful", "tasty", . . .}
Anger	{"angry", "dirty", "disturbed", "irritated", "noisy", . . .}
Sad	{"sad", "depressing", "miserable", "poor", "worried", . . .}
Surprise	{"surprised", "amazed", "shocked", "astonished", "stunned", . . .}
Fear	{"scared", "dangerous", "fear", "afraid", "horrified", . . .}
Trust	{"ensure", "trust", "make sure", "belief", "certain", . . .}
Anticipation	{"look forward", "anticipate", "expect", "predict", "hope", . . .}
Disgust	{"smelly", "disgusting", "rotten", "smell", "filthy", . . .}

The following Figure 2 illustrates the high-level process of the emotion extraction engine.

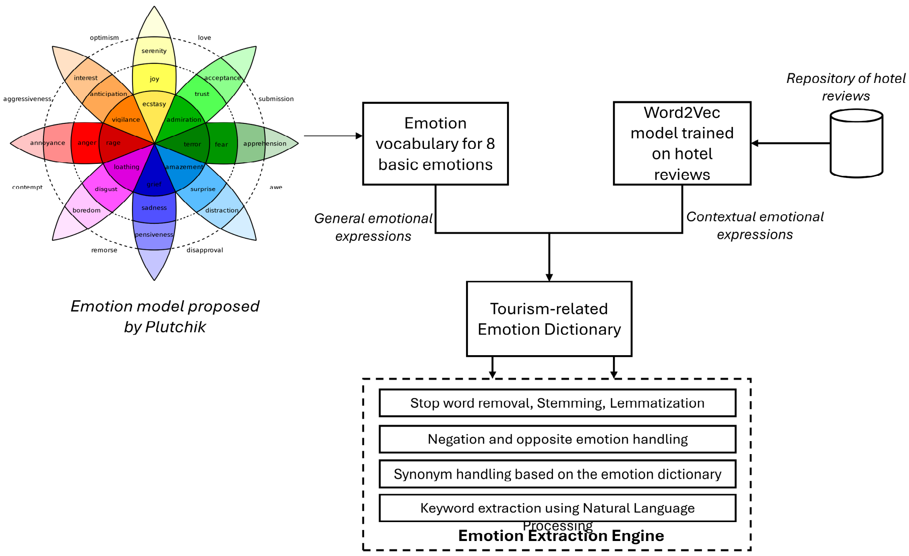


Figure 2. The high-level process of emotion extraction. The emotion model is from [38].

The extracted emotions were then aligned with hotel aspects to create informed insights related to customer experience.

3.2.1. Hotel Aspect Extraction Using Hybrid Topic Modelling

The customer's hotel experience is closely linked to various key aspects of the hotel, as outlined in the World Tourism Organization's (WTO) hotel classification report [39]. To identify these points of interest, we created an aspect model that encompasses essential facets such as "Room", "Bathroom", "Front Desk", "Location/Access", "Public Areas", "Food & Beverage", "Exterior" and "Staff". Contemporary research studies have used the aspect-based online review categorization to understand different topics using field-specific lexicon [40], frequency-based aspect topics [41] and syntactic features of the text [42]. In this study, we extract hotel mentions using a hybrid approach, combining semi-supervised topic modelling and keyword extraction through NLP techniques to enhance accuracy.

To initiate the keyword extraction process, we developed a vocabulary specific to each aspect, consisting of various properties associated with them. By expanding this vocabulary using the trained Word2Vec model, we generated a finely tuned aspect dictionary, including synonyms for each aspect. Additionally, to systematically widen the scope of aspect extraction, we incorporated a semi-supervised topic modelling approach, employing Guided Latent Dirichlet Allocation (GuidedLDA) [43]. GuidedLDA utilizes the initial seed word list, comprising the previously extracted keywords, to categorize each review into potential topics. Figure 3 illustrates the high-level process of aspect extraction applied in this study.

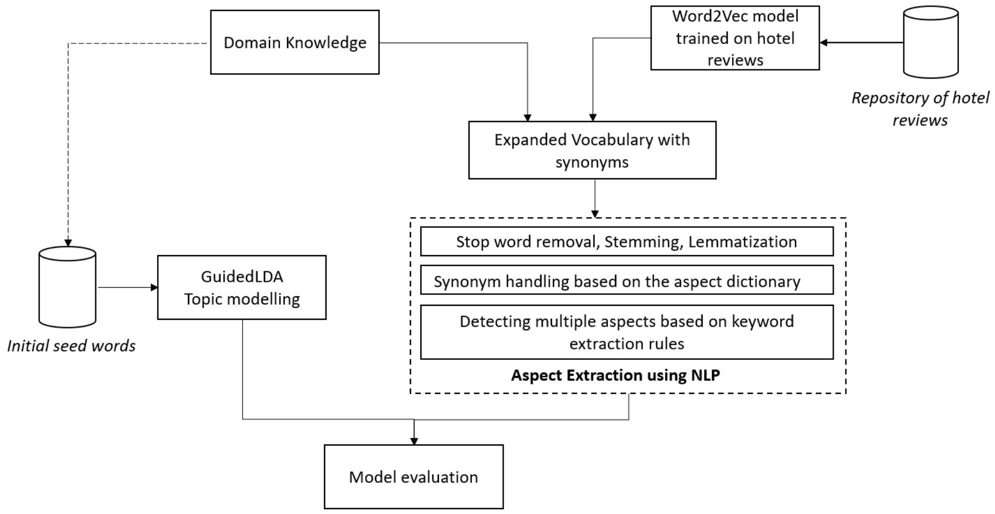


Figure 3. The high-level process of the hybrid aspect extraction approach.

3.2.2. Generating 'Emotion-Aspect Profile'

The integration of emotion extraction (Section 3.2.1) and aspect extraction (Section 3.2.1) methodologies allowed for the development of an insightful 'Emotion-Aspect profile' for each online review, as depicted in Figure 4. This profile presents a structured and quantitative representation, encompassing hotel aspects and profound emotions, making it easily interpretable. In the context of emotion modelling and psychology research, the concept of emotion dynamics outlines emotion intensity as a feature that is essential to modelling emotion behaviours [44]. The emotion intensity profile, a key element of emotion dynamics features, portrays the variation in emotions at a specific moment. The derived emotion intensity profile for a given review, customer or hotel demonstrates the aggregation of emotions in a given time point.

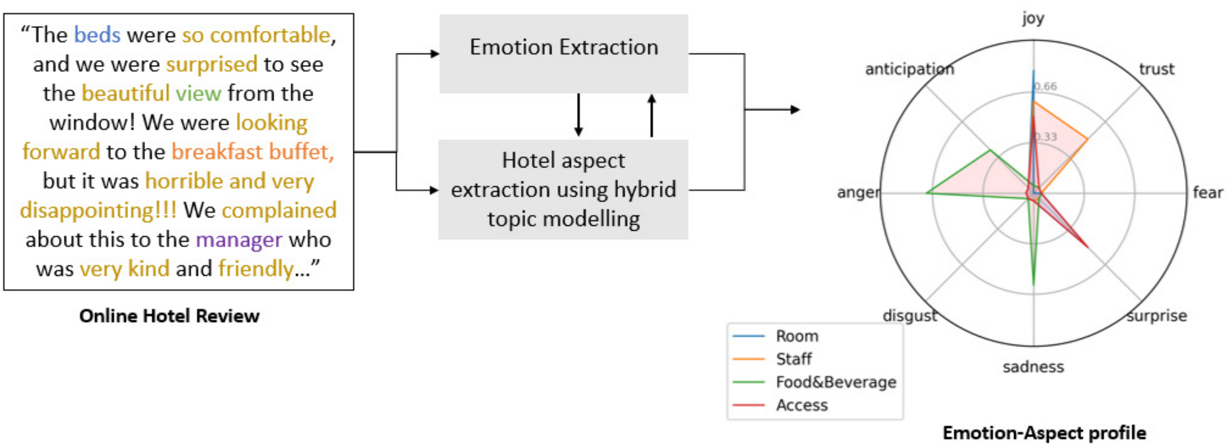


Figure 4. The combined approach in creating the emotion-aspect profile for online reviews. Emotion expressions are noted in Yellow while hotel aspects are denoted in different colours.

Since these techniques can be efficiently scaled to handle large volumes of data, the proposed platform is convenient for analysing content within social media and can be adjusted for diverse domains by fine-tuning the vocabulary. As this representation can be constructed at the review level (the most granular form), the emotion-aspect profiles can be aggregated based on time, individual customers or hotels to form a comprehensive representation of their behaviours. This quantified representation, with its variables of emotions and hotel aspects, can be transformed into a vector representation, serving as input for many analytics and machine learning tasks. In our study, we employed this representation for emotion profiling.

3.3. An Unsupervised Clustering Approach for Benchmarking Hotels Using Customer Experiences on Social Media

Based on the emotion-aspect representations derived from the above approach, further evaluation was carried out to segment reviews, customers and hotels based on digital traces on social media. Given the unlabelled nature of the data, profiling was carried out using an unsupervised clustering algorithm. By aggregating emotion-aspect profiles at different levels, the following profiling analyses were carried out.

- Profiling online reviews;
- Profiling customers;
- Profiling hotels.

The unsupervised clustering of hotel emotion-aspect profiles was formed using an improved variant of the GSOM [11]. The GSOM introduces a map topography that self-structures by adapting its size and shape. The GSOM network structure starts with a minimal number of nodes (typically four nodes) and grows on boundaries based on heuristics and input representation. The GSOM consists of two phases; first, the growing phase where the neural network grows new nodes and adjusts node weights to sufficiently represent the input space, and second, the smoothing phase in which the node weights are fine-tuned. The growth of the GSOM is determined by the number of dimensions in the input space and the spread factor (SF), where the SF can be utilized to control the spread of the neural network structure independent of the dimensionality of the dataset [11]. In contrast to other clustering algorithms, the GSOM preserves the structural representation of the underlying data. Most of other clustering algorithms require a defined map representation or pre-defined number of clusters. However, the GSOM is able to grow the nodes based on features of the input data space. It also has the ability to discard the outdated information and restrict overfitting knowledge in its knowledge acquisition, therefore preserving the stability [45].

Since the GSOM provides a spatial representation of reviews, it provides an indication of groupings with similar characteristics based on the emotions and hotel aspects. Further, this allows positioning and benchmarking of one hotel against the rest of the hotels based on visitor experiences shared through online platforms.

4. Results

The results outline the emotion-aspect representations and segmentation at three dimensions, the online review level, customer level and hotel level, by aggregating reviews at respective stages.

4.1. Emotion-Aspect Representation and Hotel Segmentation

This approach allows for creating an individualized visualization for each hotel based on the emotions and hotel aspects providing customized insights. The following Figure 5 shows the emotion-aspect profile of two hotels based on their customer comments recorded on social media.

A comparison of emotion-aspect profiles of multiple hotels enables the identification of the unique characteristics of each hotel. In Hotel A, there is increased negativity expressed towards ‘Room’ given the amplified intensities in ‘Sad’, ‘Disgust’ and ‘Anger’. As all reviews are structured to represent emotion variations, this approach allows managers to drill down to each emotion level to understand the causes for such negativity. In contrast, Hotel A shows increased positive emotions for ‘Staff’ as explained by higher intensities in ‘Trust’ and ‘Joy’. Compared to Hotel A, Hotel B showcases a varied emotion profile with distinct characteristics. More customers have expressed an increased level of 'Anger' and 'Sad' for aspects 'Staff' and 'Food&Beverage'. Notably, the aspects 'Public Areas' and 'Bathroom' have recorded higher intensity of 'Surprise'. Such analysis highlights the best-performing and poor-performing hotel aspects, thereby providing insights into the aspects that need attention.

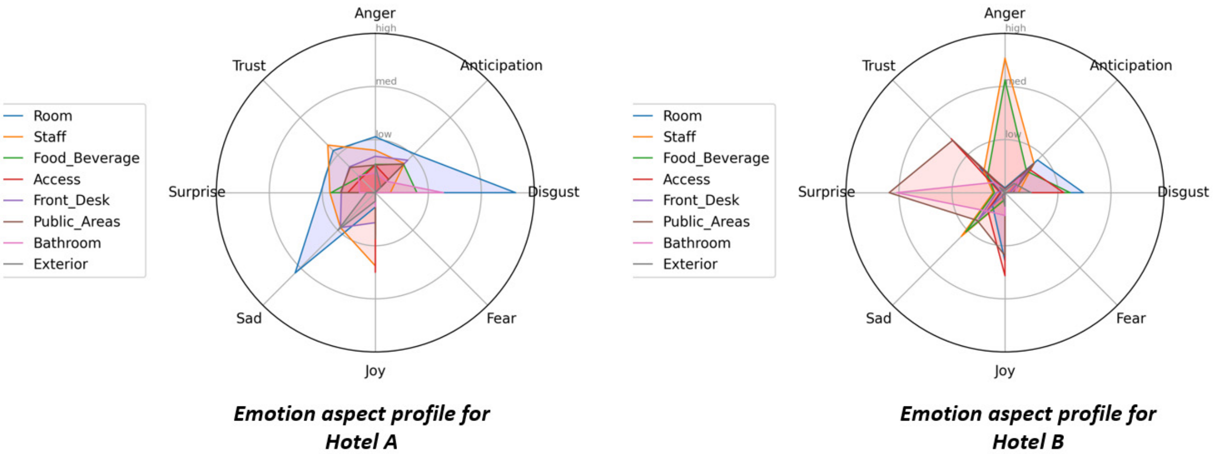


Figure 5. Comparison of emotion-aspect profiles for two hotels.

In this experiment, we utilised the emotion intensity profile to model the emotion variation corresponding to hotel aspects over time. At each time point, it provides a holistic view of the customer emotions directed towards the hotel, as shown in Figure 6. This illustrates the intensities of emotions for each hotel aspect over three years. As observed, the constellation of emotions has changed over time for each hotel aspect. Similar intensity profiles can be generated for granular levels such as days, weeks or months by aggregating the corresponding reviews in that period. Once integrated with a real-time social media data stream, this enables visualisation of the emotion fluctuations in near real-time, thus enabling a closer inspection of emotions presented on social media.

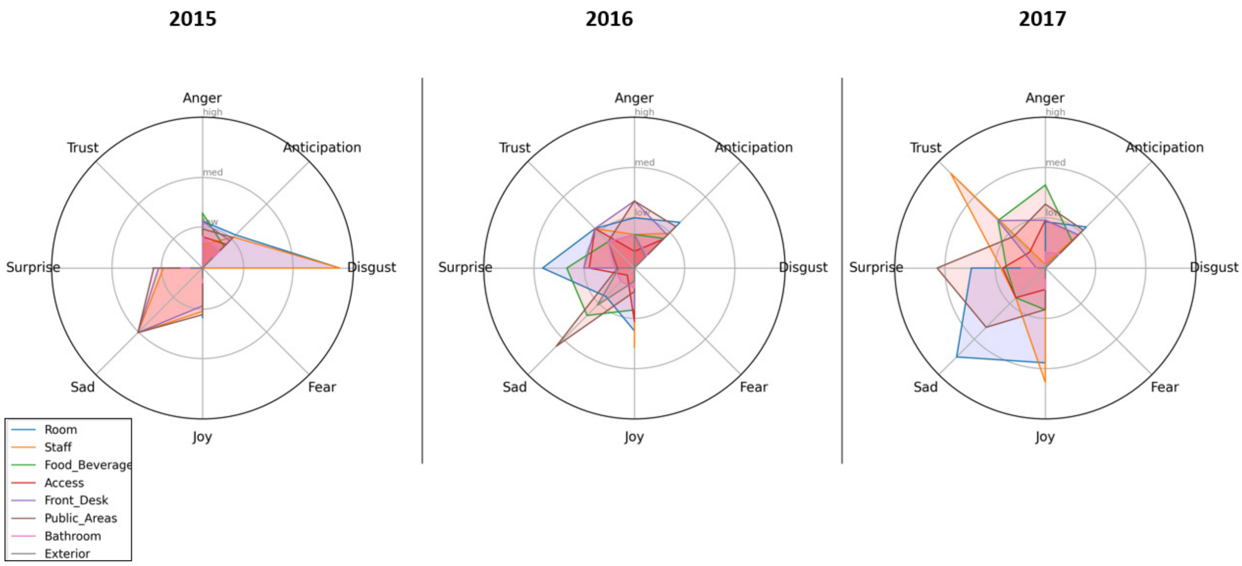


Figure 6. Demonstration of customer emotion fluctuations over time for a hotel.

Based on the derived emotion-aspect profiles, further analysis was conducted using the GSOM algorithm to create a segmentation of all 44 hotels based on the derived deep emotion intensity values. The 44 hotels' emotion variations per each aspect were fed into the GSOM to create a benchmarking visualisation. It was noted that there were four distinct clusters (profiles) based on their emotion variations. As the hotels that have been grouped showcase similar characteristics, the hotels can identify other similar hotels that perform similarly. Figure 7 illustrates the identified profiles and their distinct characteristics based on customer emotions.

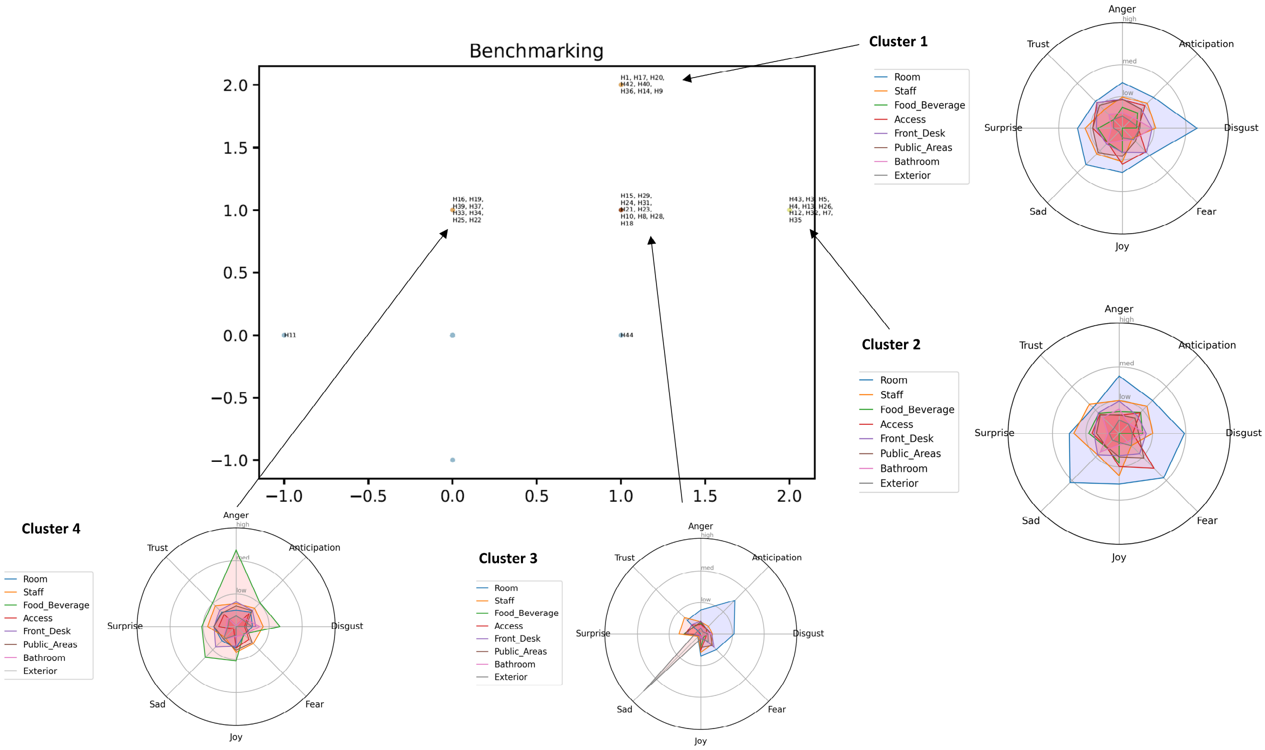


Figure 7. Cluster-based segmentation of hotels based on emotions.

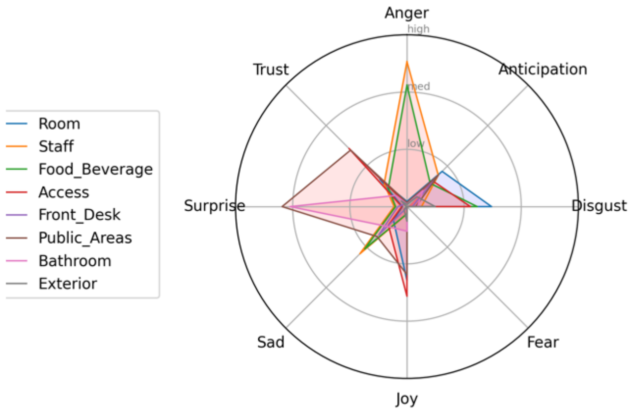
The most prominent aspect of hotels in Cluster 1 is ‘Room’, given its high intensity in all emotions compared to other aspects. It is also noted that higher negativity is shown for this aspect. Cluster 2 also shows increased negativity for ‘Room’ as denoted via high intensities in ‘Anger’, ‘Sad’, ‘Fear’ and ‘Disgust’. However, in this group, there is increased positivity shown for the aspect ‘Staff’. Cluster 3 shows a different distribution of emotions with higher intensity in ‘Sad’ for ‘Exterior’, which represents the design, appearance and ambience of the hotel. There is also an increased level of ‘Anticipation’ directed towards the aspect ‘Room’. The hotels in Cluster 4 have received an increased level of ‘Anger’, ‘Disgust’ and ‘Sad’ for ‘Food&Beverage’, representing customers’ dissatisfaction with the refreshments. Other aspects do not exhibit a clear difference.

In a holistic view, the model provides an overview of different groupings of hotels in the industry. If dimensions such as location, managing organization information, etc., are added, it is possible to create a similar benchmarking visualisation to identify different hotel profiles in a given location.

Moreover, the GSOM can be used as a dynamic visualization tool for exploratory analytics to uncover hidden associations. The positioning on the map indicates a visual representation of the similarity between the elements in the input space, thereby creating an automated visualization for insight generation. The hotels that are positioned closely on this map demonstrate similar behaviours based on customer experiences recorded on social media. This can be effectively integrated into decision-making platforms and dashboards and can be used as a key visualization component to construct a holistic view of the industry.

4.2. Emotion-Aspect Representation and Segmentation of Reviews

Figure 8 displays the emotion aspect profile derived from a comprehensive online review, showcasing the intensities of emotions towards both best-performing and poor-performing hotel aspects. The discernible differences in emotion variations and the distinct characteristics of emotions and aspects underscore the emotion-aspect profile's role as a unique and granular representation of customer satisfaction. Hence, this representation can be aptly described as a "digital representation of customer satisfaction".



	NEGATIVE				POSITIVE			
	Anger	Disgust	Fear	Sad	Anticipation	Joy	Surprise	Trust
Room	1.80%	29.60%	0.00%	6.70%	17.40%	25.80%	1.60%	13.40%
Staff	50.60%	5.10%	0.00%	23.30%	15.50%	2.60%	4.50%	11.30%
Food and Beverage	42.40%	24.60%	0.00%	20.90%	11.50%	2.50%	4.00%	9.20%
Access	0.80%	22.20%	0.00%	9.20%	12.70%	31.30%	1.80%	28.70%
Front Desk	1.20%	1.90%	0.00%	13.90%	14.40%	0.30%	3.20%	1.30%
Public Areas	1.40%	3.10%	0.00%	14.80%	15.60%	23.70%	43.70%	27.90%
Bathroom	0.90%	3.70%	0.00%	9.80%	7.90%	8.60%	40.90%	5.50%
Exterior	0.80%	9.90%	0.00%	1.50%	4.90%	5.10%	0.20%	2.80%

Figure 8. Emotion-aspect profile of a hotel based on online reviews. Red colour denotes the negative emotions while positive emotions are shown in green.

It can be observed that the aspects 'Staff' and 'Food&Beverage' have a higher intensity of 'Anger' and 'Sad' expressed in reviews compared to other aspects. Moreover, the aspect 'Room' has shown higher intensity in 'Disgust'. In positive emotions, it can be seen that 'Public Areas' and 'Bathroom' have increased intensity in 'Surprise', denoting a positive experience. The aspects 'Access' and 'Room' have mixed emotions of 'Disgust', 'Joy' and 'Trust'. The spectrum of emotions per each aspect denotes how customers position their experiences and enables them to identify the best- and worst-performing services of a hotel.

Subsequently, the emotion-aspect profiles of the reviews were vectorized and utilized as input data for the GSOM algorithm. For the experiment, we aggregated reviews specific to a particular hotel and generated a GSOM representation. The segmentation output by the GSOM is depicted in Figure 9. In the GSOM visualization, each node represents an online review and the nodes positioned in closer proximity (within the same neighbourhood) indicate similarity in their content. Based on this premise, we can identify distinct profiles of reviews for a given hotel.

As depicted in Figure 9, the GSOM algorithm distinctly segregates the reviews based on the different aspects being discussed. Clusters 1, 2 and 3 predominantly exhibit negative emotions, yet there is a clear separation in concerns, with Cluster 1 focusing on public areas, Cluster 2 on location and Cluster 3 on bathroom. On the left side of the GSOM map, the clusters reflect more positive emotions, with Cluster 4 showing mixed emotions for room and bathroom, and Cluster 5 displaying heightened positive emotions towards staff.

Notably, the positive emotions cluster is positioned far from the other clusters, indicating a distinct separation. This GSOM representation not only facilitates the segmentation of online reviews but also offers a holistic view of the reviews for a specific hotel. It enables an insightful understanding of how emotions and concerns are related to different aspects of the hotel experience, providing valuable information for decision-making and customer service enhancements.

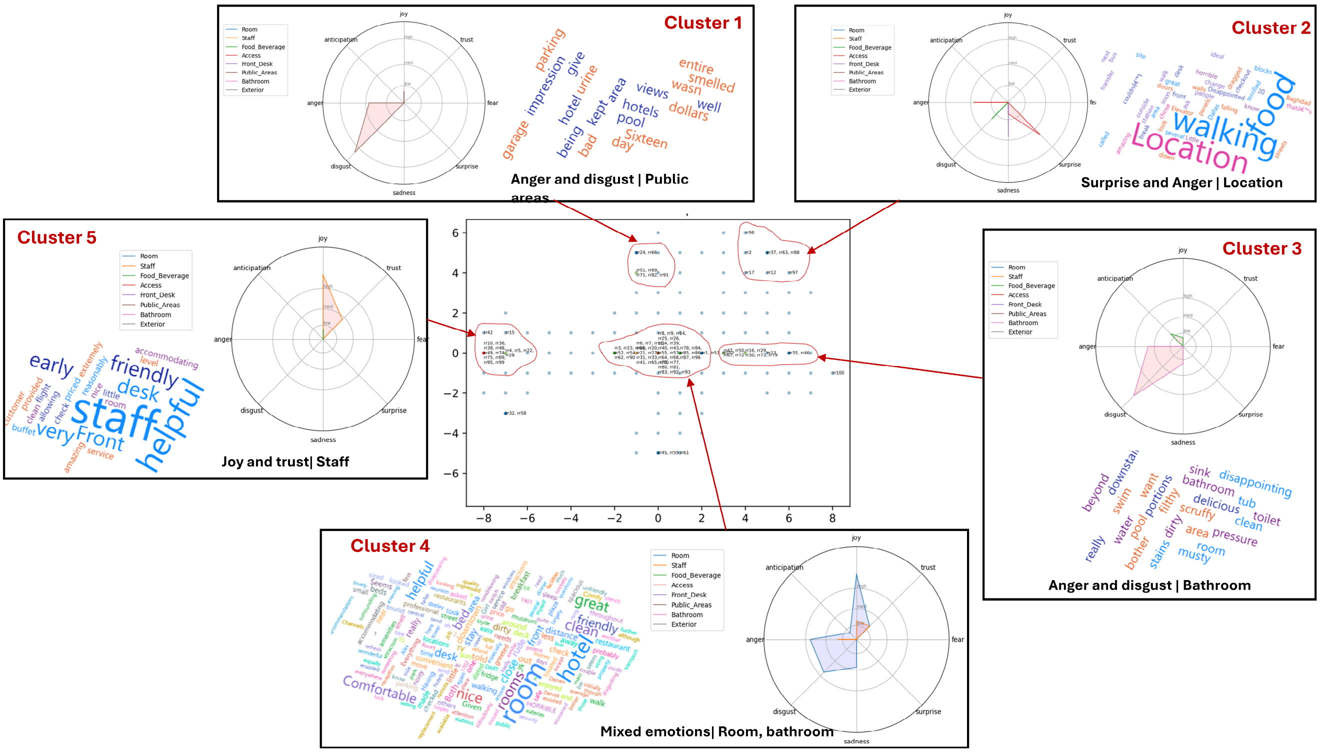


Figure 9. Segmentation of online reviews of one selected hotel using the GSOM.

4.3. Emotion-Aspect Representation and Segmentation of Customers

In prior research within the services sector, emotions have been utilized as a segmentation variable to identify meaningful customer groups. However, due to the inherently subjective nature of emotions, they can vary significantly among individual customers. Therefore, it is essential to model emotions at both the individual level and aggregate level to uncover similar emotional behaviours among customers. In this analysis, we showcase both approaches.

Figure 10 displays the emotion-aspect profiles of two customers, based on their online reviews for the same hotel during the same time period (month). The two profiles reveal distinct experiences encountered by these customers, with varying emotions and mentions of hotel aspects. Customer 1 exhibits a predominantly positive emotional state compared to Customer 2, who expresses high levels of anger and sadness towards the hotel staff. Customer 1 reflects joy and trust emotions concerning the staff and room, while expressing anger towards the food and beverage aspect.

Analysing emotions at an individual level enables the identification of specific pain points experienced by individual customers, thereby facilitating the development of targeted customer management strategies. By understanding the emotions and concerns of customers at an individual level, hospitality providers can tailor their services to meet individual needs, enhancing overall customer satisfaction and loyalty.

Subsequently, customers' emotion-aspect profiles were vectorized and used as input data for the GSOM algorithm, which generated groupings providing an aggregated view of customers. Each data point in this map represents a customer, and the profiles reveal groupings of customers with similar behaviours in terms of the emotions and hotel aspects mentioned in their reviews (Figure 11).

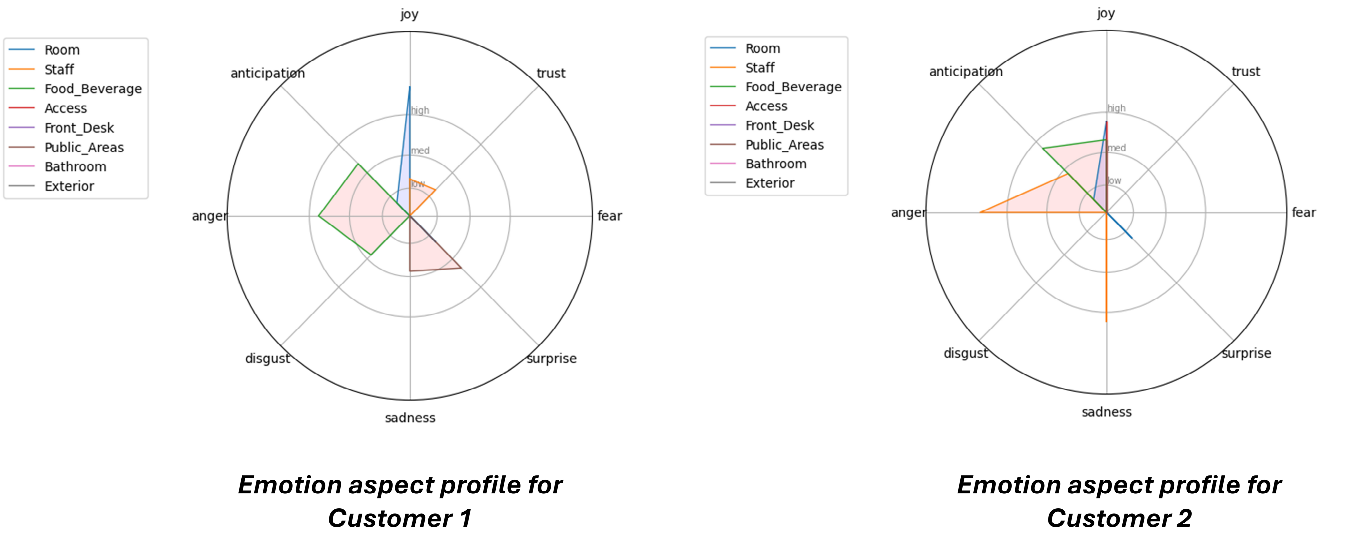


Figure 10. Emotion-aspect profiles of individuals based on their online reviews for the same hotel.

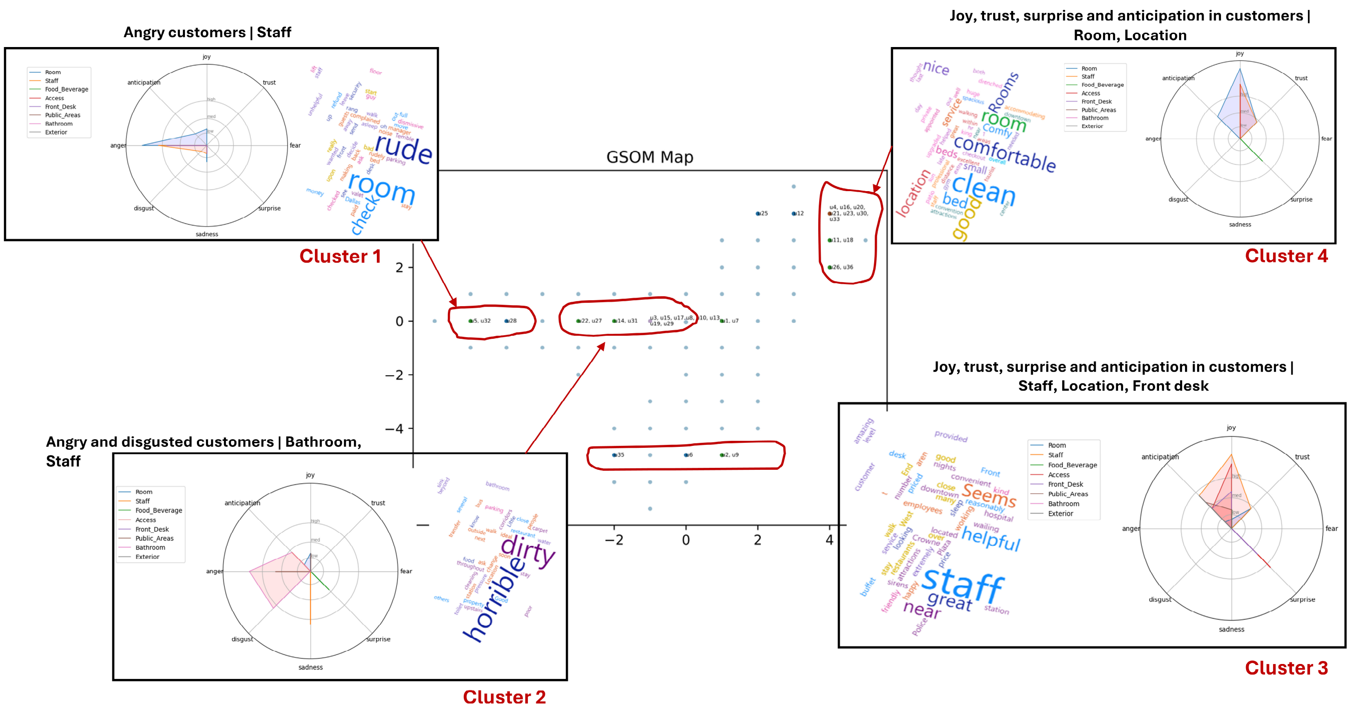


Figure 11. Segmentation of customers using the GSOM.

As observed, Cluster 1 exhibits angry customers with an intense expression of anger directed towards staff. In Cluster 2, customers express both anger and disgust, voicing complaints about staff and bathroom aspects. However, on the right side of the map, in Clusters 3 and 4, customers showcase more positive emotions. Cluster 3, comprised of positive customers, highlights mentions of staff and location in their reviews. In contrast, customers grouped in Cluster 4 have expressed high intensity of joy, complimenting the room and location.

Notably, the use of emotion as a segmentation variable results in a clear separation of customers, even when considering the same hotel aspect. It demonstrates that there can be significant variations in emotions expressed by customers, allowing for a nuanced understanding of their experiences and preferences. This approach enables the identification of distinct customer groups with shared emotional characteristics, contributing to enhanced customer segmentation and personalized service strategies in the hospitality industry.

The holistic view of customers of a hotel with such a visualisation provides direct insights for decision-makers and a deeper investigation would reveal details at a granular level. Hotels should incorporate such analytics into their customer management strategies to improve targeted customer care.

Based on this premise, we have shown the applicability of applying computational modelling of deep emotions to determine the emotions of customers over time. This has been made possible due to the digital traces of customers that have accumulated over a long period. Continuous monitoring of such outcomes could be extended to developing time-series models to forecast the emotional fluctuations of customers. Given that the generation of emotion-aspect profiles could be automated without high computational resources, hotel profiles can be updated regularly based on the frequency of data to derive a timely analysis and understanding of hotel experiences.

5. Discussion

5.1. Conclusions

The advent and the prevalence of social media platforms have transformed customer communication into a digital form, thereby empowering social media analytics to be a major component of modern organizations [46]. Therefore, it is required to systematically analyse social media content to extract different facets of information such as topics of discussion, sentiment and emotion using combined approaches in machine learning, NLP and analytics. As this information may not translate into informed insights naturally, they should be methodically transformed, combined, and analysed to generate useful insights and visualizations.

This study presented a machine learning and NLP approach that is able to systematically extract topics (hotel aspects) and emotion modelling from online reviews to generate an interpretable and structured form of customer experiences as a digital footprint. The derived emotion-aspect profiles represent a holistic view of customer experiences, to support customer emotion analysis and segmentation based on the online presence of hotels. Such analysis could have significant potential value for managerial decision-making in the tourism and hospitality industry.

5.2. Theoretical Implications

Compared to the previous literature, this study makes several contributions to the current research space. The generation of an emotion-aspect profile to characterise online reviews in a structured, interpretable and quantitative form as a customer's digital footprint provides the foundation for multiple analyses. This profile developed at an online review level provides a granular breakdown of an individual customer's entire experience. When aggregated at a customer level, it represents the individual characteristics of a customer's experience over time. Lastly, when aggregated at a hotel level, the emotion-aspect profile provides a snapshot of the hotel's digital footprint at a given time. This theorises the capturing of a hotel's digital traces and transforms it into a quantifiable structure that can be used as digital feedback to an organization's processes or decision-making. In this study, we have shown the applicability of using the derived emotion-aspect profile as the basis to conduct a segmentation study. Based on different aggregated levels, we have showcased the basis for review segmentation, customer segmentation and hotel segmentation, thus providing a holistic view of the hospitality industry at a given point in time.

Using the automated analytical functionalities proposed in this approach, it is possible to analyse a large quantity of customer feedback, compared to interviews and questionnaires that are focused on a limited number of customers. Therefore, the framework supports the digital transformation of the tourism and hospitality industry by converting the opinions and emotions of customers expressed in the digital space into actionable insights.

5.3. Managerial Implications

As practical implications of this study, the hotel industry could adopt the techniques to extract emotions and hotel aspects mentioned in online reviews, thereby creating a digital representation of customer satisfaction. The derived emotion-aspect profile will highlight the pain points of customers, which can be accounted for when improving services. This analysis can be directly linked to improving services to enhance the hotel aspects with limitations. Proper and timely management of responses and identifying sentiments expressed towards the hotel is crucial for hotel managers to disseminate positive eWOM to customers.

Another significant implication of this study lies in the potential to conduct segmentation and benchmarking using data from digital sources. This capability can even extend to an individual level, where visitors' unique individualities and emotional characteristics can be derived based on their previous responses to other hotels over time. Identifying negative customers is crucial for effective response management, enabling proactive communication and action to mitigate negativity in online reviews towards the hotel. By distinguishing clusters of discontented customers or individuals with specific concerns, hotels can direct their efforts on resolving these issues and improving the overall customer experience. This strategic approach to customer segmentation facilitates resource allocation and prioritization, ensuring that developments are directed where they are most required.

Furthermore, the self-adapting GSOM can be positioned as a valuable visualization tool to analyse data from multiple dimensions, providing insights that can be seamlessly integrated into current decision-making systems, thereby offering a competitive advantage to the hotel. Such analysis offers an opportunity for hotel managers to compare customer emotions towards competing hotels and take action to benchmark and enhance their position among similarly performing establishments. This approach facilitates the automatic detection of similarly performing hotels based on both aspects and emotions and as this is computationally less costly, integrating this into digital marketing strategies would be beneficial for continuous service improvement. Over time, hotels will be able to observe a shift in these profiles, thus identifying key characteristics that emerge over time.

Therefore, the proposed approach offers the opportunity to fully leverage online reviews by supporting multiple analysis tasks using the derived representation of emotions and aspects mentioned in the reviews. This allows the hotel industry to tap into the wealth of digital data created on social media platforms to inform decision-making and strategy development. Consequently, the outcomes of this study can drive data-driven digital marketing strategies, as online reviews serve as a critical communication medium between customers and hotel managers.

5.4. Limitations and Future Work

We acknowledge the following limitations and future work that can be carried out to further enhance this study. Fake review detection and handling ambiguous emotional expressions are not addressed in this approach and should be investigated further. Moreover, the analysis of customer emotions can be further expanded by including information related to specific events that caused certain emotional responses and reporting service failures. In addition, we postulate the possibility of incorporating hotel information such as the type of the hotel, location and special facilities to derive informed insights that could drive managerial decision-making. Another research direction would be to model the dynamic nature of segments. As emotion-based segmentation is dependent on time, it is important to note that each customer segment will have distinct emotional motivators, further contributing to the dynamic nature of segmentation based on emotions. This could be addressed by continuously monitoring online reviews and updating the GSOM with real-time data.

Author Contributions: Conceptualization: A.A., S.N., R.N., D.D.S. and D.A.; Methodology: A.A., S.N., R.N., D.D.S. and D.A.; Formal analysis: A.A., S.N. and R.N.; Writing—original draft: A.A., S.N., R.N., D.D.S. and D.A.; Writing—review & editing: A.A., S.N., R.N., D.D.S. and D.A.; Supervision: S.N., D.D.S. and D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from TripAdvisor online platform and are available from https://www.tripadvisor.com/ (accessed on 26 February 2024) with the permission of data extraction policies of the platform.

Conflicts of Interest: The authors declare that there are no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

1. Tsai, C.-F.; Chen, K.; Hu, Y.-H.; Chen, W.-K. Improving Text Summarization of Online Hotel Reviews with Review Helpfulness and Sentiment. *Tour. Manag.* **2020**, *80*, 104122. [CrossRef]
2. Muhammad, S.S.; Dey, B.L.; Weerakkody, V. Analysis of Factors That Influence Customers' Willingness to Leave Big Data Digital Footprints on Social Media: A Systematic Review of Literature. *Inf. Syst. Front.* **2018**, *20*, 559–576. [CrossRef]
3. Al-Mutawa, R.F.; Al-Aama, A.Y. User Opinion Prediction for Arabic Hotel Reviews Using Lexicons and Artificial Intelligence Techniques. *Appl. Sci.* **2023**, *13*, 5985. [CrossRef]
4. Jia, Y.; Liu, I.L.B. Do Consumers Always Follow "Useful" Reviews? The Interaction Effect of Review Valence and Review Usefulness on Consumers' Purchase Decisions. *J. Assoc. Inf. Sci. Technol.* **2018**, *69*, 1304–1317. [CrossRef]
5. Chatterjee, S. Drivers of Helpfulness of Online Hotel Reviews: A Sentiment and Emotion Mining Approach. *Int. J. Hosp. Manag.* **2020**, *85*, 102356. [CrossRef]
6. Chang, V.; Liu, L.; Xu, Q.; Li, T.; Hsu, C.-H. An Improved Model for Sentiment Analysis on Luxury Hotel Review. *Expert Syst.* **2023**, *40*, e12580. [CrossRef]
7. Gharzouli, M.; Hamama, A.K.; Khattabi, Z. Topic-Based Sentiment Analysis of Hotel Reviews. *Curr. Issues Tour.* **2022**, *25*, 1368–1375. [CrossRef]
8. Hu, N.; Zhang, T.; Gao, B.; Bose, I. What Do Hotel Customers Complain about? Text Analysis Using Structural Topic Model. *Tour. Manag.* **2019**, *72*, 417–426. [CrossRef]
9. Plutchik, R. A Psychoevolutionary Theory of Emotions. *Soc. Sci. Inf.* **1982**, *21*, 529–553. [CrossRef]
10. Li, H.; Liu, H.; Zhang, Z. Online Persuasion of Review Emotional Intensity: A Text Mining Analysis of Restaurant Reviews. *Int. J. Hosp. Manag.* **2020**, *89*, 102558. [CrossRef]
11. Alahakoon, D.; Halgamuge, S.K.; Srinivasan, B. Dynamic Self-Organizing Maps with Controlled Growth for Knoledge Discovery. *IEEE Trans. Neural Netw.* **2000**, *11*, 601–614. [CrossRef] [PubMed]
12. Park, E.; Kang, J.; Choi, D.; Han, J. Understanding Customers' Hotel Revisiting Behaviour: A Sentiment Analysis of Online Feedback Reviews. *Curr. Issues Tour.* **2020**, *23*, 605–611. [CrossRef]
13. El-Said, O.A. Impact of Online Reviews on Hotel Booking Intention: The Moderating Role of Brand Image, Star Category, and Price. *Tour. Manag. Perspect.* **2020**, *33*, 100604. [CrossRef]
14. Varkaris, E.; Neuhofer, B. The Influence of Social Media on the Consumers' Hotel Decision Journey. *J. Hosp. Tour. Technol.* **2017**, *8*, 101–118. [CrossRef]
15. Geetha, M.; Singha, P.; Sinha, S. Relationship between Customer Sentiment and Online Customer Ratings for Hotels—An Empirical Analysis. *Tour. Manag.* **2017**, *61*, 43–54. [CrossRef]
16. De Pelsmacker, P.; van Tilburg, S.; Holthof, C. Digital Marketing Strategies, Online Reviews and Hotel Performance. *Int. J. Hosp. Manag.* **2018**, *72*, 47–55. [CrossRef]
17. Barreda, A.; Bilgihan, A. An Analysis of User-generated Content for Hotel Experiences. *J. Hosp. Tour. Technol.* **2013**, *4*, 263–280. [CrossRef]
18. Sparks, B.A.; Browning, V. The Impact of Online Reviews on Hotel Booking Intentions and Perception of Trust. *Tour. Manag.* **2011**, *32*, 1310–1323. [CrossRef]
19. Yan, Q.; Zhou, S.; Wu, S. The Influences of Tourists' Emotions on the Selection of Electronic Word of Mouth Platforms. *Tour. Manag.* **2018**, *66*, 348–363. [CrossRef]
20. Xie, K.L.; Zhang, Z.; Zhang, Z. The Business Value of Online Consumer Reviews and Management Response to Hotel Performance. *Int. J. Hosp. Manag.* **2014**, *43*, 1–12. [CrossRef]
21. Giatsoglou, M.; Vozalis, M.G.; Diamantaras, K.; Vakali, A.; Sarigiannidis, G.; Chatzisavvas, K.C. Sentiment Analysis Leveraging Emotions and Word Embeddings. *Expert Syst. Appl.* **2017**, *69*, 214–224. [CrossRef]

22. Herzig, J.; Shmueli-Scheuer, M.; Konopnicki, D. Emotion Detection from Text via Ensemble Classification Using Word Embeddings. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, Amsterdam, The Netherlands, 1–4 October 2017; ACM: New York, NY, USA, 2017; pp. 269–272.
23. Abeysinghe, S.; Manchanayake, I.; Samarajeewa, C.; Rathnayaka, P.; Walpola, M.J.; Nawaratne, R.; Bandaragoda, T.; Alahakoon, D. Enhancing Decision Making Capacity in Tourism Domain Using Social Media Analytics. In Proceedings of the 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 26–29 September 2018; pp. 369–375.
24. Abdaoui, A.; Azé, J.; Bringay, S.; Poncelet, P. FEEL: A French Expanded Emotion Lexicon. *Lang. Resour. Eval.* **2017**, *51*, 833–855. [CrossRef]
25. Mohammad, S.M.; Turney, P.D. Crowdsourcing a Word–Emotion Association Lexicon. *Comput. Intell.* **2013**, *29*, 436–465. [CrossRef]
26. Puh, K.; Bagić Babac, M. Predicting Sentiment and Rating of Tourist Reviews Using Machine Learning. *J. Hosp. Tour. Insights* **2022**, *6*, 1188–1204. [CrossRef]
27. Li, H.; Yu, B.X.B.; Li, G.; Gao, H. Restaurant Survival Prediction Using Customer-Generated Content: An Aspect-Based Sentiment Analysis of Online Reviews. *Tour. Manag.* **2023**, *96*, 104707. [CrossRef]
28. Li, R.; Li, Y.-Q.; Ruan, W.-Q.; Zhang, S.-N.; Wang, M.-Y. Sentiment Mining of Online Reviews of Peer-to-Peer Accommodations: Customer Emotional Heterogeneity and Its Influencing Factors. *Tour. Manag.* **2023**, *96*, 104704. [CrossRef]
29. Chen, J.; Becken, S.; Stantic, B. Travel Bubbles to Maintain Safe Space for International Travel during Crisis—Emotions Reflected in Twitter Posts. *Curr. Issues Tour.* **2022**, *26*, 2479–2493. [CrossRef]
30. Ahn, J.; Kwon, J. Green Hotel Brands in Malaysia: Perceived Value, Cost, Anticipated Emotion, and Revisit Intention. *Curr. Issues Tour.* **2020**, *23*, 1559–1574. [CrossRef]
31. Calvo-Porral, C.; Lévy-Mangin, J.-P. An Emotion-Based Segmentation of Bank Service Customers. *Int. J. Bank Mark.* **2020**, *38*, 1441–1463. [CrossRef]
32. Maheswaran, D.; Shavitt, S. Issues and New Directions in Global Consumer Psychology. *J. Consum. Psychol.* **2000**, *9*, 59–66. [CrossRef]
33. Plutchik, R. *The Emotions*; University Press of America: Lanham, MD, USA, 1991; ISBN 978-0-8191-8286-9.
34. Mohammad, S.M.; Turney, P.D. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, CA, USA, 5 June 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 26–34.
35. Adikari, A.; Gamage, G.; de Silva, D.; Mills, N.; Wong, S.-M.J.; Alahakoon, D. A Self Structuring Artificial Intelligence Framework for Deep Emotions Modeling and Analysis on the Social Web. *Future Gener. Comput. Syst.* **2021**, *116*, 302–315. [CrossRef]
36. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
37. 515K Hotel Reviews Data in Europe. Available online: https://kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe (accessed on 19 April 2019).
38. Putting Some Emotion into Your Design—Plutchik's Wheel of Emotions. Available online: https://www.interaction-design.org/literature/article/putting-some-emotion-into-your-design-plutchik-s-wheel-of-emotions (accessed on 6 May 2023).
39. World Tourism Organization (UNWTO) (Ed.) *Hotel Classification Systems: Recurrence of Criteria in 4 and 5 Stars Hotels*; World Tourism Organization (UNWTO): Madrid, Spain, 2015; ISBN 978-92-844-1664-6.
40. Bagherzadeh, S.; Shokouhyar, S.; Jahani, H.; Sigala, M. A Generalizable Sentiment Analysis Method for Creating a Hotel Dictionary: Using Big Data on TripAdvisor Hotel Reviews. *J. Hosp. Tour. Technol.* **2021**, *12*, 210–238. [CrossRef]
41. Xu, X. Examining an Asymmetric Effect between Online Customer Reviews Emphasis and Overall Satisfaction Determinants. *J. Bus. Res.* **2020**, *106*, 196–210. [CrossRef]
42. Zhang, J.; Lu, X.; Liu, D. Deriving Customer Preferences for Hotels Based on Aspect-Level Sentiment Analysis of Online Reviews. *Electron. Commer. Res. Appl.* **2021**, *49*, 101094. [CrossRef]
43. Jagarlamudi, J.; Daumé III, H.; Udupa, R. Incorporating Lexical Priors into Topic Models. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, Avignon, France, 23–27 April 2012; pp. 204–213.
44. Kuppens, P.; Verduyn, P. Emotion Dynamics. *Curr. Opin. Psychol.* **2017**, *17*, 22–26. [CrossRef] [PubMed]
45. Nawaratne, R.; Alahakoon, D.; De Silva, D.; Chhetri, P.; Chilamkurti, N. Self-Evolving Intelligent Algorithms for Facilitating Data Interoperability in IoT Environments. *Future Gener. Comput. Syst.* **2018**, *86*, 421–432. [CrossRef]
46. Fan, W.; Yan, X. Novel Applications of Social Media Analytics. *Inf. Manag.* **2015**, *52*, 761–763. [CrossRef]