

The Reason behind the Rating: Text Mining of Online Hotel Reviews

Abstract— Online customer reviews are a significant marketing resource for products and service business, often serving as a key reference point for potential customers. This study analyses a set of hotel reviews to find out what aspects of hotel experience are taken into consideration by customers when rating a hotel. We use text mining of hotel reviews to identify main features referred to by customers and determine their correlation with customer satisfaction or dissatisfaction. Our data set comprises reviews of ten hotels located in the vicinity of “Khao San Road” in the centre of Bangkok in Thailand, a popular holiday destination known as the centre of the “backpacking universe”. The results highlight several main aspects affecting customer satisfaction, with friendly and helpful staff being the most influential one, whilst negative experiences with complementary services provided by the hotel, such as pool and Wi-Fi have the strongest impact on customer dissatisfaction. Other factors impacting customer satisfaction include cleanliness, room and bathroom interior, sleep quality and location. Our findings can be used to design effective feedback gathering and social media monitoring systems, and they also underpin a set of managerial implications regarding managing and marketing customer experiences in the hotel industry.

Keywords— *Online Reviews, Text mining, Topic Modelling, Hotel aspects, Rating Factors, Preference Models*

I. INTRODUCTION

Online travel platforms encourage users to share their experiences and opinions about hotels and travel destinations, and this has created a massive amount of online user generated reviews [1]. At the same time, an increasing number of travellers are using online platforms to search, read and consume travel related information. As such, user generated content serves as a new form of electronic *Word of Mouth (eWOM)*, and is treated as a valuable source of information regarding customer satisfaction with different aspects of the hotel service.

Previous studies have examined the content of user complaints and compliments to identify factors of customer satisfaction, covering a broad range of topics such as determinants of satisfaction using online reviews [2], analysing customer complaints and managerial responses to those reviews [3], integrating online review and ratings [4] and customer delight [5]. These studies conclude that an important aspect in gauging customer satisfaction is to identify the main attributes of hotel service which customers consider to match their needs and expectations [6]. For example tourists would evaluate whether or not their expectations are fulfilled based on

specific aspects of the actual experience, for example availability of entertainment, or proximity to cultural landmarks (depending on the type of tourism), facilities and location [7]. Previous studies have shown that tourist satisfaction influences intentions to recommend and revisit [8][9]. Experiences shared online, in this context, are influential not only to gain awareness of customer preferences but these can also impact future service performance by attracting or repelling future customers [10]. Overall customer satisfaction score summarises the individual assessments of performance against the different attributes that define customer expectations of a product or a service [11]. For service sectors such as hotels, the overall customer satisfaction is assessed through the performances of different attributes of service quality against customer expectations.

This study aims to understand better the customer expectations regarding different attributes of hotel service in the context of tourism, and the way they impact their overall assessment of the hotel service in online reviews. This will not only impact the way in which guest experiences are managed, but will also guide the information models and the user interface design of effective online review systems and social media monitoring and analysis tools.

Within this aim, our analysis of online hotel reviews focuses on the following specific research questions:

RQ1: What attributes of the hotel service do guests consider important when rating their experiences with hotels?

RQ2: What are the most meaningful words in hotel ratings as indicators of user experience aspects? What is the relationship between these indicators and the overall hotel rating?

In this study, we use Text Mining to discover patterns and relationships within the text of customers’ online reviews [12]. The specific tool within the text mining toolkit we have chosen is Topic Modelling [12], an advanced software and mathematical technique originating from the fields of data mining and natural language processing. Topic Modelling allows identifying patterns in word frequencies and uses the patterns to define “topics”. These topics serve as indicators for the hotel service attributes considered important by customers. In the next step, regression analysis is carried out to verify the impact these service attributes, or factors, have on the overall customer satisfaction rating.

The rest of this article proceeds as follows. In the next section, we review related studies. We then describe the data and the process of developing our research method. In the fourth section, we present the empirical results and discuss their implications for hotel managers. In the final section, we conclude by setting out the limitations of this study and future topics for research.

II. LITERATURE REVIEW

Online customer reviews, also known as electronic Word of Mouth, or eWOM; provide information about various aspects of a product or a service, including value for money and overall evaluation [1]. The quality of the content of the consumer reviews is an important factor towards consumer purchasing behaviour [2]. The overall average ranking score of a hotel as well as the average rankings of specific hotel features benefit customers in obtaining quantitative information about quality and performance of hotels' products and services. Online hotel reviews are still important due to focusing on qualitative dimensions of hotels' products or services. This is mainly because of two main characteristics of review quality: review relevancy and "factualness" of reviews.

Online hotel reviews serve two purposes: travellers obtain information guiding their selection; whilst hotel operators can analyse such reviews to monitor customers' feedback about their service and product and evaluate their popularity [3]. Such eWOM tends to be a trustworthy source for customers looking for information about the product and/or service rather than messages from the owners of the product or service [4].

In the tourism field, scholars have examined online customer reviews from different perspectives. First, helpfulness of hotel reviews is among major research streams which reveals the classification among online customer sentiments (e.g. [5]). A second perspective investigates the impact of hotel reviews on sales [1]; on online booking [6]; and on customer choices [7]. A third perspective examines customer satisfaction. Indeed it is generally accepted that eWOM is shaped by the two extremes: the most satisfied and the most dissatisfied customers. Thus, reviewers who frequently write on websites, blogs or online communities tend to be those who are either extremely happy or extremely disappointed with the services or products [8]. Moreover, the majority of purchasers do not directly complain to service providers, but write negative eWOM comments and no longer purchase those services or products [8].

Customer satisfaction in the tourism industry is a function of customers assessing their experience of different service aspects against a set of expectations. Knowing which aspects influence customer satisfaction is thus very useful from both business and IT perspectives, helping to design effective feedback capturing and monitoring systems. Indeed, recent studies have illustrated that user-generated content is a potential source of fine-grained information about user behaviour, opinions and sentiment [6] which play an important role in motivating customers loyalty or giving a recommendation.

A number of studies have aimed to identify significant factors for customer satisfaction, aiming to guide the provision

of services that are preferred by customers. Li, Cui and Peng [9] summarise as influential hotel image, staff manner, facilities and quality of service. Shanka and Taylor [10] conclude that front desk staff, check-in and check-out process, restaurant and bar provided by the hotel are significant factors in customer satisfaction. They also indicate the factors of lowest importance including entertainment amenities, business services and the mini bar provided in guest rooms. Research conducted by [11] reveals that cleanliness, price value of the room, friendly staff and security are the factors focused on in selecting hotel accommodation. Besides, there is a direct relationship between hotel hygiene factor and guests' intention to come back [12]. Vermeulen and Seegers [7] found that exposure to online reviews, both positive and negative; enhance consumers' consideration of hotels, and the effect is stronger for lesser-known hotels. However, only a few studies investigate factors which impact the overall review scores in the hotel industry ([13] and [14]). In both [23] and [24] studies, the attributes of value and cleanliness were found to be positively associated with the overall satisfaction scores. On the flip side, if the hotel property's facilities are old or if the service is poor compared to what is generally expected for the particular star rating category, customers will feel dissatisfied [15]. Customer service, cleanliness, location, value, facility, size and décor of guestroom, amenities, amenities, food, quietness are important attributes identified through analysing travel blogs [16].

Previous studies all have highlighted the factors impacting customers and usually these are not clearly described by the overall hotels' ranking. Xu, X. and Li, Y., 2016 [17] found that the determinants that create either customer satisfaction or dissatisfaction toward hotels are different and are specific to particular type of hotels. Generally, they found that the determinants of customer satisfaction toward various hotel types are generally same: location, staff performance and nice and clean room, and good restaurant. Determinants for customer dissatisfaction are WiFi, parking, bathroom, unfriendly and unhelpful staff. Radojevic, T. et al., 2015 [18] examines the customer satisfaction factors for a sample of 6768 hotels located in 47 capital city in Europe. The factors are either city-specific factors or hotel-specific factors. Their findings show that the hotel star rating is the single most important factor that influences customer experience. the presence of air-conditioning devices in rooms, a bar located within the hotel area, access to free Wi-Fi, membership in a branded hotel chain and price have significant positive associations with customer satisfaction (*ceteris paribus*). Variables that appear to be adversely associated with customer satisfaction are distance from the city centre, size of the hotel, and general hotel price level in the city where the hotel is located.

Overall, researchers recognise that measuring customer satisfaction can be a demanding challenge because of the complexity of customer experience aspects [19]. In addition, [20] raises the difference in perspectives when thinking about hotel evaluation between a hotel owner's beliefs and customers' experience. Previous studies have measured customer satisfaction and provided models through survey [21], case studies and interviews [22] or content analysis [1].

However, these studies suffer from the subjective nature of self-reported data and limited scope of content analysis. In contrast, text mining allows us to establish statistically sound dependencies between elements of user-generated content and summary satisfaction rankings, yet such studies are very seldom [23].

III. RESEARCH METHODOLOGY

A. Application context

Thailand is one of the countries where the main industry is tourism and hospitality. Khao San Road is one of the most attractive tourist destinations in Thailand. It is a historical short road in Bangkok, Thailand. In last two decades, Khao San Road has been the main attraction for backpackers and is thus called the “backpacker ghetto” [24]. However, since 2015 the type of travellers staying there has diversified away from the stereotypical “backpacker” [25]. At Khao San Road, there are many coffee shops, restaurants, bars and pubs, but also street food traders selling deep fried insects and exotic foods, and also nightclubs [25]. Consistent with this profile is the proliferation of low-cost accommodation in this area. Overall, Khao San Road is considered as one of the benchmarks of the tourism industry in Thailand, and is thus used as the focus of our sample of mid-range hotels.

B. Data collection and Sampling

Data are collected from Booking.com. Booking.com, unlike other hotel review websites, allows only customers who have been to the hotel they booked to write reviews, which can prevent deceitful comments from marketing teams and competitors [25]. 47% of reviewers are from Europe, followed by 33% from Asia. Analysing customers’ type shows that 52% of the reviewers were couple, 19% solo, 15% group and 12% family with young children, 2% family with older children and 1% people with friends. 97% of reviewers travelled for leisure which is mainly reflect on the location. As shown in Figure 1 below, Booking.com structures reviews into (1) negative and (2) positive sections, and provides information about (3) date of review, (4) trip type, (5) traveller type, (6) room type, (7) number of nights per customer stay in that hotel/hostel, (8) of traveller, (9) number of reviews written by user and (10) rating that they give the accommodation.



Figure 1. Example of a review on Booking.com

The rating of each hotel is calculated from the average of review ratings from all travellers who wrote about and visited it. The website also breaks down the overall rating into a rating for each factor to provide more information as follow: (i) Cleanliness, (ii) Comfort, (iii) Location, (iv) Facilities, (v) Staff, (vi) Value of money, (vii) Free WiFi.

Data has been gathered from 200 reviews of 10 hotels. Hotels have been selected based on following criteria. Firstly, all the considered hotels are 3-star. Secondly, the review dates cover a period of around two and a half years, from 1 January 2015 to 15 July 2017; because reviews posted before 2015 no longer affect hotel ratings [26]. Thirdly, all the hotels are located near the Khao San Road area in Bangkok, Thailand, but not on the main Khao San Road. It is to ensure that location aspect would not cover the influence of another aspect of customer’s hotel satisfaction. Lastly, they are located less than a mile from Khao San Road to ensure that travellers go to the same range of attractions and eliminate any bias which might come from different purposes of travelling. All the variables collected through data collection are listed in Table 1.

The following steps were included in this research design.

- Conversion of unstructured data to structured data using text mining
- Extraction of useful words from structured data
- Definition of segment of those words from step 2
- Definition of topic composed of those words from step 2
- Identification of the relevant words and factors which affect hotel ratings and represent customer satisfaction
- Obtaining interesting insights from groups of essential words and factors from step 5.

TABLE 1. DATA SET DESCRIPTION

Variable	Description
Review ID	Review ID
Hotel ID	Hotel ID
Hotel name	Hotel name
Hotel rating	The overall rating out of 10 for hotel
Hotel rating in the following factors: (i) Cleanliness, (ii) Comfort, (iii) Location, (iv) Facilities, (v) Staff, (vi) Value of money, (vii) Free WiFi.	Rating of each factor provided by Booking.com
Date	The date that the traveller shared experience on web platform
Review rating	Review rating score
Negative comment	Negative comment
Positive comment	Positive comment
Number of stay	Number of nights traveller stayed in hotel
Customer type	Type of traveller, i.e. solo, couple, family, group, family with children, family with older children, and people with friends.
Trip type	Type of trip, i.e. business or leisure trips
Customer country of residence	
Customer number of review	Number of reviews written by customer

C. Data analysis

In order to improve the accuracy of the model in this study, two more variables are generated as follow:

Target variable: customer satisfaction

If review rating > 7 then ‘Satisfy’

Else ‘Dissatisfy’

Hotel rating description:

If hotel rating > 9 then ‘Superb’

If hotel rating > 7 then ‘Good’

If hotel rating > 5 then ‘Okay’

If hotel rating > 3 then ‘Poor’

Else ‘Very poor’

The Hotel rating description is the overall hotel rating, in this case, 2 out of 10 hotels are “Superb”, 5 out of 10 are “Good” and 3 out of 10 are “Okay”; while, reviewers have rated the hotels along their reviews. Their rating is varying from poor to Superb. The distribution of reviewers’ rating is summarised in Table 2.

TABLE 2. DISTRIBUTION OF REVIEWERS’ RATING

Reviewers’ rating	# Reviews	% Reviews
Superb	73	37
Good	81	41
Okay	30	15
Poor	12	6
Very Poor	4	2

The comparison of the three levels of hotel review description is also set out in the bar chart in Figure 2. The ‘superb’ review level (78%) performs the highest percentage of customer review as well as there is no review in ‘poor’ and ‘very poor’ levels in ‘superb’ hotels. Looking back at hotels rated ‘good’ and ‘okay’, the majority of customers who stayed at those hotels rated their experiences as good, at 48% and 43% respectively, and very few customers of hotels labelled ‘good’ and ‘okay’ wrote reviews indicating the hotel was poor or very poor.

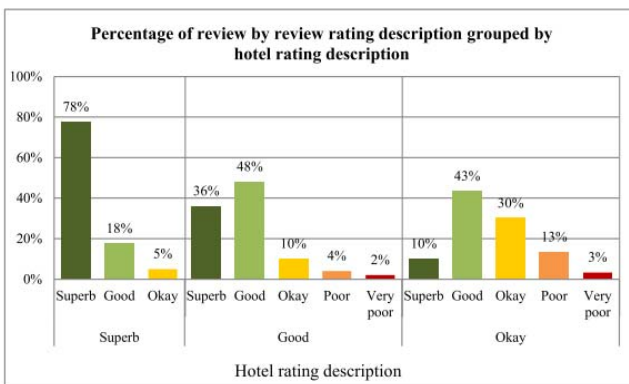


FIGURE 3. THE PERCENTAGE OF EACH VALUE IN REVIEW RATING DESCRIPTION

We use text mining to analyse both positive and negative comments and extract useful patterns about different aspects of hotels. SAS Enterprise Miner has been employed to perform text mining analysis. Textual data is typically available only in readable document format such as customer reviews. Text mining is the process which generates a quantitative representation of documents. Once the text is transformed into a set of numbers that capture the patterns in the textual data, any statistical model, machine learning or data mining model can be used to for generating insights. A typical text mining project involves (i) text parsing and transformation, (ii) text filtering, (iii) text mining.

The first step of text mining is parsing textual data by applying a natural language processing algorithm. In general, the aim of this step is to transform customer reviews, unstructured data, into structured data (quantitative representation) for ease of further analysis. Text parsing is a step for extracting words from the whole dataset, cleaning the dataset and creating a dictionary of extracted words. Parsing text also identifies sentences within individual customer reviews, and defines part of speech and stem words [24]. Normalisation and stemming of words is performed to reduce the complexity of inputs. This process also involves semantically equivalent words and synonyms. For example, “staff” and “receptionist” are grouped in the same group as they have a similar meaning. Removing auxiliary including stop words such as “a” and “the” is another task performed in this stage. This study then removes abbreviations, auxiliaries or modals, conjunctions, determiners, interjections, numbers or numeric expressions, infinitive markers, negative participles, possessive markers, prepositions and pronouns. It also eliminates term characters, including numbers and terms including a punctuation character. By doing this, the frequency of unique words and complexity will be reduced without loss of information.

The following step is to assign a part of speech tag to each word to undertake linguistic analysis and extract more sophisticated features by labelling a word with the appropriate part of speech in its sentence.

Text filter is a task performed after the text parsing step. In a corpus of good size documents, it is very likely to have many terms that are irrelevant to either differentiating documents from each other or to summarising the documents. Text filter attempts to reduce the number of parsed words analysed in the next step. Unnecessary information is eliminated from the data. Therefore, only valuable and pertinent information is taken into the further analysis [27]. Terms which are not relevant to differentiate and summarise documents are eliminated in this step. The process of checking word spelling is also carried out in this step as users can manually add additional synonyms in order to increase analysis precision. This research obtained a dictionary file from the OpenOffice US English spelling dictionary (http://extensions.services.openoffice.org/en/project/en_US-dict), as mentioned in SAS (2014). Terms which are not relevant to differentiate and summarise documents are eliminated in this step. The algorithm in this step calculates the frequency of each word and then deletes those which are too simple, as per the recommendation by Chakraborty, Pagoluand Garla (2013). Words with either too

high or too low frequency are considered as bad discriminators. On the other hand, the best discriminator is a word that has neither high nor low frequency; ideally, the frequency should be between a number of documents/10 and number of documents/100 (Chakraborty, Pagolu and Garla, 2013). Moreover, documents composed of none of the filtered words are removed.

In the next step, Text Topic is performed on the output of previous steps. The Text topic procedure is used to analyse, summarise and create a topic to represent similar theme or ideas in the dataset. The algorithm in the background of this functionality is rotated singular value decomposition (SVD) to create multi-term topics. The number of topics is equal to the number of SVD dimensions.

At the end of this stage, information is transformed from unstructured to structured format. Therefore, it is possible to build and evaluate a model which shows the main hotel aspects and how this reflects on the provided rating on Booking.com.

Correlation analysis between outputs from transformed textual data is performed to reduce the number of inputs and eliminate redundancy of all factors. Statistical models will be built subsequently. Two techniques are initially selected to perform this part: decision tree and logistic regression. Imbalanced data were adjusted before applying models. Because the minority of target variables is ‘dissatisfied’ at 22.3%, the probability was adjusted to 50%. The target variable is binary, therefore logistic regression has been chosen. Various models have been built and analysed to find the best set of models which can predict the important factors affecting customer satisfaction (dependent variable).

In the next stage, model evaluation is carried on using Model Comparison functionality in SAS Enterprise Miner. This study needs to investigate which aspects affect customer satisfaction with hotels, it was necessary to predict ‘Satisfied’ as ‘Satisfied’ as well as ‘Dissatisfied’ as ‘Dissatisfied’. Thus, the AUC (area under the Receiver Operating Characteristic (ROC) curve) will be a criterion for selecting the best model.

IV. RESULTS

After applying Text Parsing step on both positive and negative comments, all actionable terms which could help us to identify the hotels’ aspects and the associated sentiment are identified and separated based on their role. It is clear that noun, noun expressions and adjectives are the main roles in sentiment analysis. The results of the Text Parsing step is shown in Table 3 for both positive and negative comments.

TABLE 3. FREQUENCY TABLE OF WORDS EXTRACTED FROM TEXT PARSING STEP

Role	Negative comments		Positive comments	
	#Terms	%Terms	#Terms	%Terms
Adjective	190	15.22	243	16.51
Adverb	59	4.73	121	8.22
Noun	464	37.18	509	34.58
Noun expressions	283	22.67	404	27.44
Verbs	252	20.20	195	13.25
Total	1,248	100	1,472	100

After applying Text Filter and Text Topic steps, interesting topics in both compliment and complaint reviews are generated. Topics are generated using Term Cutoff and Document Cutoff metrics. Term Cutoff means minimum topic weight that a term must have to be used as a term for this topic. The weight of any term that has an absolute value weight less than this is effectively set to zero. Document Cutoff means minimum topic membership that a document must have to be included in this topic.

By using [14] and [15] as references for naming aspects of customer satisfaction, seven aspects for negative comments were found in our 10 random three-star hotels located near Khao San Road: ‘Hotel physical attributes’, ‘Sleep quality’, ‘Human resources’, ‘Payment’, ‘Room and bathroom interiors’, ‘Complementary service’ and ‘Cleanliness’. Six aspects of customer satisfaction are revealed from positive comments: ‘Location’, ‘Human resources’, ‘Sleep quality’, ‘Room and bathroom interiors’, ‘Cleanliness’ and ‘Stay value’.

Examples of Text Topic step is provided in Table 4.

TABLE 4. EXAMPLE OF RESULTS FROM TEXT TOPIC STEP

Words under topic	Topic Description	Satisfaction Aspect	Positive/negative review
+night, +noise, +room, +window, +sleep	Noise	Sleep Quality	Negative review
+clean, +old, +clean, +room, +window, +night	Cleanliness	Cleanliness	Negative review
+helpful, +staff, +attraction, +price, +stay	Helpful staff	Human Resource	Positive review
+price, + as well, attraction, +clean room, + great, + good	Hotel provides good price	Stay Value	Positive review

A. Topic comparison:

To sum up, after both negative and positive comments implement text mining process, nine aspects are output from both sides of the reviewed data: ‘Hotel physical attributes’, ‘Sleep quality’, ‘Human resources’, ‘Payment’, ‘Room and bathroom interiors’, ‘Complementary service’, ‘Cleanliness’, ‘Location’ and ‘Stay value’. Seven topics are found in negative comments and six in positive comments. Four topics are identified in both sides of comments; sleep quality, human resource, room and bathroom interiors, and cleanliness. Complementary service, hotel physical attributes and payment are only found in the negative topic and stay value is discovered in the positive topic only.

Human resource aspect is appeared in both types of reviews. Surely, the negative human resource topic is about complaining staff’s manner, especially poor attitude, rudeness, and not helpful, and some customers felt neutral. Notwithstanding, the customer tends to satisfy when staff is willing to help them, friendly, polite as detected in positive topics.

The following topic discovered from compliment and complaint is room and bathroom interiors. Size and cleanliness of room are mentioned in positive and negative reviews whilst decoration can be found in only positive reviews.

The last topic which is revealed in both ways of comments is cleanliness. Room hygiene is mentioned in both positive and negative comments. Cleanliness of the whole hotel is seen in only positive reviews.

It is interesting that customers wrote only complaints regarding complementary services such as Wi-Fi or pool and did not compliment them. They tend to feel neutral when the services meet their expectation. Also, hotel physical attribute is found only in a negative review. It means that old building made the customer feel unhappy. On the contrary, location is revealed only in positive comments.

Next, three decision tree models and 12 logistic regression models are performed, where target variable is “customer satisfaction”. The dataset is divided into two parts: 70% is a training dataset and 30% is a validation dataset. The training dataset is used as an input to develop the model using several techniques, especially regression, and decision tree; while the validation dataset is used to choose the best models among those built by the techniques mentioned above in the model evaluation step. Due to the level of target variables, *review_rating* (*Review_rating_desc2lv*) is binary, and the partitioning method is set to Stratified. The next step is to deal with missing values and outliers.

A ROC Index of validation dataset is selected as a criterion to finalise the best model. A ROC Curve is a plot illustrating how well the model classifies binary target variables between true positive rate (TP) and false positive error rate (FP). The higher the ROC, the better the model performance. Therefore, the best model is logistic regression with ROC= 0.905. This model is built by using a backward model selection technique; only 19 topic factors can be utilized. It is implied that this regression model is excellent in order to separate ‘Dissatisfy’ from the whole dataset.

The output of regression model is presented in Table 5. Four elements from the figure above are emphasised in this section: ‘Variable’, ‘Level’, ‘Coefficient’ and ‘Effect Number’. Firstly, the variable column shows the list of factors affecting customer satisfaction with hotel service. This model is composed of 19 variables: 10 topics from negative comments (variable named *TextTopic2_x*) and nine from positive

comments (variable named *TextTopic6_x*). Secondly, the level column displays the value from the variable column. Each topic column contains two distinct levels: 0 and 1. ‘0’ means that a particular review is not categorised in that topic and ‘1’ means that a review is categorised in that topic. It should be noted that a review can be categorised in more than one topic [27]. Thirdly, the coefficient shows the direction of effect from factors to the customer’s hotel satisfaction: a positive value displays a positive effect, and a negative value displays a negative effect. For example, ‘*TextTopic6_5*’ variable or human resources – friendly staff topic, which its level value is equal to 0 (review is not grouped in this topic), has a negative effect. It can be implied human resources (friendly staff) has a positive association with the target variable.

1) *Human resources aspect*

It can be easily seen that human resources, focusing on how friendly the staff is (effect number=1), has the biggest positive effect on hotel rating, and helpful staff (effect number=18) slightly impacts on hotel satisfaction. This study also obtained some evidence to support Berezina et al.’s idea, as shown in negative topic 4. Reference [23] argue that this factor can indicate both positive and negative comments. This topic is about complaining about staff’s manners and negatively affects hotel satisfaction rate [23].

2) *Complementary service*

The first important pessimistic rank concerns the complementary service provided by the hotel. Two services were mentioned in this study: pool and Wi-Fi. Both types of complementary service are detected in complaints, and are causes of customer dissatisfaction. They both negatively impact hotel ratings. Reference [14] also report that Wi-Fi is a cause of dissatisfaction and [15] experienced the same issue about pools from disappointed customers.

3) *Room and bathroom interiors*

Room and bathroom interiors, especially complaints about rooms, are the next important factor with a negative relationship with hotel rating. As reference [28] found, it is expected that room and bathroom interiors will be found, as this topic can be counted as a core of the hotel industry.

TABLE 5. INFLUENTIAL ASPECTS

Topic ID	Satisfaction aspect	Coefficient	Effect number
Positive Topic 5	Human resources- friendly staff	6.086	1
Negative Topic 8	Complementary service- pool	-1.454	3
Negative Topic 6	Room and bathroom interiors	-1.16	4
Positive Topic 1	Location- near Khao San Road	-1.061	6
Positive Topic 8	Cleanliness	1.052	7
Positive Topic 2	Location- Good location	-0.979	8
Negative Topic 1	Hotel physical attributes	-0.918	9
Positive Topic 7	Location- attraction	0.896	10
Negative Topic 4	Human resources	-0.816	11
Positive Topic 9	Sleep quality- noise	-0.754	12
Negative Topic 3	Sleep quality- bed	-0.454	13
Negative Topic 2	Sleep quality- noise	0.366	14
Negative Topic 10	Cleanliness	-0.259	15
Positive Topic 6	Room and bathroom interiors- bathroom	0.197	16
Negative Topic 5	Payment	0.184	17
Positive Topic 3	Human resources- helpful staff	0.152	18
Positive Topic 10	Stay value	0.098	19
Negative Topic 7	Room and bathroom interiors- bathroom	0.005	20

Comments about room and bathroom interiors focusing on rooms are detected in two sides of review as [23] mentioned in their studies. As a result, negative room and bathroom interiors focusing on room contribute to customer dissatisfaction, and the positive one has a positive effect on rating hotel. However, discussion of bathrooms is found more frequently in complaints [27]; [14], there is a minimal positive impact from the negative aspect of room and bathroom interiors as regards bathrooms (Negative topic 7) as can be seen from the value of the coefficient

4) Location

References [28], [27] and [14] propose similar results, that location has a positive influence on customer satisfaction. [14] also suggest that location has much more influence than room and bathroom interiors.

However, only one location aspect focusing on attractions nearby positively affects hotel ratings among the three sub-aspects in the location topic. It might be because all 10 random hotels in this study are located in the same area, Khao San Road. Therefore, the location of the hotel has no positive impact on increasing customer satisfaction

5) Cleanliness

Another common concern aspect of customers' expectation is cleanliness [14]. Cleanliness can drive both positive and negative ratings [27], as shown above. The positive cleanliness aspect definitely contributes to higher hotel rating. Similarly, the negative cleanliness aspect is negatively associated with hotel rating.

6) Sleep quality

Another interesting aspect is sleep quality. Sleep quality is one of the customers' needs [15]. Reference [15] suggest that sometimes sleep quality cannot contribute much to satisfaction, but its absence will affect dissatisfaction. Negative sleep quality aspect which is commented about bed (negative topic3) and positive sleep quality focusing on noise (negative topic9) contribute hotel dissatisfaction of customer. The interesting thing is that the negative sleep quality aspects (negative topic2), which focused on noise, has a positive relationship with customer satisfaction due to the fact that Khao San Road is a night street and travellers might know this before booking these hotels. Therefore, it is acceptable to travellers that they hear some noise from outside.

Table 6 below is a summary table of the analysis mentioned above. It includes customer's hotel satisfaction aspects and the direction of their relationship with hotel rating. Certain aspects can contribute both positive and negative association between such aspects and hotel rating.

V. DISCUSSION AND IMPLICATIONS OF RESULTS:

The results obtained here demonstrate that different aspects of perceived service quality play different roles in determining the overall satisfaction from the service. The presence of some aspects is a strong predictor for satisfaction, the presence of others is a strong predictor of dissatisfaction, and a third type influence strongly both ways. Table 6 above illustrates this finding in the target domain of hotel service.

TABLE 6. SUMMARY TABLE OF DIRECTION OF RELATIONSHIP BETWEEN CUSTOMER SATISFACTION ASPECT AND HOTEL RATING

Topic	Satisfaction Aspects	Relationship with hotel rating	
		Positive	Negative
1	Human resource – friendly and helpful staff	√	
2	Human resource – staff's manner		√
3	Complementary service – pool and Wi-Fi		√
4	Room and bathroom interiors - room	√	√
5	Room and bathroom interiors - bathroom	√	
6	Location - attraction	√	
6	Cleanliness	√	√
7	Sleep quality - bed		√
8	Sleep quality - noise	√	√

This analysis allows us to construct on-line information systems for gathering and processing customer feedback. The findings also allow us to construct effective systems for monitoring and analysis of social media, which focus on the important topics of eWOM to deliver up-to-date information about customer satisfaction to hotel management.

Indeed, from the hotels managers' perspective, our findings represent important steps toward understanding traveller satisfaction factors based on online consumers' perceptions of their hotel and ultimately, understand the importance of online user reviews of their hotel. Once hotels' managers have better understanding of customers' satisfaction or dissatisfaction; they can improve their service quality. Specifically, customer satisfaction leads to customer loyalty and positive eWOM and by extension improving hotel's profitability [29]. These factors are very important for hotel industry which is highly competitive and hotels need to distinguish themselves among competitors in a very specific location.

Kerr, Harges and Kim's study [27] establishes that decision making of customers tends to rely mostly on the negative reviews and less on the positive ones. Hotel managers should therefore take special care about aspects of the hotel service which are linked with customer dissatisfaction. Ensuring faultless service along these aspects will increase the likelihood of positive reviews and allow the sustained growth of the business. For example, complimentary facilities should be paid close attention to regardless of the fact that their use is not chargeable to hotel guests. Areas to monitor closely are the cleanliness around the pool area and speed and reliability of Wi-Fi, which should be complimentary since giving free service has a greater association with high hotel rating [15].

Staff performance is the most influential topic contributing to customer satisfaction [11]; [28]; [14]; [23] and indeed it is considered as a core product of any hotel [28]. Hoteliers should therefore offer comprehensive training to their employees in order to improve their personality, ability to serve customer, professionalism, and manner.

The appearance of cleanliness topic and room and bathroom interiors topic are useful for the hotel manager to properly allocate employees to take responsibility for cleanliness and neatness of bathroom, room and in-room amenities especially

bed and bedsheets. Moreover, cleanliness of the whole hotel is also important.

The final topic which can inform management action regards improving the customers' sleep by soundproofing or other related measures such as blackout curtains.

VI. CONCLUSION

Motivated by the ever increasing information on hotels on the Internet, we investigated the degree of customer satisfaction and the impact on that satisfaction borne by different hotel aspects. This was done through the innovative text mining technique of topic modelling, allowing us to form a holistic picture of these dependencies without any preconceptions about what these important aspects could be.

Forming such a holistic picture allows us to create information systems and managerial procedures of systematic gathering, monitoring and analysing online customer reviews and social media posts for key aspects, and estimating how changes along these key aspects will result in overall customer satisfaction or otherwise. Controlling these aspects will assist hotels in building brand and customer acquisition with low cost. Hotel managers can understand customers' reaction to specific aspects of their hotels and fix the area that they underperform.

Finally, this study had some limitations in terms of scope which was limited to a single tourist area thus achieving homogeneous sample. As a side-effect of our scope limitation and the need to achieve homogeneous sample of hotels, we also obtained a limited sample of only 10 hotels. These limitations should be addressed in future studies.

A. References

- [1] M. Sigala, "Web 2.0, social marketing strategies and distribution channels for city destinations: enhancing the participatory role of travellers and exploiting their collective intelligence," in *M. Gascó-Hernández, T. Torres-Coronas (Eds.), Information communication technologies and city marketing: digital opportunities for cities around the world.*, IDEA Publishing, 2008.
- [2] H. Li, Q. Ye, and R. Law, "Determinants of Customer Satisfaction in the Hotel Industry: An Application of Online Review Analysis," *Asia Pacific J. Tour. Res.*, vol. 18, no. 7, pp. 784–802, 2013.
- [3] S. E. Levy, W. Duan, and S. Boo, "An Analysis of One-Star Online Reviews and Responses in the Washington, D.C., Lodging Market," *Cornell Hosp. Q.*, vol. 54, no. 1, pp. 49–63, 2013.
- [4] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tour. Manag.*, vol. 59, pp. 467–483, 2017.
- [5] V. P. Magnini, J. C. Crotts, and A. Zehrer, "Understanding customer delight: An application of travel blog analysis," *J. Travel Res.*, vol. 50, no. 5, pp. 535–545, 2011.
- [6] C. C. Yang, Y. T. Jou, and L. Y. Cheng, "Using integrated quality assessment for hotel service quality," *Qual. Quant.*, vol. 45, no. 2, pp. 349–364, 2011.
- [7] T. Lockyer, "The perceived importance of price as one hotel selection dimension," *Tour. Manag.*, vol. 26, no. 4, pp. 529–537, 2005.
- [8] C. H. Lin and B. Z. L. Kuo, "The Behavioral Consequences of Tourist Experience," *Tour. Manag. Perspect.*, vol. 18, pp. 84–91, 2016.
- [9] X. Sun, C. Geng-Qing Chi, and H. Xu, "Developing destination loyalty: The case of hainan island," *Ann. Tour. Res.*, vol. 43, no. 40871060, pp. 547–577, 2013.
- [10] B. A. Sparks and V. Browning, "The impact of online reviews on hotel booking intentions and perception of trust," *Tour. Manag.*, vol. 32, no. 6, pp. 1310–1323, 2011.
- [11] I. K. W. Lai and M. Hitchcock, "Sources of satisfaction with luxury hotels for new, repeat, and frequent travelers: A PLS impact-asymmetry analysis," *Tour. Manag.*, vol. 60, pp. 107–129, 2017.
- [12] V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog, *Text Mining in Organizational Research*. 2017.
- [13] D.-H. Park, J. Lee, and I. Han, "The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement," *Int. J. Electron. Commer.*, vol. 11, no. 4, pp. 125–148, 2007.
- [14] D. Buhalis and R. Law, "Progress in information technology and tourism management: 20 years on and 10 years after the Internet-The state of eTourism research," *Tour. Manag.*, vol. 29, no. 4, pp. 609–623, 2008.
- [15] M. P. O'Mahony and B. Smyth, "A classification-based review recommender," *Res. Dev. Intell. Syst. XXVI Inc. Appl. Innov. Intell. Syst. XVII*, vol. 23, no. 4, pp. 49–62, 2010.
- [16] Q. Ye, R. Law, B. Gu, and W. Chen, "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," *Comput. Human Behav.*, vol. 27, no. 2, pp. 634–639, 2011.
- [17] I. E. Vermeulen and D. Seegers, "Tried and tested: The impact of online hotel reviews on consumer consideration," *Tour. Manag.*, vol. 30, no. 1, pp. 123–127, 2009.
- [18] A. M. Susskind, "I Told You So ! Restaurant Customers ' Word-of-Mouth Communication Patterns I Told You So ! Restaurant Customers ' Word-of-Mouth Communication," *Hotel Restaur. Adm. Quartely*, vol. 43, pp. 75–85, 2002.
- [19] C. Li, G. Cui, and L. Peng, "The signaling effect of management response in engaging customers: A study of the hotel industry," *Tour. Manag.*, vol. 62, pp. 42–53, Oct. 2017.
- [20] T. Shanka and R. Taylor, "An Investigation into the Perceived Importance of Service and Facility Attributes to Hotel Satisfaction," *J. Qual. Assur. Hosp. Tour.*, vol. 4, no. 3–4, pp. 119–134, Feb. 2004.
- [21] C. Greathouse, K. R., Gregoire, M. B., Shanklin, C. W. and Tripp, "Factors considered important in hotel accommodations by travelers stopping at visitor information centers," *Hosp. Res. J.*, vol. 19, no. 4, pp. 129–140, 1996.
- [22] T. Lockyer, "Hotel cleanliness - How do guests view it? Let us get specific. A New Zealand study," *Int. J. Hosp. Manag.*, vol. 22, no. 3, pp. 297–305, 2003.
- [23] P. Racherla, D. J. Connolly, and N. Christodoulidou, "What

- Determines Consumers' Ratings of Service Providers? An Exploratory Study of Online Traveler Reviews," *J. Hosp. Mark. Manag.*, vol. 22, no. 2, pp. 135–161, 2013.
- [24] A. Barreda and A. Bilgihan, "An analysis of user-generated content for hotel experiences," *J. Hosp. Tour. Technol.*, vol. 4, no. 3, pp. 263–280, 2013.
- [25] R. B. Woodruff, E. R. Cadotte, and R. L. Jenkins, "Modeling Consumer Satisfaction Processes Using Experience-Based Norms," *J. Mark. Res.*, vol. 20, no. 3, p. 296, 1983.
- [26] Z. Xiang, Z. Schwartz, J. H. Gerdes, and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?," *Int. J. Hosp. Manag.*, vol. 44, pp. 120–130, 2015.
- [27] C. H. J. Wu and R. Da Liang, "Effect of experiential value on customer satisfaction with service encounters in luxury-hotel restaurants," *Int. J. Hosp. Manag.*, vol. 28, no. 4, pp. 586–593, 2009.
- [28] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, "Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews," *J. Hosp. Mark. Manag.*, vol. 25, no. 1, pp. 1–24, 2016.
- [29] G. Chakraborty, M. Pagolu, and S. Garla, "PREVIEW: Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS," *Text Min. Anal. Pract. Methods, Examples, Case Stud. Using SAS*, pp. 1–38, 2013.
- [30] Wikitravel, "Bangkok/Khao San Road,," 2017. [Online]. Available: http://wikitravel.org/en/Bangkok/Khao_San_Road.
- [31] Booking, "About Booking.com," 2017. [Online]. Available: <https://www.booking.com/content/about.en-gb.html?label=gen173nr-1FCAEoggJCAIhYSDNiBW5vcmVmaFCIAQGYAS7CAQp3aW5kb3dzIDEwyAEM2AEB6AEB-AELkgIBeagCAw;sid=18a0bff7d70261e767b0d8abff9f5829>.
- [32] P. M. Herr, F. R. Kardes, and J. Kim, "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *J. Consum. Res.*, vol. 17, no. 4, p. 454, 1991.
- [33] P. O'Connor, "Managing a hotel's image on Tripadvisor," *J. Hosp. Mark. Manag.*, vol. 19, no. 7, pp. 754–772, 2010.
- [34] B. Bartikowski and S. Llosa, "Customer satisfaction measurement: Comparing four methods of attribute categorisations," *Serv. Ind. J.*, vol. 24, no. 4, pp. 67–82, 2004.